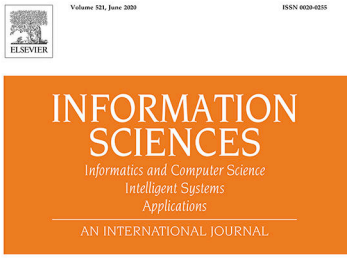
Journal Pre-proofs



Spatial rough set-based geographical detectors for nominal target variables

Hexiang Bai, Deyu Li, Yong Ge, Jinfeng Wang, Feng Ca

PII: S0020-0255(21)01240-8
DOI: https://doi.org/10.1016/j.ins.2021.12.019
Reference: INS 17098

To appear in: *Information Sciences*

Available online at www.sciencedirect.com
ScienceDirect

Received Date: 26 January 2020
Revised Date: 4 December 2021
Accepted Date: 6 December 2021

Please cite this article as: H. Bai, D. Li, Y. Ge, J. Wang, F. Ca, Spatial rough set-based geographical detectors for nominal target variables, *Information Sciences* (2021), doi: https://doi.org/10.1016/j.ins.2021.12.019

Spatial rough set-based geographical detectors for nominal target variables

Hexiang Bai[a,*], Deyu Li[a,b], Yong Ge[c], Jinfeng Wang[c], Feng Cao[a]

[a]*School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China.*
[b]*Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China.*
[c]*State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.*

Abstract

Spatial rough set theory is an extension of classical rough set theory that is designed to handle spatial data. Different from its classical analog, spatial rough set theory approximates nominal target variables in each local region, thereby enabling the local explanatory power of various factors on these nominal target variables to be determined. Based on spatial rough set theory, this paper describes two new uncertainty indices for measuring the spatial heterogeneity of the local explanatory power on nominal target variables. Moreover, three new spatial rough set-based geographical detectors are proposed. With the help of the uncertainty indices, these detectors measure the spatial explanatory power of different factors, compare their importance, and detect feature interactions. Experiments are conducted using two publicly accessible datasets to demonstrate the effectiveness of the proposed geographical detectors and evaluate their performance against that of geographical detectors based on q-statistics. The results show that the proposed detectors are effective in processing nominal variables and can be used to complement existing q-statistics-based geographical detectors.

Keywords: Rough set, Uncertainty measure, Spatial heterogeneity, Spatial analysis, Geographical detector

1. Introduction

In modeling our world, observations of the same condition may lead to different responses; that is, when the conditions are not sufficient to explain the variable of interest completely, there may be some inconsistency between the condition features and the resulting decisions. A typical example encountered when interpreting remotely sensed images is that different objects may have the same spectrum. Undoubtedly, this introduces uncertainties in explaining the target objects.

[31] introduced the concept of roughness to explain this type of uncertainty, and proposed the use of rough sets to handle roughness in data. Rough sets construct knowledge granules (equivalence classes) using conditional features and explain the target objects using a lower

*Corresponding author. Tel.: +86 351 7010566; fax: +86 351 7018176. E-mail: baihx@sxu.edu.cn.

approximation, which includes knowledge granules that can completely explain the target objects, and an upper approximation, which includes knowledge granules that can completely or partly explain the target objects [28]. The idea of rough sets has been extended to incorporate the characteristics of different types of data, such as rough fuzzy sets and fuzzy rough sets [30], variable precision rough sets [45], probability rough sets [48], and composite rough sets [46, 47].

Various applications of rough set theory to spatial data have been described in the literature. Similar to other types of data, the application of rough sets to spatial data has mainly focused on three aspects. First, the reduct in rough sets is an effective feature selection tool for spatial data. For instance, [20] showed that the reduct is superior to principal component analysis in the classification of remote sensing images, and presented hyperspectral band selection methods based on the reduct that offer advantages over traditional methods. In addition, [15] simplified the reduct of fuzzy rough sets in terms of the requirement for quantitative decisions in cartographic generalization.

Second, rough set theory is an effective tool for extracting decision rules and classification tasks. For example, [43] used classical rough sets to extract spatiotemporal rules for characterizing river eutrophication, while [42] applied the rules extracted by rough sets to generate intelligent initial map scales. In addition to classical rough sets, other rough set extensions have been used for rule extraction and classification, such as tolerance relation-based rough sets [44], variable precision rough sets [29], and rough fuzzy sets [5]. Moreover, rough sets have been extended to the classification of spatial data from the perspective of multiple scales [41, 22] and spatial heterogeneity [6].

Third, rough set theory provides effective tool sets for measuring the roughness uncertainty in spatial data, that is, quantifying the explanatory power of the conditional features on the target variable. Many traditional roughness measures, such as the approximation quality [12], information entropy [12, 23], combination entropy [32], and multi-classification of quality [8], have been improved by considering the characteristics of spatial data. For example, [2, 3, 1] proposed a series of uncertainty measures for rough classification and used them to translate between land cover taxonomies. [17] proposed a rough set-based sample quality measure for remote sensing images, while [33, 7] used a similarity measure of rough sets to compare the roughness between areas.

Measuring roughness uncertainty is the foundation of applications of rough set theory to spatial data. For example, the reduct algorithm mainly depends on the measure selected for characterizing the roughness uncertainty. However, current roughness measures characterize the explanatory power of conditional features in the entire study area. They ignore differences in the explanatory power of conditional features in different local regions, that is, the spatial heterogeneity of conditional features' explanatory power.

Spatial heterogeneity refers to the variation of geographical phenomena over space [21]. This is an important topic in the study of geographical phenomena, such as the analysis of populations, communities, ecosystems, and landscapes [34]. Spatial heterogeneity has different manifestations in spatial data. For example, the spatial heterogeneity of point pattern datasets can be defined as the degree of the aggregation-type deviation from complete spatial

randomness [35]. The spatial heterogeneity of a surface pattern refers to the variation of a qualitative or quantitative value over space [13, 18]. The spatial heterogeneity of conditional features' explanatory power indicates that the explanatory power of conditional features at one site/stratum is different from that at other sites/strata [38]; that is, the conditional feature set explains the target variable to varying degrees at different locations.

Currently, two spatial statistical methods can be used to characterize the spatial heterogeneity of the explanatory power of conditional features. One is geographically weighted regression (GWR) [16, 49], which quantifies the spatial local heterogeneity. GWR uses the local correlation between the dependent variable and target variable, or local coefficients of different dependent variables, to inspect the distribution of the local explanatory power of conditional features in different areas [26]. However, GWR is only suitable for continuous variables and suffers from collinearity issues when there are many nominal conditional features [40]. Furthermore, GWR does not provide an overall measure of the spatial heterogeneity of the explanatory power of conditional features.

The second approach involves the use of q-statistics-based geographical detectors (q-GD) [37, 38], which have been employed in many real-life applications [27, 39, 36]. The q-statistic calculates whether the target variable differs greatly in each geographical stratum (equivalence class) formed by the conditional features [37]. A smaller standard deviation of the target variable in each stratum indicates that the conditional variables provide a better determination (explanation) of the target variable [38]. [37] showed that q-GD can be used to detect the most relevant conditional features, compare their importance, and detect the interactions between them. However, although q-GD provides an overall measure, it is designed for continuous target variables and is not suitable for nominal target variables. Moreover, q-GD only measures the explanatory power from the viewpoint of strata, and ignores the variation of the explanatory power at different locations.

Although GWR and q-GD inspect the spatial explanatory power from different perspectives, neither is designed for nominal target variables. However, nominal variables or categorical data are inevitable in representations of geographical data, such as for the distribution of different types of crime, vegetation, soil type, and human species. Recently, a new spatial extension of rough set theory, named spatial rough set theory [6], has provided an option for measuring the spatial heterogeneity of the local explanatory power of conditional features on nominal target variables. Unlike other extensions of rough set theory, spatial rough set theory does not approximate the target variable over the entire study area, but instead focuses on each local region. This enables the local explanatory power of conditional features on nominal target variables to be quantified and its spatial heterogeneity to be determined.

Using the concept of spatial rough sets, this paper describes two new measures for quantifying the spatial heterogeneity of the local explanatory power of nominal target variables. Three new spatial rough set-based geographical detectors (SRS-GD) are proposed using these two measures. SRS-GD and q-GD inspect the explanatory power from two different perspectives. SRS-GD is designed for nominal target variables, and inspects the spatial heterogeneity of conditional features' local explanatory power from the perspective of local roughness, whereas

q-GD is designed for continuous target variables and inspects the global explanatory power from the perspective of the overall difference in variance before and after the study area is stratified. Similar to q-GD, SRS-GD can perform the following tasks for nominal target variables:

(1) Measure the local explanatory power and detect its spatial heterogeneity.
(2) Compare the average local explanatory power of conditional features.
(3) Detect the interactions among conditional features.

The remainder of this paper is organized as follows. First, some basic knowledge about spatial rough sets is reviewed. The local positive region-based approximation quality is then used to establish indices for the average local explanatory power and its spatial heterogeneity. Three new geographical detectors are proposed for the above-mentioned spatial data analysis tasks. Finally, two publicly accessible datasets are used to further illustrate the performance of the proposed geographical detectors and compare them with q-GD.

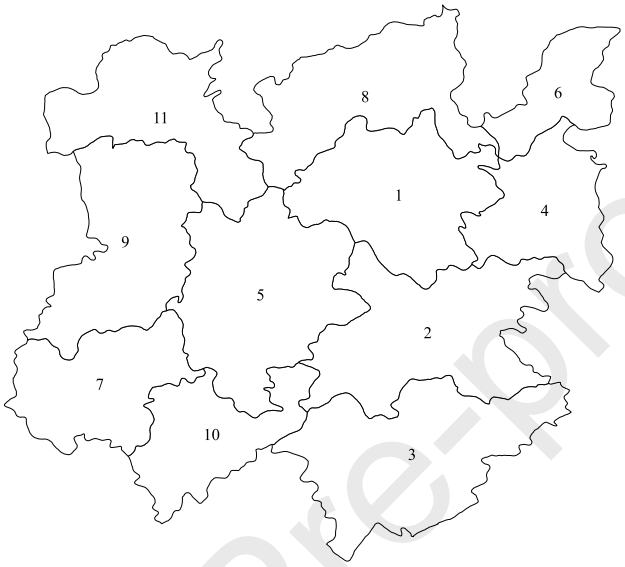
2. Spatial rough set model [6]

Spatial rough set theory uses a spatial information system (SIS) to model spatial data. An SIS consists of an information table and an adjacency matrix. In the information table, each row represents a geographical object and each column represents one feature of the geographical objects. The adjacency matrix is used to model the spatial relations among objects. Formally, an $SIS = \{U, A, d, M\}$ consists of the following:

- a universe U, which is the set of all the geographical objects,
- a set V consisting of feature values,
- a set A of named functions from U to V,
- a function $d : U \to V_d$ (the decision feature or target variable, having domain V_d),
- an adjacency function $M : U \times U \to 0, 1$, which records whether two objects are considered to be neighbors in an SIS.

By a "named function" in A, we mean a pair (a, a_f), where a belongs to a set of feature names and a_f is a function $a_f : U \to V$. To avoid complicating the notation, the same symbol is used for the feature name and for the function itself. That is, A is not a set of functions, but a multi-set with named elements. Given an ordering of the elements of U, the adjacency can be represented by a matrix M. M_{ij} denotes the ijth entry in M. If the ith and jth objects are adjacent, then $M_{ij} = 1$; otherwise, $M_{ij} = 0$.

Figure 1 shows an example of an SIS. Figure 1(a) is an example study area from which an adjacency matrix (see Figure 1(c)) can be established. In the figure, two objects are neighbors if they are touching. Figure 1(b) is the corresponding information system. Items 'a1' to 'a3' in the first row represent the conditional feature names; 'd' represents the decision feature; 'a' to 'c' in the remaining rows are the corresponding feature values for different objects. The information system and adjacency matrix form an SIS of the study area.



(a) Map of an example study area

ID	a1	a2	a3	d
1	a	b	a	a
2	b	a	a	a
3	b	a	a	a
4	b	c	a	b
5	b	c	a	a
6	c	b	c	a
7	c	c	b	b
8	d	c	b	b
9	c	b	c	b
10	a	b	a	b
11	a	b	a	b

(b) Example of an information system

$$M = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

(c) Adjacency matrix

Figure 1: Example of an SIS: (a) map of the geographical objects of the study area; (b) information system corresponding to the geographical objects in (a); (c) adjacency matrix of geographical objects in (a).

In an SIS, the spatial rough set model uses an indiscernibility relation, which is closely related to the object locations. For an x-centered region $C(x) = \{y \in U : M_{xy} = 1\} \cup \{x\}$, two objects y and z are said to be x-local indiscernible if $a(y) = a(z)\, \forall a \in A$ and $y, z \in C(x)$. Using the local indiscernibility relation, objects that are spatially distant from each other can be discerned, regardless of whether they have different feature values. Based on the x-local indiscernibility relation, an x-local indiscernible set of y using $A' \subseteq A$ can be defined as

$$LInd(y, x, A') = \begin{cases} \varnothing, & \text{if } y \notin C(x) \\ \{z \in C(x) : \forall a \in A'(a(y) = a(z))\} = [y]_{A'} \cap C(x), & \text{if } y \in C(x) \end{cases}$$

where $[y]_{A'} = \{z \in U : a(y) = a(z), \forall a \in A'\}$.

Similar to classical rough sets, spatial rough set theory uses x-local indiscernible sets to approximate the target concept $X \subseteq U$ in each x-centered region. The lower approximation contains all of the x-local indiscernible sets that belong to the target concept, whereas the upper approximation contains all x-local indiscernible sets that have a non-empty intersection with the target concept in each x-centered region. Formally,

$$\underline{appr}^{x}_{A'}(X) = \{y \in C(x) : LInd(y, x, A') \subseteq X\}$$
$$\overline{appr}^{x}_{A'}(X) = \{y \in C(x) : LInd(y, x, A') \cap X \neq \varnothing\}.$$

The x-local rough set of target concept X using $A' \subseteq A$ is a pair

$$(\underline{appr}^{x}_{A'}(X), \overline{appr}^{x}_{A'}(X)).$$

Based on the x-local rough set, the x-local positive region using $A' \subseteq A$ is defined as

$$POS^{x}_{A'}(d) = \bigcup_{d_i \in V_d} \underline{appr}^{x}_{A'}(\{u \in U : d(u) = d_i\}), \tag{1}$$

where V_d is the domain of d.

Thus, x-local rough sets are local descriptions of the target concept. To obtain an overall description of the target concept in the entire study area, all x-local rough sets of target concept X are aggregated to form an overall approximation of target concept X (spatial rough set of target concept X):

$$SRS_{A'}(X) = \{(\underline{appr}^{x}_{A'}(X), \overline{appr}^{x}_{A'}(X)) | x \in U\}.$$

3. Spatial rough set-based geographical detectors

To measure the explanatory power of conditional features, a number of uncertainty measures for rough sets have been developed from various perspectives. However, when applied for spatial analysis, these indices mainly measure the overall explanatory power in the study area. As they ignore the existence of spatial heterogeneity, they are insufficient for analyzing geographical phenomena. Accordingly, it is necessary to migrate existing roughness measures to spatial rough sets to measure the local explanatory power and its heterogeneity.

Compared with other roughness measures, the approximation quality is easy to interpret. Hence, it is one of the most commonly used roughness measures in analyzing spatial data. The approximation quality uses the proportion of objects in the positive region of the universe to measure the explanatory power of conditional features. A larger value of the approximation quality indicates a lower possibility of observing any inconsistency between the condition and the target variable. In the following, this measure is migrated to x-local rough sets.

Definition 1. *For an object $x \in U$, the x-local approximation quality of decision d using feature set $A' \subseteq A$ is*

$$\gamma(x, A', d) = |POS^x_{A'}(d)|/|C(x)|. \tag{2}$$

The x-local approximation quality uses the proportion of elements that are uniquely classified by the target concept, i.e., all elements in the corresponding x-local indiscernible set that have the same target value, to evaluate the quality with which the target concepts are approximated. A larger positive region produces a larger value of $\gamma(x, A', d)$, indicating that the conditional features have greater local explanatory power. When $\gamma(x, A', d) = 1$, all elements in $C(x)$ are uniquely classified to one target concept and d can be precisely approximated using feature set A' in $C(x)$. This means that A' completely explains d in the x-centered region. When $\gamma(x, A', d) = 0$, there are no elements in the positive region. This means that it is not possible to determine the target category using only the conditional features in $C(x)$. Similar to classical rough sets, the x-local approximation quality increases monotonically as features are added to A'.

Property 3.1. *For an $SIS = (U, A, d, M)$, if $B' \subseteq A' \subseteq A$, then $\gamma(x, A', d) \geq \gamma(x, B', d)$.*

Proof. If $B' \subseteq A'$, then $[y]_{A'} \subseteq [y]_{B'} \Rightarrow [y]_{A'} \cap C(x) \subseteq [y]_{B'} \cap C(x) \Rightarrow LInd(y, x, A') \subseteq LInd(y, x, B')$. Suppose that $z \in POS^x_{B'}(d)$ and $z \in LInd(z, x, B')$. If $z \notin POS^x_{A'}(d)$, then $LInd(z, x, A') \not\subseteq \{x \in U : d(x) = d_i\}$ for any d_i. However, there exists some d_i such that $LInd(z, x, A') \subseteq LInd(z, x, B') \subseteq \{x \in U : d(x) = d_i\}$ because $z \in POS^x_{B'}(d)$. This leads to a contradiction. Accordingly, any $z \in POS^x_{B'}(d) \Rightarrow z \in POS^x_{A'}(d)$. Therefore, $POS^x_{B'}(d) \subseteq POS^x_{A'}(d)$ and $\gamma(x, A', d) \geq \gamma(x, B', d)$. □

In addition to a local approximation, it is important to inspect the overall spatial approximations in certain real-life applications. Based on the x-local approximation quality, a new index is proposed for measuring the overall local explanatory power of the conditional features.

Definition 2. *For an $SIS = \{U, A, d, M\}$, $A' \subseteq A$, the approximation quality-based average local explanatory power of A' is*

$$\mathscr{D}(A', d) = \sum_{x \in U} \gamma(x, A', d)/|U| \tag{3}$$

and the spatial entropy of the local explanatory power is

$$SE(A', d) = -\sum_{x \in U} \gamma_N(x, A', d) log_2 \gamma_N(x, A', d), \tag{4}$$

where $\gamma_N(x, A', d) = \gamma(x, A', d)/\sum_{x \in U} \gamma(x, A', d)$.

$\mathscr{D}(A', d)$ is the average of all local approximation measure values over the entire study area. A larger value of $\mathscr{D}(A', d)$ indicates that A' has a larger average local explanatory power on d. $\mathscr{D}(A', d)$ is different from the conventional approximation quality in classical rough sets. The conventional approximation quality is $|POS_{A'}(d)|/|U|$, which uses the proportion of objects in the positive region of the entire study area. Clearly, $\mathscr{D}(A', d)$ takes account of the local approximation quality at every location of the study area, whereas the conventional form only uses the feature space to approximate the target concepts. Similar to x-local measures, $\mathscr{D}(A', d)$ increases monotonically as features are added to A'.

Property 3.2. *For $SIS = (U, A, d, M)$, given $B' \subseteq A' \subseteq A$, $\mathscr{D}(A', d) \geq \mathscr{D}(B', d)$.*

Different from $\mathscr{D}(A', d)$, $SE(A', d)$ can be used to measure the degree of spatial heterogeneity of the approximation of target concepts. A large value of $SE(A', d)$ indicates a small difference in the local explanatory power of A' on the target concepts at different locations. When the approximation measure is equal at all locations, $SE(A', d)$ reaches its maximum of $-log_2(1/|U|)$.

Based on these indices, three new geographical detectors are proposed to perform three different tasks on spatial datasets. The spatial rough set-based factor detector is similar to the factor detector in q-GD and is denoted by SRS_F. It can be used to measure the average local explanatory power and detect the existence of spatial heterogeneity in the local explanatory power of features. The spatial rough set-based ecological detector is similar to the ecological detector in q-GD and is denoted by SRS_E. This provides an efficient means of comparing the average local explanatory power of conditional features. The spatial rough set-based interaction detector is similar to that in q-GD and is denoted by SRS_I. It detects the local explanatory power added by new features.

3.1. SRS_F: Measuring the average local explanatory power and detecting its spatial heterogeneity

Given a conditional feature set A', SRS_F is $\{\mathscr{D}(A', d), SE(A', d)\}$. $\mathscr{D}(A', d)$ is used to measure the conditional features' explanatory power on the target geographical phenomena. Larger values of $\mathscr{D}(A', d)$ indicate that the conditional features provide a better explanation of the target nominal variable. The maximum value of $\mathscr{D}(A', d)$ is one, which means that A' completely explains d. $SE(A', d)$ is used to measure the spatial heterogeneity of the explanatory power on the target nominal variable. Smaller values of SE reflect greater spatial heterogeneity in the spatial dataset. The maximum value of SE is $-log_2(1/|U|)$, which is attained when A' has the same explanatory power at all locations.

When the first component of SRS_F (i.e., $\mathscr{D}(A', d)$) is calculated, it is important to test whether the average local explanatory power is significantly greater than zero. For this task, a small fraction of objects U' is randomly drawn from U and the value of $\gamma(x, A', d)$ is calculated for each object in U'. This constitutes a simple random sample of $\gamma(x, A', d)$. A Student's t-test is then used to test whether the mean of $\gamma(x, A', d)$ ($\mathscr{D}(A', d)$) is significantly greater than zero. If the result is not statistically significant, then A' has no explanatory power on d.

Compared with q-GD, the SRS_F uses two indices: $\mathscr{D}$ and SE. The former denotes the average local explanatory power and the latter reflects the degree of spatial heterogeneity of

the local explanatory power. From $\mathscr{D}$, it is difficult to determine whether the local explanatory power is evenly distributed over space or has large differences between different locations. However, spatial heterogeneity is an important information source for analyzing spatial data [21]. Accordingly, it is important to calculate SE in addition to $\mathscr{D}$. This is an advantage of SRS_F over q-GD, in which there are no tools for detecting spatial heterogeneity.

3.2. SRS_E: Comparing the average local explanatory power of conditional features

Generally, in identifying the cause of the target geographical phenomenon, there are many candidate features. The question is, does one spatial feature play a more important role than other features? This issue can be solved by comparing the average local explanatory powers of different features. Given two feature sets A' and B', SRS_E is $\mathscr{D}(A', d) - \mathscr{D}(B', d)$. If the difference is greater than zero, there are more objects in the lower approximation of the x-local rough sets, and there exist more x-local indiscernible sets that can completely explain the target variable in the x-local rough sets, when A' is used. When $\mathscr{D}(\{a\}, d) = 1$, feature a can completely explain the target variable. To summarize, larger values of $\mathscr{D}(\{a\}, d)$ indicate that feature a has greater importance.

To test whether $\mathscr{D}(A', d) - \mathscr{D}(B', d)$ is statistically significantly different from zero, a fraction U' of U is randomly selected. The local measures for these random samples are calculated using different features, for example, $\gamma(x, A', d)$ and $\gamma(x, B', d)$ for $x \in U'$. There are two sets of local measures for two features, that is, $\{\gamma(x, A', d)|x \in U'\}$ and $\{\gamma(x, B', d)|x \in U'\}$. A Student's t-test is used to test whether there is a statistically significant difference between the mean of these two sets. If the difference is statistically significant, then $\mathscr{D}(A', d)$ is statistically significantly different from $\mathscr{D}(B', d)$.

3.3. SRS_I: Detecting the interactions among conditional features

$\mathscr{D}$ can be used to measure the increase in explanatory power that comes from additional features. Suppose that the original model only uses the feature set A_1. A new feature set A_2 is added to the model to approximate the target geographical phenomenon. SRS_I, that is, $\mathscr{D}(A_1 \cup A_2, d) - \mathscr{D}(A_1, d)$, reflects the explanatory power added by A_2. The difference from q-GD is that new features cannot weaken the explanatory power for the target geographical phenomenon because $\mathscr{D}$ is monotonically increasing in terms of Property 3.2. A Student's t-test can be used to test whether $\mathscr{D}(A_1 \cup A_2, d) - \mathscr{D}(A_1, d)$ is statistically significantly greater than zero using a small random sample from U.

As well as measuring the additional explanatory power, SE can be used to measure whether the spatial heterogeneity of the local explanatory power is weakened or enhanced by the addition of new features. $SE(A_1 \cup A_2, d) > SE(A_1, d)$ indicates that feature set A_2 reduces the spatial heterogeneity of the explanatory power. If $SE(A_1 \cup A_2, d) < SE(A_1, d)$, feature set A_2 increases the spatial heterogeneity of the explanatory power.

All three geographical detectors depend on two indices, $\mathscr{D}$ and SE. These two indices are easy to compute when $|POS^x_{A'}(d)|$ is calculated. The first step in calculating $|POS^x_{A'}(d)|$ is to construct $C(x)$ using x and all its neighbors. Second, all the equivalence classes in $C(x)$ are calculated. Finally, the numbers of elements of the equivalence classes with only one decision

in $C(x)$ are aggregated among all these equivalence classes to calculate $|POS^x_{A'}(d)|$. Because $C(x)$ generally contains a small fraction of U, no heuristic methods are used in calculating $|POS^x_{A'}(d)|$.

4. Empirical study

The three proposed detectors were empirically evaluated using two datasets. Moreover, the factor detector, ecological detector, and interaction detector forms of q-GD were compared with the SRS-GDs on these two datasets. Finally, all SRS-GDs were tested on a dataset with more than two decision values to demonstrate their effectiveness when applied to nominal target variables. Each SRS-GD was implemented using C++, and the experiments were performed on a computer with an Intel®Core™ i7-7200U CPU and 16 GB memory. The operating system was Ubuntu 18.04.

4.1. Experimental data

To validate the proposed method, two publicly accessible datasets were used in the experiments. These two datasets are available from GeoDa [4]. The first dataset consists of Baltimore house sale prices and hedonics[1]; this dataset is referred to as Baltimore for simplicity. Baltimore contains point-pattern spatial data, as shown in Figure 2. There are 211 objects in the dataset. Four attributes of the houses were selected as the conditional features: 'whether it is a detached unit' (DWELL), 'whether it has a patio' (PATIO), 'whether it has a fireplace' (FIREPL), and 'number of stories' (NSTOR). The decision feature was constructed using the 'sale price of the house' (PRICE). If the sale price was more than $40,000, then the decision value was set to one; otherwise, the decision value was set to zero. In the experiment, points less than 1.5 km away from the current point were considered as the current point's neighbors. For simplicity, this distance is referred to as the neighboring distance.

The second dataset is taken from the "2008 Cincinnati Crime + Socio-Demographics" repository[2], and is referred to as Cincinnati for simplicity. This dataset contains spatial data on an irregular lattice, as shown in Figure 3. There are 457 objects in the dataset. The 'male population' (MALE), 'female population' (FEMALE), 'median age' (MEDIAN_AGE), 'average family size' (AVG_FAMSIZ), and 'population density' (DENSITY) were selected as the conditional features. The occurrences of (THEFT_D) were used as the decision feature. In the experiment, objects less than 0.5 km away from the current object were considered as the current object's neighbors.

4.2. Experimental design

Each experiment consisted of two steps. First, each continuous-valued condition feature was discretized into five categories using the equal-width discretization method. This step is necessary for both q-GD and SRS-GD. Equal-width and equal-frequency methods are two simple and easy-to-interpret methods. The equal-width method generally divides the universe

[1] https://geodacenter.github.io/data-and-lab/baltim/
[2] https://geodacenter.github.io/data-and-lab/walnut_hills/

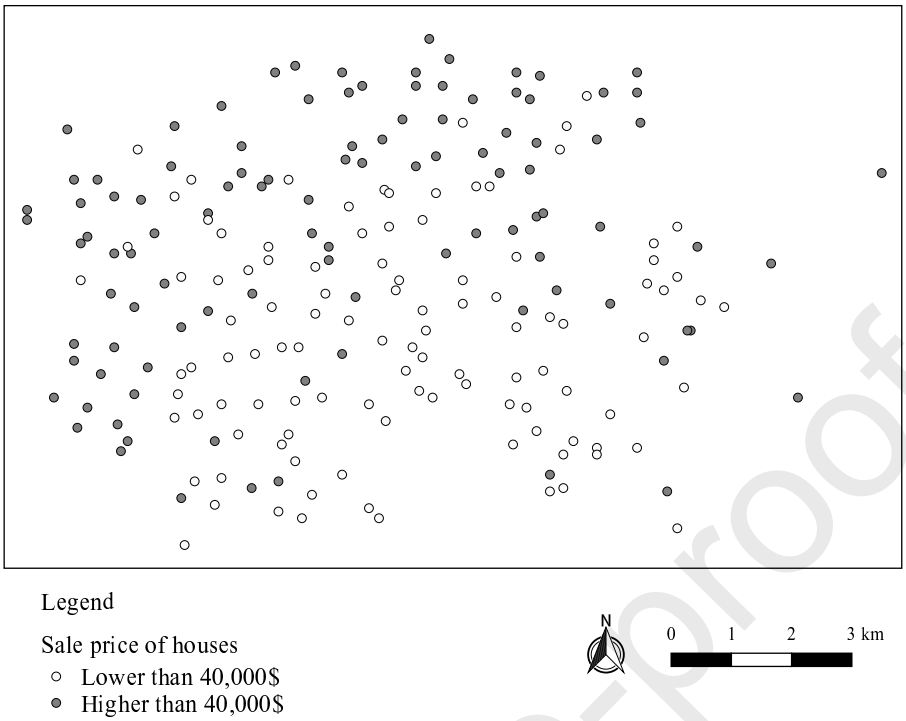


Figure 2: Map of the Baltimore house price data

Figure 3: Map of the occurrences of theft in Cincinnati in 2008

into several subsets of different sizes. This is more similar to the case of categorical features. Therefore, the equal-width method was used.

However, in real-life applications, it is important to select the appropriate discretization method in terms of application requirements. For example, the minimum description length principle [14] is suitable for minimizing the information loss using as few cuts as possible. In some situations, it is important to discretize the continuous features manually according to their physical meanings. A simple example is the gross domestic product. A comparison of different discretization methods for spatial data can be found in [24, 9].

Next, SRS-GD and q-GD were used to analyze the explanatory power of the conditional feature set for each dataset. In all experiments, the significance levels for the Student's t-test and F-test were set to 0.05 (5%).

When SRS-GD was used, all three new factors (SRS_F, SRS_E, and SRS_I) were calculated. SRS_F was applied to each feature in addition to all features, that is, $\mathscr{D}(A, d)$, $SE(A, d)$, $\mathscr{D}(\{a_i\}, d), SE(\{a_i\}, d), a \in A$ were calculated to evaluate SRS_F. Second, SRS_E was used to detect which conditional feature best explains the target nominal variable. Finally, $\mathscr{D}(\{a_1, a_{i\neq 1}\}, d) - \mathscr{D}(\{a_1\}, d)$ was calculated to inspect the additional explanatory power produced by adding feature a_i to feature a_1, that is, the interaction among features (SRS_I). In the Baltimore dataset, DWELL was selected as a_1. In the Cincinnati dataset, MALE was selected as a_1. $SE(\{a_1, a_{i\neq 1}\}, d)$ was calculated to measure whether the spatial heterogeneity of the local explanatory power was weakened or enhanced. In the experiment, the first conditional feature was selected as a_1 to demonstrate the effectiveness of SRS_I. Regardless of which conditional feature was selected as a_1, the influence of adding new conditional features using SRS_I could always be detected.

To compare SRS-GD with q-GD, the binary nominal target variables were used as the continuous target variables in performing q-GD analysis. The strata formed by all conditional features, that is, intersections of strata formed by each feature, were used to calculate the q-statistics for all features, denoted by q_A, to measure the associated explanatory power. The factor detector and ecological detector forms of q-GD were applied to compare the explanatory power of different conditional features on the target variable. Moreover, the interaction detector forms of q-GD were calculated to inspect whether the introduction of feature a_i enhanced or weakened the explanatory power of feature a_1.

4.3. Results and discussion

4.3.1. Measuring the local explanatory power and its spatial heterogeneity

The second row of Table 1 contains the q-statistics of each individual conditional feature and all conditional features for the Baltimore dataset. The q-statistics (factor detector in q-GD) show that each conditional feature and all the conditional features can explain the target variable to some extent. The F-tests for all q-statistics indicate statistical significance. This means that the stratified population variance based on the strata formed by features is statistically significantly different from the population variance. In some cases, the q-statistics failed to pass the F-test. For example, there are three q-statistics in the second row of Table 2

(which presents the q-statistics for the Cincinnati dataset) that failed to pass the F-test. Non-significant q-statistics suggest that the explanatory power of the feature set is so weak that the stratified population variance is almost the same as the population variance.

A'	DWELL	PATIO	FIREPL	NSTOR	ALL FEATURES
q-statistic	0.2693	0.0824	0.1233	0.1043	0.3859
$\mathscr{D}(A', d)$	0.1861	0.1374	0.1061	0.2764	0.6121
$SE(A', d)$	5.996	6.511	5.677	6.939	7.589
q-cont	0.2769	0.2064	0.2760	0.1770	0.6174

Table 1: $\mathscr{D}$, SE, and q-statistics of conditional features in the Baltimore dataset. The fifth row contains the q-statistics for the original continuous-valued target variable.

A'	MALE	FEMALE	MEDIAN _AGE	AVG_ FAMSIZ	DENSITY	ALL FEATURES
q-statistic	0.0370^{+}	0.0231^{+}	0.0624	0.0831	0.0114^{+}	0.1975
$\mathscr{D}(A', d)$	0.0645	0.0424	0.0402	0.0426	0.0586	0.3781
$SE(A', d)$	8.274	8.188	7.298	7.204	7.874	8.735
q-cont	0.01312	0.0696^{+}	0.0325	0.0365	0.0080^{+}	0.3470^{+}

$^{+}$ failed to pass the F-test.

Table 2: $\mathscr{D}$, SE, and q-statistics of conditional features in the Cincinnati dataset. The fifth row contains the q-statistics for the original continuous-valued target variable.

Additionally, the q-statistics for the original continuous-valued target variable, which is denoted by q-cont for simplicity, were also calculated. The fifth row of Tables 1 and 2 contains the q-cont values of the Baltimore and Cincinnati datasets, respectively. Clearly, the q-statistics and q-cont are not consistent with each other. For example, the q-cont for all features in the Cincinnati dataset is no longer statistically significant. Simultaneously, some of the original conditional features with small or insignificant explanatory power have a large significant q-cont. In fact, the q-statistics measure the explanatory power for the occurrence of target geographical phenomena, whereas q-cont measures the explanatory power of the degree of the target geographical phenomena. Therefore, they have different values that cannot be used in place of one another.

Unlike the factor detector in q-GD, SRS_F has two components, $\mathscr{D}(A', d)$ and $SE(A', d)$. The $SE(A', d)$ component reflects the degree of spatial heterogeneity of the local explanatory power. The fourth row in Tables 1 and 2 contains the SE values for each feature and all features. Larger values of SE indicate a smaller difference between the local explanatory power among all x-centered regions.

The other component, $\mathscr{D}(A', d)$, describes the explanatory power using the average proportion of objects in the x-local indiscernible set that can completely explain d using A' in each x-centered region, that is, $\gamma(x, A', d)$. The third row in Tables 1 and 2 contains $\mathscr{D}$ for each feature and all features for the Baltimore and Cincinnati datasets. The Student's t-tests for all $\mathscr{D}$ indicate that they are statistically significantly greater than zero. This means that each individual feature and all the features together can explain the target nominal variable to some extent.

Because $\mathscr{D}$ and the q-statistics measure the explanatory power from two different perspectives, conditional features with no statistically significant q-statistic may have a statistically significant $\mathscr{D}$. For example, the non-significant q-statistics for MALE, FEMALE, and DENSITY indicate that these three features have no explanatory power for the occurrence of thefts in the strata. However, as the $\mathscr{D}$ values for these three features are statistically significant (according to the Student's t-test), each feature's average local explanatory power is statistically significantly greater than zero and can partly explain the target nominal variable.

Although q_A gives the explanatory power of A on d, it cannot replace $\mathscr{D}(A, d)$. Indeed, q_A mainly takes the strata into account, and ignores the local explanatory power of conditional features. In some situations, it is difficult to determine whether there is any spatial heterogeneity in the local explanatory power of conditional features using only q-GD.

For example, given a spatially randomly distributed conditional feature a, if the q-statistic of a is 0.9, it is reasonable to conclude that the target variable depends on a to a great extent, but the distribution of the local explanatory power of a on the target variable might be spatially randomly distributed. Taking the Baltimore dataset as an example, a spatially reshuffled[3] dataset was used to calculate the q-statistics and $\mathscr{D}(A, d)$. In this case, q_A was unchanged and passed the F-test. This means that q-GD does not reflect changes in the distribution of objects. However, $\mathscr{D}(A, d)$ decreased to 0.54, which shows that $\mathscr{D}$ is sensitive to the objects' spatial distribution.

Another important reason for using $\mathscr{D}(A, d)$ is that conditional features may only be consistent with the target variable in different local regions. For example, assume that the objects in an equivalence class $[x]_A$ are evenly distributed in two local regions, 'a' and 'b', in the study area. In region 'a', all objects in $[x]_A$ have decision '1', whereas all objects in region 'b' have decision '0'. Clearly, feature set A can explain the target variable in both local regions, but is inconsistent with the target variable over the entire study area. When q-statistics are used in such a situation, it is clear that the stratum corresponding to $[x]_A$ cannot explain the target variable. Despite this, $\mathscr{D}(A, d)$ may be significantly greater than zero because it inspects the local explanatory power in each x-centered region.

4.3.2. Comparison of the explanatory power of different conditional features

Any two features may have a different explanatory power on the target variable. Taking DWELL and NSTOR in the Baltimore dataset as an example, the q-statistic for DWELL is greater than that for NSTOR, which indicates that DWELL has greater explanatory power than NSTOR from the perspective of the spatial strata. Rows 2–4 in Tables 3 and 4 demonstrate the significance of the difference between conditional features for the Baltimore and Cincinnati datasets using the ecology detector form of q-GD.

For the Baltimore dataset, the ecological factor form of q-GD shows that the explanatory power of conditional features runs in the following order: DWELL > FIREPL = NSTORE > PATIO. Although the q-statistic of FIREPL is greater than that of NSTOR, the ecological factor indicates that the difference in explanatory power between FIREPL and NSTOR is

[3]This means that the variables' values randomly change the polygon or point to which they are attached.

		PATIO	FIREPL	NSTOR
q-GD	DWELL	TRUE	TRUE	TRUE
	PATIO		TRUE	TRUE
	FIREPL			FALSE
SRS-GD	DWELL	TRUE	TRUE	TRUE
	PATIO		FALSE	TRUE
	FIREPL			TRUE

Table 3: Significance of the difference between conditional features for the Baltimore dataset using SRS-GD and q-GD.

		FEMALE	MEDIAN_AGE	AVG_FAMSIZ	DENSITY
q-GD	MALE	NA[#]	TRUE	TRUE	NA[#]
	FEMALE		TRUE	TRUE	NA[#]
	MEDIAN_AGE			TRUE	TRUE
	AVG_FAMSIZ				TRUE
SRS-GD	MALE	TRUE	FALSE	TRUE	FALSE
	FEMALE		FALSE	FALSE	TRUE
	MEDIAN_AGE			TRUE	FALSE
	AVG_FAMSIZ				TRUE

[#] Because the q-statistics of MALE, FEMALE, and DENSITY failed to pass the F-test, the explanatory power of these three features is very weak. It is not necessary to compare their explanatory power any more.

Table 4: Significance of the difference between conditional features for the Cincinnati dataset using SRS-GD and q-GD.

not statistically significant. For the Cincinnati dataset, because the q-statistics of MALE, FEMALE, and DENSITY failed to pass the F-test, the explanatory power of these three features is very weak. Thus, it is no longer meaningful to compare their explanatory power. All other feature pairs have significant differences in their explanatory power. The order of the explanatory power is AVG_FAMSIZE > MEDIAN_AGE > MALE = FEMALE = DENSITY.

Unlike the q-statistics, SRS_E compares the explanatory power between features from the perspective of the local approximation quality. Rows 5–7 in Tables 3 and 4 demonstrate the significance of the difference between conditional features using SRS_E for the two datasets. For the Baltimore dataset, SRS_E gives the following order for the explanatory power of conditional features: NSTORE > DWELL > PATIO = FIREPL. PATIO = FIREPL because SRS_F shows that the difference in the explanatory power between FIREPL and PATIO is not statistically significant.

The situation in the Cincinnati dataset is somewhat complicated. According to SRS_E, when MEDIAN_AGE is excluded, it is clear that MALE=DENSITY>AVG_FAMSIZE=FEMALE. The conditional feature MEDIAN_AGE has no statistically significant difference in explanatory power compared with all other conditional features, except that FAMSIZE>MEDIAN_AGE is statistically significant.

Clearly, the relation between features may be different when using the factor detector form of q-GD and SRS_E. For example, the MALE feature in the Cincinnati dataset has the largest

explanatory power according to SRS-GD, whereas it has no explanatory power in terms of q-GD. The reason is that q-GD and SRS-GD inspect the explanatory power using the entire study area and local areas, respectively. Although MALE cannot explain the target nominal variable because there is no significant difference between the stratified population variance and the population variance in the entire study area, it may explain the target random variable to some extent in most x-centered regions. For the same reason, the order of the explanatory power of different features varies between SRS-GD and q-GD.

Furthermore, SRS-GD can be used as a complement to q-GD. From the local perspective of the explanatory power, SRS-GD identifies certain relations between features that are missed by q-GD. For example, the relation between MALE and FEMALE is statistically significant, although q-GD failed to detect this relation from the spatial strata perspective. Additionally, $SE(A', d)$ demonstrates the spatial heterogeneity of the local explanatory power of different features.

4.3.3. Detecting feature interactions

Tables 5 and 6 summarize $\mathscr{D}$, SE, and the q-statistics for the feature sets {DWELL, PATIO}, {DWELL, FIREPL}, and {DWELL, NSTOR} (Baltimore dataset) and {MALE, FEMALE}, {MALE, MEDIAN_AGE}, {MALE, AVG_FAMSIZ}, and {MALE, DENSITY} (Cincinnati dataset). Compared with single features, the average local explanatory power increases when another feature is introduced, as proven by Property 3.2. The Student's t-test also suggests that all $\mathscr{D}(\{a_1, a_{i\neq 1}\}, d) - \mathscr{D}(\{a_1\}, d)$ are statistically significantly greater than zero. This indicates that adding new features significantly increases the average explanatory power.

	DWELL $\wedge$ PATIO	DWELL $\wedge$ FIREPL	DWELL $\wedge$ NSTOR
$\mathscr{D}(\{a_i, a_j\}, d)$	0.2705	0.3009	0.4658
$SE(\{a_i, a_j\}, d)$	6.997	7.062	7.361
q-statistics	0.3048	0.3271	0.3163

Table 5: $\mathscr{D}$, SE, and q-statistics of the Baltimore dataset for the feature sets {DWELL,PATIO}, {DWELL,FIREPL}, and {DWELL,NSTOR}.

	MALE $\wedge$ FEMALE	MALE $\wedge$ MEDIAN_AGE	MALE $\wedge$ AVG_FAMSIZ	MALE $\wedge$ DENSITY
$\mathscr{D}(\{a_i, a_j\}, d)$	0.0757	0.1087*	0.1084	0.1261
$SE(\{a_i, a_j\}, d)$	8.315	8.370	8.188	8.354
q-statistics	0.0468^{+}	0.0984^{+}	0.1095	0.0537^{+}

$^{+}$ failed to pass the F-test.

Table 6: $\mathscr{D}$, SE, and q-statistics of the Cincinnati dataset for the feature sets {MALE,FEMALE}, {MALE,MEDIAN_AGE}, {MALE,AVG_FAMSIZ}, and {MALE,DENSITY}.

In most cases, SE increased when new features were added. This indicates that the introduction of new features weakens the spatial heterogeneity of the local explanatory power. An exception is that $SE(\{\text{MALE, AVG\_FAMSIZ}\}, d) > SE(\{\text{MALE}\}, d)$. According to $SE(\{$

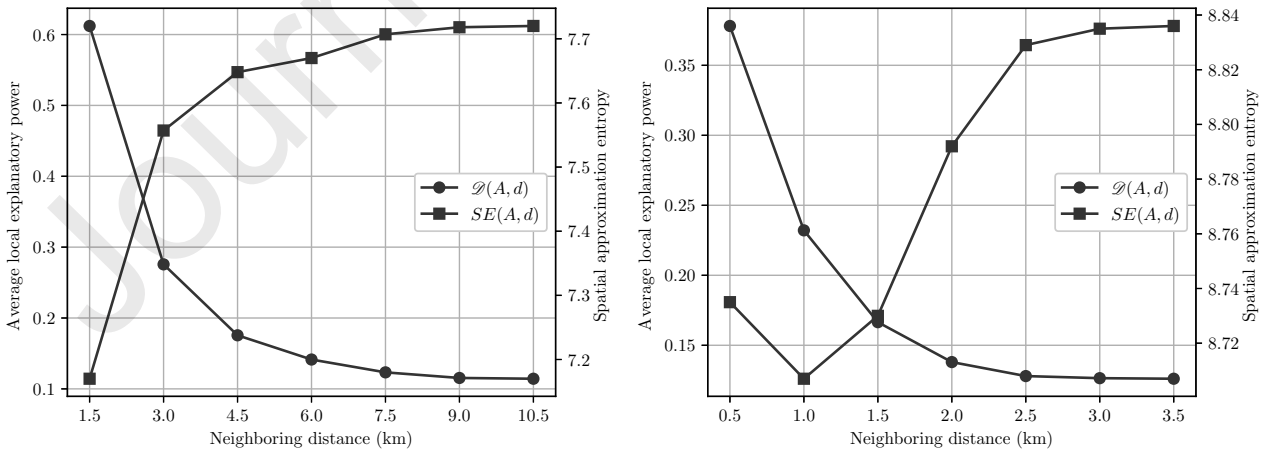
AVG_FAMSIZ}) < $SE(\{\text{MALE}\}, d)$, the degree of spatial heterogeneity of the local explanatory power of AVG_FAMSIZ is less than that of MALE. Thus, introducing AVG_FAMSIZ to {MALE} may increase the spatial heterogeneity of the local explanatory power.

Although the interaction detector form of q-GD detected the interaction between features in terms of the strata formed by different feature sets, it was consistent with SRS_I in both datasets. The q-statistics also increased when adding new features. However, unlike SRS_I, which increased monotonously, the q-statistics sometimes decreased when new features were added. An example is given in [37].

4.4. Neighboring distance and local explanatory power

The neighboring distance is an important factor for SRS-GD. If the neighboring distance is too small, most objects will not have any neighbors, which results in a meaningless SRS-GD. However, if the neighboring distance is too large, most objects will be considered as neighbors of the current object. The local explanatory power will then be concealed by the global information, preventing SRS-GD from measuring the local explanatory power and its spatial heterogeneity.

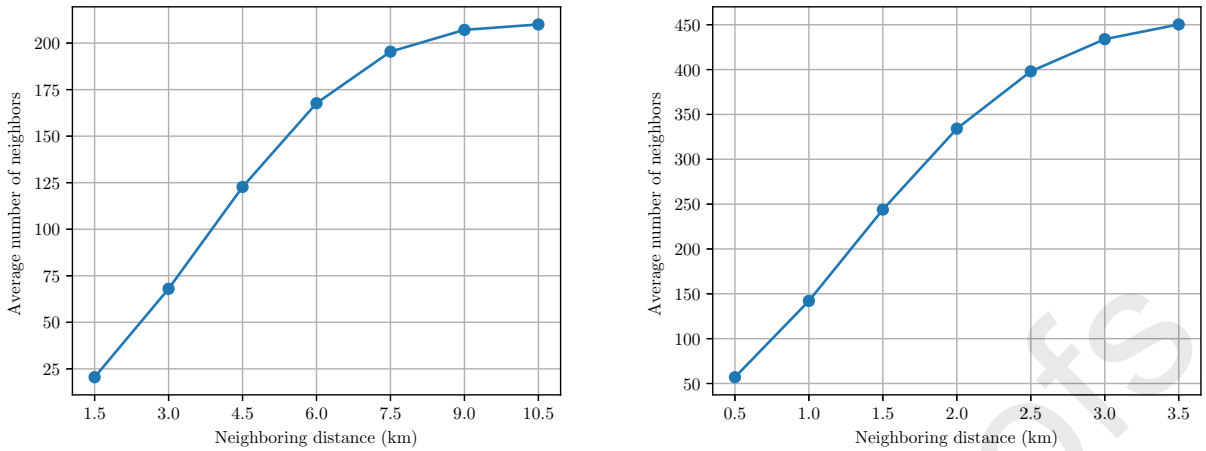
Figure 4 shows the variation of the average local explanatory power of all conditional features and its spatial heterogeneity. The neighboring distances were varied from 1.5 km to 10.5 km for the Baltimore dataset and from 0.5 km to 3.5 km for the Cincinnati dataset. Although $SE(A, d)$ is greater for 0.5 km than for 1.0 km in the Cincinnati dataset, a larger distance generally produces a larger value of $SE(A, d)$ and lower spatial heterogeneity. The reason is that, when the neighboring distance is too large, $|C(x)|$ approaches $|U|$, as shown in Figure 5. Then, $\gamma(x, a_1, d)$ approaches the approximation quality of the classical rough sets for each object, and $SE(A, d)$ approaches its maximum as the neighboring distance increases.



(a) Baltimore (b) Cincinnati

Figure 4: Variation in the average local explanatory power of all conditional features and its spatial heterogeneity using different neighboring distances.

For $\mathscr{D}(A, d)$, the trend is different. This metric decreases as the neighboring distance increases. The reason is that the local indiscernible sets that are originally in local positive regions



(a) Baltimore (b) Cincinnati

Figure 5: Average number of neighbors for different neighboring distances.

might have more than one decision as the neighboring distance increases. Accordingly, many objects that are originally in the positive region move to the boundary region. However, as $|C(x)|$ approaches $|U|$, $\gamma(x, a_1, d)$ becomes the same as the approximation quality of the classical rough sets. $\mathscr{D}(A, d)$ also approaches the minimum as the neighboring distance increases.

From the rough set theory perspective, an increase in the neighboring distance enhances the average number of neighbors and eventually causes $\mathscr{D}$ and SE to approach their limits. However, distance (that is, spatial lag) is more commonly used in analyzing spatial heterogeneity because this is a natural and essential metric in depicting and explaining geographical phenomena [11]. Finally, the same geographical phenomena can be approximated using points or polygons of different densities in different applications, and some geographical phenomena may be represented using different densities of points or polygons in different areas, which is also the situation in the two datasets used in this study [10, 19, 25, 18]. Accordingly, if the number of neighbors is fixed for each object, the proposed SRS-GD may not correctly reflect the target geographical phenomena.

In fact, $\mathscr{D}(A, d)$ and $SE(A, d)$ demonstrate the spatial explanatory power on the target variable at different spatial scales under different neighboring distances. Thus, they provide a multi-scale view of the explanatory power of conditional features. The optimal neighboring distance should be determined by balancing the explanatory power and its spatial heterogeneity according to the application. We will address this issue in future work.

4.5. Multiple-category experiment

To show the effectiveness of SRS-GD on nominal target variables, an experiment was performed on the Cincinnati dataset. First, the target variable was replaced by the major crime type. The equation $t = argmax\{\ [NR_{BURGLARY},\ NR_{ASSAULT}, 0.2 \times NR_{THEFT}]\}$[4] was used

[4]Because there are far more thefts than burglaries or assaults in most places on the map, the number of thefts was assigned a small weight to prevent the other two crimes being concealed.

to determine the major type of crime, where NR denotes the number of instances and BURGLARY, ASSAULT, and THEFT are the three types of crime: $t = 1, 2, 3$ indicates that BURGLARY, ASSAULT, and THEFT are the major type of crime, respectively. NONE indicates that no crimes were reported. Figure 6 shows the distribution of the four categories in the Cincinnati dataset.

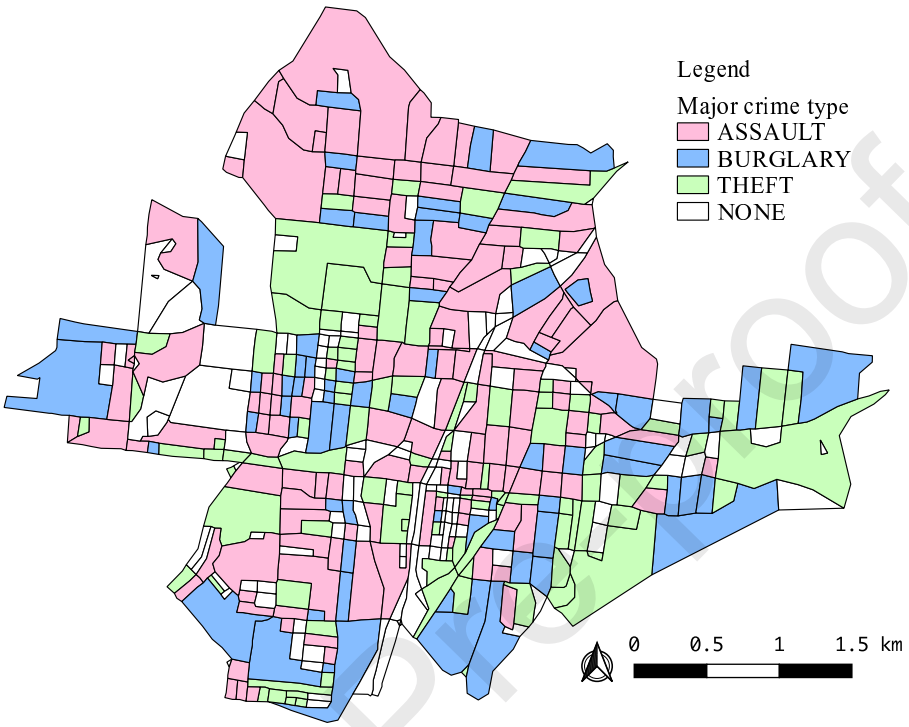


Figure 6: Map of the major crime types in Cincinnati in 2008.

Table 7 shows the SRS_F value for each feature and all features. According to the Student's t-test, all $\mathscr{D}$ are statistically significantly greater than zero. Accordingly, each feature and all features can explain the major crime type to some extent. Table 8 presents the results for whether the relation between features is statistically significant (SRS_E). Clearly, only DENSITY and AVG_FAMSIZ have significant differences in explanatory power. This means that almost all features have similar explanatory power when used individually.

A'	MALE	FEMALE	MEDIAN_AGE	AVG_FAMSIZ	DENSITY	ALL FEATURES
$\mathscr{D}(\{a_i\}, d)$	0.0221	0.0301	0.0210	0.0264	0.0276	0.2769
$SE(\{a_i\}, d)$	7.585	7.627	7.242	7.050	7.685	8.7

Table 7: $\mathscr{D}$ and SE of each conditional feature for the multiple-category Cincinnati dataset.

Similar to the two-category cases, SRS-GD was also effective in detecting feature interactions (SRS_I). Table 9 presents the $\mathscr{D}$ and SE values for the multiple-category Cincinnati dataset using the feature sets {MALE,FEMALE}, {MALE,MEDIAN_AGE}, {MALE,AVG_FAMSIZ}, and {MALE,DENSITY}. Compared with the single-feature $MALE$, the average local explanatory power and SE both increase when another feature is introduced. Moreover, the Student's

	FEMALE	MEDIAN_AGE	AVG_FAMSIZ	DENSITY
MALE	FALSE	FALSE	FALSE	FALSE
FEMALE		FALSE	FALSE	FALSE
MEDIAN_AGE			FALSE	FALSE
AVG_FAMSIZ				TRUE

Table 8: Significance of the relation between conditional features for the multiple-category Cincinnati dataset using SRS-GD.

t-test shows a significant increment in the explanatory power when adding another feature. Accordingly, introducing new features increases the average local explanatory power and decreases the degree of spatial heterogeneity.

	MALE $\wedge$ FEMALE	MALE $\wedge$ MEDIAN_AGE	MALE $\wedge$ AVG_FAMSIZ	MALE $\wedge$ DENSITY
$\mathscr{D}(\{a_i, a_j\}, d)$	0.0530	0.0653	0.0638	0.0703
$SE(\{a_i, a_j\}, d)$	8.109	8.102	7.972	8.181

Table 9: $\mathscr{D}$ and SE of the multiple-category Cincinnati dataset for the feature sets {MALE,FEMALE}, {MALE,MEDIAN_AGE}, {MALE,AVG_FAMSIZ}, and {MALE,DENSITY}.

5. Conclusion

The explanatory power or the power of determinant [37] is an effective tool for selecting important features that most affect the target variable. In the case of q-GD, the difference between the population variance and strata population variance is used to measure the explanatory power, whereas classical rough set theory and its non-spatial extensions use roughness measures. These methods only inspect the explanatory power in the entire study area. Spatial rough sets take into account spatial heterogeneity, quantifying the roughness at different locations using neighboring objects. This extension makes it possible to inspect the explanatory power from a local perspective.

Based on two measures, $\mathscr{D}$ and SE, three spatial rough set-based geographical detectors have been proposed for measuring the average local explanatory power, comparing the explanatory power between features, and detecting the interactions among conditional features. SRS-GD is an important complement to the widely used q-GD, as it inspects the explanatory power from a local perspective. Experiments showed that the three new geographical detectors are effective in processing datasets that have nominal target variables, and can identify some relations between features that are ignored by q-GD.

Although the proposed SRS-GD model effectively explores and compares the local explanatory power of conditional features, it may require further reinforcement before being applied in other situations. In particular, future research should focus on the following four aspects:

(1) Using other roughness measures to explore the local explanatory power: The approximation quality only describes the proportion of objects in the positive region and ignores the inner structure of the positive region. This is insufficient in some situations [12]. Accordingly,

roughness measures such as the information entropy or combination entropy could be migrated to spatial rough sets to measure the local explanatory power in such situations.

(2) Exploring the influence of discretization methods: Different discretization methods convert the continuous conditional features to different categorical features. This in turn influences the results given by SRS-GD. In future work, it is important to compare different discretization methods thoroughly, as this will guide users to select appropriate discretization methods in terms of applications.

(3) Extending SRS-GD for both nominal and continuous-valued target variables: The current SRS-GD is only effective for nominal target variables. A possible solution is to extend the concept of fuzzy rough sets to spatial data, because fuzzy rough sets handle mixed data very effectively.

(4) Using non-spatial extensions of rough sets to compare the global explanatory power of conditional features: It is natural to use uncertainty measures in non-spatial rough set extensions to measure the explanatory power of conditional features. These measures inspect the explanatory power from different perspectives and provide more comprehensive comparisons. However, these extensions have not provided tools for determining whether the difference between the explanatory power of different conditional feature sets is statistically significant and whether the explanatory power of a feature set is significantly greater than zero. Accordingly, it is important to inspect the statistical characteristics of these measures in future work.

Acknowledgments

The work is supported by the National Key Research and Development Program of China (Grant No. 2017YFB0503501), the National Natural Science Foundation of China (Grant Nos. 41871286, 62072294), and the National Natural Science Foundation for Distinguished Young Scholars of China (Grant No. 41725006).

References

[1] Ahlqvist, O., 2005. Using uncertain conceptual spaces to translate between land cover categories. International Journal of Geographical Information Science 19, 831–857.

[2] Ahlqvist, O., Keukelaar, J., Oukbir, K., 2000. Rough classification and accuracy assessment. International Journal of Geographical Information Science 14, 475–496.

[3] Ahlqvist, O., Keukelaar, J., Oukbir, K., 2003. Rough and fuzzy geographical data integration. International Journal of Geographical Information Science 17, 223–234.

[4] Anselin, L., Syabri, I., Kho, Y., 2006. Geoda: an introduction to spatial data analysis. Geographical analysis 38, 5–22.

[5] Bai, H., Ge, Y., Wang, J., Li, D., Liao, Y., Zheng, X., 2014. A method for extracting rules from spatial data based on rough fuzzy sets. Knowledge-Based Systems 57, 28–40. doi:10.1016/j.knosys.2013.12.008.

[6] Bai, H., Li, D., Ge, Y., Wang, J., 2019. A spatial heterogeneity-based rough set extension for spatial data. International Journal of Geographical Information Science 33, 240–268. doi:10.1080/13658816.2018.1524148.

[7] Banu, S.K., Tripathy, B.K., 2018. Neighborhood-rough-sets based spatial data analytics, in: Mehdi Khosrow-Pour, D. (Ed.), Encyclopedia of Information Science and Technology, Fourth Edition. IGI Global, pp. 1835–1844. DOI: 10.4018/978-1-5225-2255-3.ch160.

[8] Bello, M., Nápoles, G., Vanhoof, K., Bello, R., 2021. Data quality measures based on granular computing for multi-label classification. Information Sciences 560, 51–67. URL: https://www.sciencedirect.com/science/article/pii/S0020025521000542, doi:https://doi.org/10.1016/j.ins.2021.01.027.

[9] Cao, F., Ge, Y., Wang, J.F., 2013. Optimal discretization for geographical detectors-based risk assessment. GIScience & Remote Sensing 50, 78–92. doi:10.1080/15481603.2013.778562.

[10] Cliff, A.D., Ord, J.K., 1981. Spatial processes: models and applications. Pion Ltd, London, United Kindom.

[11] Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.J., Jakomulska, A., Miriti, M., Rosenberg, M.S., 2002. A balanced view of scale in spatial statistical analysis. Ecography 25, 626–640. URL: http://dx.doi.org/10.1034/j.1600-0587.2002.250510.x, doi:10.1034/j.1600-0587.2002.250510.x.

[12] Düntsch, I., Gediga, G., 1998. Uncertainty measures of rough set prediction. Artificial Intelligence 106, 109–137. doi:10.1016/S0004-3702(98)00091-5.

[13] Dutilleul, P., Legendre, P., 1993. Spatial heterogeneity against heteroscedasticity: An ecological paradigm versus a statistical concept. Oikos 66, 152–171.

[14] Fayyad, U.M., Irani, K.B., 1993. Multi-interval discretization of continuous-valued attributes for classification learning., in: Bajcsy, R. (Ed.), IJCAI:International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Francisco, California, USA. pp. 1022–1029.

[15] Fiedukowicz, A., 2015. Fuzzy rough sets theory reducts for quantitative decisions – approach for spatial data generalization, in: Kryszkiewicz, M., Bandyopadhyay, S., Rybinski, H., Pal, S.K. (Eds.), Pattern Recognition and Machine Intelligence, Springer International Publishing, Cham. pp. 314–324. doi:10.1007/978-3-319-19941-2_30.

[16] Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environment and Planning A: Economy and Space 30, 1905–1927. doi:10.1068/a301905.

[17] Ge, Y., Bai, H., Wang, J., Cao, F., 2012. Assessing the quality of training data in the supervised classification of remotely sensed imagery: a correlation analysis. Journal of Spatial Science 57, 135–152. doi:10.1080/14498596.2012.733616.

[18] Ge, Y., Jin, Y., Stein, A., Chen, Y., Wang, J., Wang, J., Cheng, Q., Bai, H., Liu, M., Atkinson, P.M., 2019. Principles and methods of scaling geospatial earth science data. Earth-Science Reviews 197, 102897. URL: http://www.sciencedirect.com/science/article/pii/S0012825219301539, doi:https://doi.org/10.1016/j.earscirev.2019.102897.

[19] Haining, R.P., 1990. Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge, United Kingdom.

[20] Kelam, H., Venkatesan, M., 2019. Optimal band selection using generalized covering-based rough sets on hyperspectral remote sensing big data, in: Peter, J.D., Alavi, A.H., Javadi, B. (Eds.), Advances in Big Data and Cloud Computing, Springer Singapore, Singapore. pp. 263–273.

[21] Li, H., Reynolds, J.F., 1995. On definition and quantification of heterogeneity. Oikos 73, 280–284.

[22] Li, W., Li, J., Huang, J., Dai, W., Zhang, X., 2021. A new rough set model based on multi-scale covering. International Journal of Machine Learning and Cybernetics 12, 243–256.

[23] Liang, J., Chin, K., Dang, C., Yam, R.C., 2002. A new method for measuring uncertainty and fuzziness in rough set theory. International Journal of General Systems 31, 331–342.

[24] Liu, H., Hussain, F., Tan, C.L., Dash, M., 2002. Discretization: An enabling technique. Data Mining and Knowledge Discovery 6, 393–423. doi:10.1023/A:1016304305535.

[25] Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 2015. Geographic Information Science and Systems. 4th ed., Wiley Publishing.

[26] Losada, N., Alén, E., Cotos-Yáñez, T.R., Domínguez, T., 2019. Spatial heterogeneity in Spain for senior travel behavior. Tourism Management 70, 444–452. doi:10.1016/j.tourman.2018.09.011.

[27] Luo, L., Mei, K., Qu, L., Zhang, C., Chen, H., Wang, S., Di, D., Huang, H., Wang, Z., Xia, F., Dahlgren, R.A., Zhang, M., 2019. Assessment of the geographical detector method for investigating heavy metal source apportionment in an urban watershed of eastern china. Science of The Total Environment 653, 714 – 722. doi:https://doi.org/10.1016/j.scitotenv.2018.10.424.

[28] Niu, J., Chen, D., Li, J., Wang, H., 2022. A dynamic rule-based classification model via granular computing. Information Sciences 584, 325–341. URL: https:

//www.sciencedirect.com/science/article/pii/S0020025521010872, doi:https://doi.org/10.1016/j.ins.2021.10.065.

[29] Pan, X., Zhang, S., Zhang, H., Na, X., Li, X., 2010. A variable precision rough set approach to the remote sensing land use/cover classification. Computers & Geosciences 36, 1466–1473. doi:10.1016/j.cageo.2009.11.010.

[30] Pang, B., Mi, J.S., Xiu, Z.Y., 2019. L-fuzzifying approximation operators in fuzzy rough sets. Information Sciences 480, 14–33. URL: https://www.sciencedirect.com/science/article/pii/S0020025518309678, doi:https://doi.org/10.1016/j.ins.2018.12.021.

[31] Pawlak, Z., 1982. Rough sets. International Journal of Computer & Information Sciences 11, 341–356. doi:10.1007/BF01001956.

[32] Qian, Y., Liang, J., 2008. Combination entropy and combination granulation in rough set theory. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 16, 179–193. doi:10.1142/S0218488508005121.

[33] Sharmila Banu, K., Tripathy, B.K., 2016. Rough Set Based Similarity Measures for Data Analytics in Spatial Epidemiology. International Journal of Rough Sets and Data Analysis (IJRSDA) 3, 114–123. doi:10.4018/IJRSDA.2016010107.

[34] Shaver, G., 2005. Spatial heterogeneity: Past, present, and future, in: Lovett, G., Turner, M., Jones, C., Weathers, K. (Eds.), Ecosystem Function in Heterogeneous Landscapes, Springer, New York, NY, USA. pp. 443–449.

[35] Shu, H., Pei, T., Song, C., Ma, T., Du, Y., Fan, Z., Guo, S., 2019. Quantifying the spatial heterogeneity of points. International Journal of Geographical Information Science 33, 1355–1376. doi:10.1080/13658816.2019.1577432.

[36] Sun, D., Shi, S., Wen, H., Xu, J., Zhou, X., Wu, J., 2021. A hybrid optimization method of factor screening predicated on geodetector and random forest for landslide susceptibility mapping. Geomorphology 379, 107623. URL: https://www.sciencedirect.com/science/article/pii/S0169555X21000313, doi:https://doi.org/10.1016/j.geomorph.2021.107623.

[37] Wang, J., Li, X., Christakos, G., Liao, Y., Zhang, T., Gu, X., Zheng, X., 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, 107–127. doi:10.1080/13658810802443457.

[38] Wang, J.F., Zhang, T.L., Fu, B.J., 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 67, 250–256. doi:10.1016/j.ecolind.2016.02.052.

[39] Wei, W., Guo, Z., Shi, P., Zhou, L., Wang, X., Li, Z., Pang, S., Xie, B., 2021. Spatiotemporal changes of land desertification sensitivity in northwest china from 2000 to 2017. Journal of Geographical Sciences 31, 46–68.

[40] Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems 7, 161–187. doi:10.1007/s10109-005-0155-6.

[41] Wu, W.Z., Leung, Y., 2011. Theory and applications of granular labelled partitions in multi-scale decision tables. Information Sciences 181, 3878–3897. doi:10.1016/j.ins.2011.04.047.

[42] Yan, C., Yang, L., Gartner, G., Zhu, Q., Liu, X., 2019. Intelligent initial map scale generation based on rough-set rules. Arabian Journal of Geosciences 12, 109.

[43] Yan, H.Y., Zhang, X.R., Dong, J.H., Shang, M.S., Shan, K., Wu, D., Yuan, Y., Wang, X., Meng, H., Huang, Y., Wang, G.Y., 2016. Spatial and temporal relation rule acquisition of eutrophication in Da'ning River based on rough set theory. Ecological Indicators 66, 180–189. doi:10.1016/j.ecolind.2016.01.032.

[44] Yun, O., Ma, J., 2006. Land cover classification based on tolerant rough set. International Journal of Remote Sensing 27, 3041–3047.

[45] Zhan, J., Jiang, H., Yao, Y., 2020. Covering-based variable precision fuzzy rough sets with promethee-edas methods. Information Sciences 538, 314–336. URL: https://www.sciencedirect.com/science/article/pii/S0020025520305752, doi:https://doi.org/10.1016/j.ins.2020.06.006.

[46] Zhang, J., Li, T., Chen, H., 2014. Composite rough sets for dynamic data mining. Information Sciences 257, 81–100. doi:10.1016/j.ins.2013.08.016.

[47] Zhang, J., Zhu, Y., Pan, Y., Li, T., 2016. Efficient parallel boolean matrix based algorithms for computing composite rough set approximations. Information Sciences 329, 287–302. doi:10.1016/j.ins.2015.09.022.

[48] Ziarko, W., 2005. Probabilistic rough sets, in: Ślęzak, D., Wang, G., Szczuka, M., Düntsch, I., Yao, Y. (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 10th International Conference, RSFDGrC 2005, Regina, Canada, August 31 - September 3, 2005, Proceedings, Part I. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 283–293. doi:10.1007/11548669_30.

[49] Zolnik, E., 2021. Geographically weighted regression models of residential property transactions: Walkability and value uplift. Journal of Transport Geography 92, 103029. URL: https://www.sciencedirect.com/science/article/pii/S096669232100082X, doi:https://doi.org/10.1016/j.jtrangeo.2021.103029.

List of Tables