J. Geogr. Sci. 2023, 33(3): 618-638
DOI: https://doi.org/10.1007/s11442-023-2098-6

© 2023 Science Press Springer-Verlag

Multi-hierarchical spatial clustering for characteristic towns in China:
An Orange-based framework to integrate GIS and Geodetector

ZHANG Zuo[1], DOU Yuqian[1], LIU Xiaoge[1], *GONG Zhe[1,2]

1. School of Public Administration, Central China Normal University, Wuhan 430079, China;
2. School of Forestry and Horticulture, Hubei Minzu University, Enshi 445000, Hubei, China

Abstract: As China's economy and society continue to expand, urbanization in China has reached a new stage. In this context, China's characteristic town development plans, from the national to the local level, provide a new impetus for the expansion of towns and cities, and contribute to rural revitalization. When assessing the universality of regional dynamics, China exhibits high diversity for regional development. This highlights a complex scientific problem associated with describing the underlying linkages and influencing variables between distinct regional characteristic towns. It also complicates the application of tools that support spatial orientation and spatial decision-making. To address this problem, this study proposes a cross-platform analytical framework that unifies data, geography information systems (GIS), unsupervised analysis, visualization, and Geodetector, with Orange as the core. Based on a review of distribution patterns and multi-hierarchical spatial clustering features, this paper focuses on the rise of characteristic towns in China and investigates the primary environmental and human factors influencing spatial heterogeneity in small towns. The findings show that the development level of characteristic towns under the "city-town" system varies across China. Multi-hierarchical cluster analysis effectively reveals the intrinsic features of characteristic towns and facilitates precise spatially-oriented decision-making under different scenarios. In the framework of the "natural-humanistic" and "city-town" systems, the Geodetector effectively measures the spatial stratified heterogeneity of each indicator in the characteristic towns. This reveals an inner logic among the diverse non-linear spatial correlations. Ultimately, the study deeply investigates the individual development of characteristic towns, and the synergistic construction of "city-town" systems, arguing that characteristic towns have the potential to create "city-town" interactive spaces, and the ability to realize "Orange+GIS" cross-platform synergies.

Keywords: characteristic towns; spatial clustering; city-town system; Orange; GIS; Geodetector

1 Introduction

From urban planning, the "city-town" system refers to a group of closely connected and in-

Received: 2022-08-15 **Accepted:** 2022-11-14
Foundation: National Natural Science Foundation of China, No.72174071, No.71774066
Author: Zhang Zuo (1982–), PhD and Professor, E-mail: zhangzuocug@163.com
***Corresponding author:** Gong Zhe (1990–), E-mail: gongzhe901028@163.com

terdependent towns in a relatively complete region with different functional divisions, different hierarchical scales, and orderly spatial distribution (Hou *et al.*, 2015). Since the reform and opening up, the unprecedented rapid economic growth in China has created a tremendous gap between urban and rural development (Qian, 2017; Zhao and Bai, 2019). Exploring policy governance solutions with urban-rural interdependence to address the development imbalance between urban and rural areas plays a key role in forming a coordinated and orderly "city-town" system (Ye *et al*., 2022). Characteristic towns, with industry as the core, are viewed as one of the main carriers to promote China's urbanization and rural revitalization (Gu *et al*., 2015; Wang *et al*., 2019). Since 2016, various provinces and cities in China have been constructing different types of characteristic towns through a "one town, one characteristic" approach on a large scale under the guidance of the national strategy (Zhou *et al*., 2020). Among the "city-town" system, characteristic towns are essential nodes for urban and rural factors to flow, located in advantageous areas around urban clusters, and whether they have precise industrial positioning is crucial for achieving their sustainable development (Zou and Zhao, 2018).

There has been significant international discussion and research on the development conditions, dynamic mechanisms, and development strategies of small towns. Most studies examine the unique characteristics of small towns to establish their significance in the national land space (Filipovic *et al*., 2016). One example includes identifying small towns based on their sustainable development pillars, ultimately focusing on preserving their characteristics and functions (Vaishar *et al*., 2016). Almond (2020) categorized the common feature of American College Towns, which integrates higher education into the ordinary cultural ecosystem of small towns, where local higher education institutions tend to be underrepresented. Kwak *et al*. (2020) applied machine learning and statistical methods to investigate and evaluate different elements of Urban Heat Island Dynamics in Korean New Towns, to determine whether sustainable development policies support the town's expansion strategy. Extensive attention has been paid to foreign characteristic towns, but little research has been done on Chinese characteristic towns, most of which focus on the spatial pattern of small towns. Song and He (2022) assessed the expansion characteristics and periods of national characteristic towns about spatial scale, using the lighting dataset, and the implementation of policies significantly contributed to the growth of characteristic towns. According to Zhang *et al*. (2022), two types of newly developing industrial towns in China were compared, Characteristic Towns and Taobao Towns, examining their spatial patterns and their dependencies, and demonstrating that Characteristic towns gradually decrease from the east to west due to their extensive distribution.

Because of the different characteristic town types around the world, many studies have analyzed the feasibility of constructing a single industry-oriented research town based on a clustering method or sociological instruments (Dubnitskiy and Lunina, 2015; Baimurzina and Kabashova, 2020). These studies have assessed the effect of tourism development on the demographic attractiveness of small local towns (Stojanović *et al*., 2017; Zawadzka, 2021). For example, one study focused on the visual perceptions of the architectural elements in a historic temple town, exploring the architectural heritage (Kandasamy and Kesavaperumal, 2019). However, there has been little quantitative research on characteristic towns exploring the spatial dimension of local characteristics, particularly concerning using multi-hierarchical cluster analysis to explore policy science. Furthermore, studies have mainly focused on generic case presentations and experience summaries, with few studies examining the “city-town” system.

Recently, the Chinese central government released a series of policy documents to address issues related to characteristic towns. The policies reinforce the importance of characteristic towns at a national strategic level and encourage the proliferation of characteristic towns across the country in an “absorbing radiation” manner (Liu, 2020; Shao and Lin, 2021). During the actual creation of characteristic towns, there is a problem of imbalanced resource allocation between the two divergent driving paradigms held by local governments and private firms (Yu *et al*., 2018). At the same time, determining the types and allocation of resources needed by the towns frequently lacks an efficient and rational basis, given unequal and closed competitive processes (Liu, 2020). This highlights the importance of exploring whether the “top-down” policy direction approach conflicts with enhancing local characteristics. However, the rapid creation of characteristic towns has led to vicious inter-regional competition, which may create investment risks for less competitive cities (Zou and Zhao, 2018). Many characteristic towns are currently vaguely positioned, are imitated, or are even copied from one another in the construction process. This can lead to a high degree of homogenization in many towns, triggering real estate speculation with the goal of obtaining preferential policy benefits, and deviating from the original purpose of the “Thousand Towns, Thousand Strategies” scheme (Wu *et al*., 2018).

Given this backdrop, the first challenge in the sustainable development of China’s characteristic towns is emphasizing the town’s unique qualities (Chapman *et al*., 2015), while effectively integrating regional culture (Guan and Li, 2020). Additionally, while economic agglomeration created by characteristic towns might have spillover effects on the surrounding countryside, these benefits may differ from region to region (Courtney *et al*., 2008; Zhang *et al*., 2019). This demonstrates the necessity of considering the natural resources of towns with regional characteristics, the underlying linkages, and inequalities in their economic and social development, as well as their influencing factors. According to Stoica *et al*. (2020), policymakers should consider both the local development potential and the development conditions required for different types of small towns. That study also highlighted the need to be cautious of small towns losing their urban identity by exhibiting extremely rural characteristics.

The main methodological challenge in designing a research study in this domain is to combine many “data-model-space-visualization” elements across multiple system platforms. The research also involves complicated spatial quantitative analysis and scientific decision-making supports. Orange software is the main tool used for this study and is widely used in data mining and visual analysis (the free download web address is https://orangedatamining.com/). It was developed by the bioinformatics lab at the Faculty of Computer and Information Sciences of the University of Ljubljana in The Republic of Slovenia (Demšar *et al*., 2013). With approximately two dozen analysis modules, Orange can execute data mining through visual programming or Python scripts to develop prediction models for big data sets (Naik and Samant, 2016). For example, Godec *et al*. (2019) used Orange as a visual programming toolbox to simplify a picture analysis, by merging embedded deep learning, machine learning algorithms, and data visualization.

The principal clustering method used in this study is K-means clustering, which is a traditional cluster analytical technique (He and He, 2018). To examine the complicated properties of terrain distribution in a specific area, extended spatial clustering approaches (Morehouse, 2020) can be coupled with digital elevation model (DEM) data sources. This delivers the computational speed and information requirements required for simulation-based urban models (Ghiassi and Mahdavi, 2016). To support spatial analysis, GIS may improve the usability of systems by facilitating data integration and visualization (Shafabakhsh *et al.*, 2017). When examining the ground associated with towns, GIS and cloud computing technologies can be fully utilized to store, manage, analyze, and make judgments based on urban data (e.g., buildings, road networks, above and below ground facilities, etc.). This can help mitigate the new challenges and management challenges associated with urban planning (Gong *et al.*, 2017; Lv *et al.*, 2018). As an additional tool, Geodetector is a model for spatial analysis that detects the influence of intrinsic drivers, based on spatially stratified heterogeneity between variables (Wang *et al.*, 2016). It is extensively applied to analyze spatio-temporal evolution and to identify types of influences in the natural and social sciences (Zhao *et al.*, 2020; 2021). A review of past studies suggests that Orange's potential is not being fulfilled in the domain of spatial management and analysis in towns and cities. As such, this study combines GIS and Geodetector with Orange-based spatial clustering as the core to achieve this synergy and solve complex spatial decision-making problems in a systematic way.

This paper constructs a multi-platform and integrated analysis framework, with Orange as the core, and integrating GIS and Geodetector. The goal is to describe the complexity and heterogeneity associated with the spatial distribution features of the "city-town" system at multiple spatial scales, based on a foundational overview of the current development in China's characteristic towns. Under the "city-town" system, we mine the spatial dimensions of characteristic towns using GIS, and then further construct workflows around multi-hierarchical spatial clustering in the Orange platform. This includes data processing, unsupervised analysis, visualization, and geographical analysis. These steps reveal intrinsic correlations and heterogeneities within the spatial dimensions of the characteristic towns. In addition, we further investigate the factors influencing the spatial distribution, including the characteristic towns, using two models in Geodetector: the factor detector and interaction detector. Ultimately, this study provides a spatial multi-scale scientific orientation, as well as a foundation for measures based on the town itself, aiming at ensuring the high-quality, healthy, and sustainable development of characteristic towns.

The specific objectives of the study are as follows: (1) apply a multi-scale spatial order and systematic framework to investigate the spatial clustering relationships of characteristic towns under the "city-town" system; (2) use the classification results at different levels to explore the factors influencing the spatial distribution of characteristic towns under the "human-nature" framework; (3) investigate the application potential of Orange in spatial analysis, as well as an analytical framework integrating GIS and Geodetector, to provide scientific validity for further formulating refined and rationable spatial guidance policies considering the situation of the city or town itself. The rest of this article is structured as follows. Section 2 discusses the core data, the analytical framework, and the major approach to this study. Section 3 describes the spatial features and classification results. Sections 4 and 5 offer the discussion and conclusions

2 Materials and methods

2.1 Overall analytical framework

With the Orange-based framework as the core, this study proposes an integrated analysis framework (Figure 1), including three subframes: GIS, the kernel modules of Orange, and Geodetector. As the cornerstone of the framework, the GIS-based data analysis subframe supplies the fundamental input data for the subsequent Orange clustering analysis and the supportive environment for the Geodetector subframe with spatial analysis. The heart of this framework is the Orange-based model operation, which integrates and analyzes the relevant elements of the characteristic towns utilizing K-means calculation and visualization. The

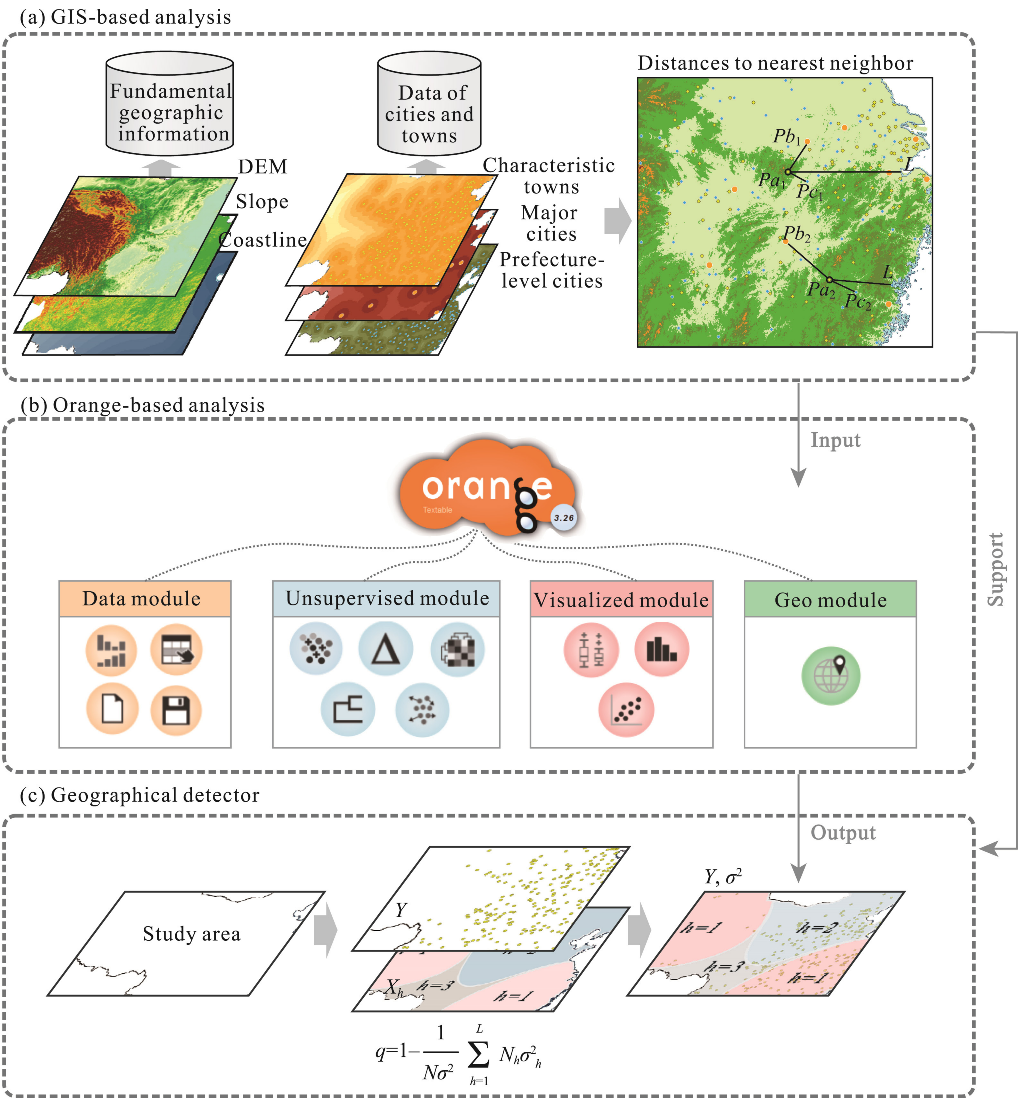

Figure 1 Integrated analytical framework. Notes: GIS-based analysis (a) generates further distances to the nearest neighbors based on the calculation of fundamental geographic information and data collated for cities and towns. Orange-based analysis (b) shows the main modules applied and their widgets. Geographical detector (c) shows the main calculation steps in Geodetector.

clustering results are then exported to the Geodetector to apply the factor detector and interaction detector. The following is a detailed explanation of these three subframes.

(1) The GIS-based subframe serves as a foundation for the in-depth investigation of the multi-dimensional spatial similarities and heterogeneity of characteristic towns. This section introduces two databases: basic geographic information data and city-town data. The first database contains DEM, Slope, and Coastline indicators, while Characteristic towns, Major cities, and Prefecture-level cities are contained in the second one. Quantified geographical features are derived by determining the distance between characteristic towns and searchable cities that are spatially adjacent to them using the city-town geographic data. Using the "city-town" system, relevant spatial elements database, and GIS analysis environment, the multi-dimensional spatial attributes and features of the characteristic towns in the "city-town" system are identified using spatial overlay analysis, spatial search, and distance calculation analysis.

(2) Data, Unsupervised, Visualized, and Geo modules are the kernel ones in the Orange-based subframe, with different colored widgets for the different modules, and are principally used for the spatial feature association and clustering of characteristic towns. In the "city-town" system, after identifying geographical patterns and mining spatial data, spatial clustering analysis is used to reveal multidimensional spatial similarity between characteristic towns. Specifically, by applying basic geographic information and diverse spatial feature data in GIS, this paper investigates the optimal alternative for spatial clustering under different linkages and levels of depth, based on the K-means core algorithm and the distance metric approach. This step serves as a foundation for the subsequent investigation of spatial influencing factors, as well as the comparison and optimization of policy classifications.

(3) The Geodetector subframe detects elements that produce spatial category heterogeneity in characteristic towns. To further investigate the driving factors behind the spatial linkages of characteristic towns and their influential roles, this paper utilizes Geodetector to quantify the effect of different geographical elements and spatial distance elements on spatial clustering heterogeneity. This process is based on the segmentation of characteristic towns in traditional regions and is combined with the multi-hierarchical spatial clustering results obtained in Orange. By examining the differences that exert by impact factors and mutual influence with respect to distinct characteristic towns under spatial clustering, this study attempts to explore the intrinsic logic of the pluralistic nonlinear spatial correlation embedded in the "city-town" system.

There are comprehensive impacts imposed on characteristic towns, which include natural, ecological, economic, and social impacts. Table 1 shows the 11 metrics analyzed by GIS in this study; all are highly related to advancements in characteristic towns. The indicators are divided into geographical and non-geographical indicators, including natural geographical conditions, human development environments (Liu *et al*., 2015; Choi *et al*., 2019; Grave *et al*., 2019; Zhou *et al*., 2021), and a more comprehensive collection of assessment indicators. Specifically, the geographical indicators processed using basic data in GIS include: distance to prefecture-level city (DisPC), distance to large and medium cities (DisLMC), distance to the coastline (DisC), mean elevation (DEM), and mean slope (Slope). Non-geographical indicators include the total number of people (POP), total population aged 15 to 64 (POPY), proportion of residents (PLR), per capita GDP (GDPPC), the proportion of non-agricultural population (PNP), and the proportion of non-agricultural output value (PNV). All of these are generated from basic data recorded in the statistics yearbook and official websites.

Table 1 The system of influencing factors

Dimension	Influencing factors	Abbreviation	Unit
Non-geographical indicators	Total number of people	POP	Person
	Total population aged 15 to 64	POPY	Person
	Proportion of local residents	PLR	%
	Per capita GDP	GDPPC	Ten thousand
	Proportion of non-agricultural population	PNP	%
	Proportion of non-agricultural output value	PNV	%
Geographical factors	Distance to prefecture-level city	DisPC	km
	Distance to large and medium cities	DisLMC	km
	Distance to the coastline	DisC	km
	Mean elevation	DEM	m
	Mean slope	Slope	°

2.2 Orange-based analysis

Based on the module type and operational flow, Figure 2 shows the Orange workflow diagram. Figure 2a illustrates how Orange's modules are combined to form a complete workflow for spatial clustering analysis. In addition, we cut the workflow from the perspective of the modules and the analysis phases to understand it clearly. Figure 2b depicts the critical calculation step, which includes loading all of the city-town data into the data module, briefly processing it to filter out the characteristic town data, and then dispatching it to the Unsupervised module for core computation and visualization. Figure 2c shows a detailed view of the operational flow of Orange, which is generally compatible with the process of "input – data visualization and statistics – core models and results – output". It is also one of the basic workflows of data mining and processing in Orange.

This study applies KNN (K-Nearest Neighbors) for clustering calculations, with K-means ++ algorithm as the initialization method and the Manhattan distance to assess distance in space. The goal of KNN is to locate the top k data items in the database that are almost parallel with the target, based on data comparability (Jia and Richards, 2005). K-means ++ can detect the mutual proximity of point elements in geographic space. It lowers computing instability by increasing iteration times, thereby eliminating the impact of the initial point selection in K-means on classification outcomes. The Manhattan distance is a straight line-based distance measurement. The absolute difference between coordinates is calculated to assess the similarity of objects and to obtain cluster membership (Chiu *et al.*, 2016). The characteristic towns have map-viewable geo-coordinate attributes that can be plotted using Geo Map widget in the Geo module.

2.3 GIS-based analysis

The cross-platform integrated analytics framework designed for this study uses Orange as its

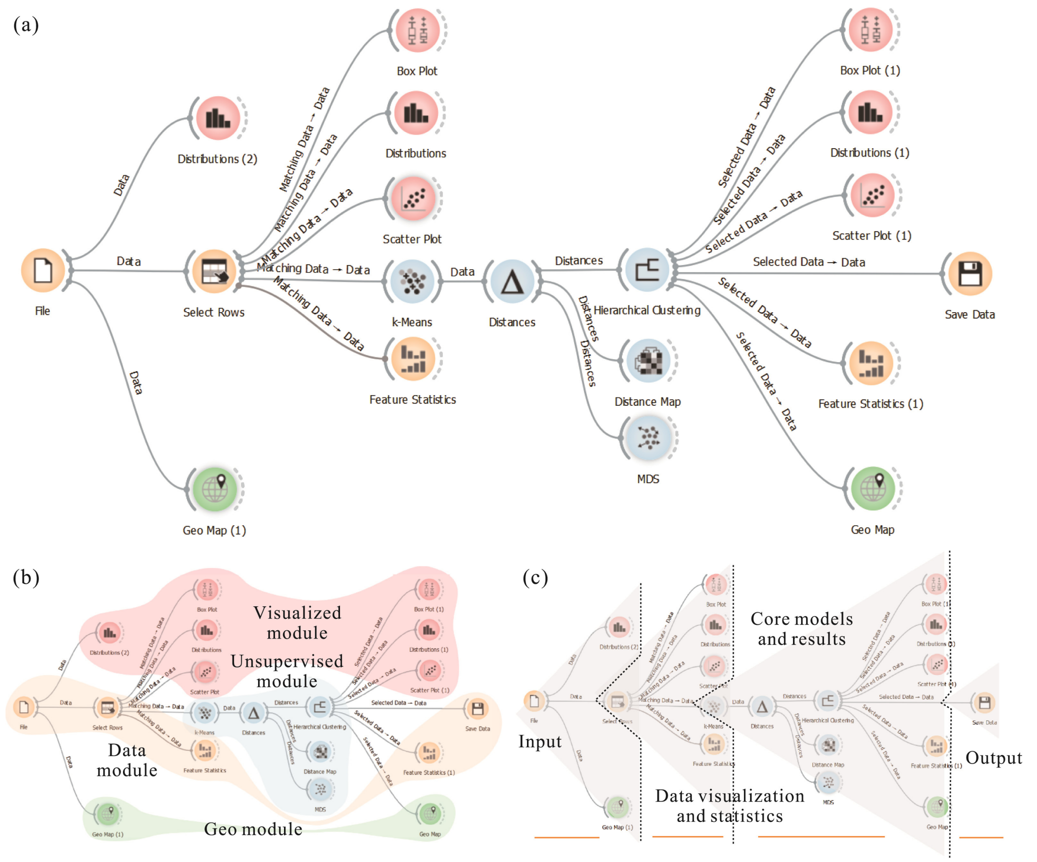

Figure 2 Analytical framework in Orange. *Notes:* (a) is the original workflow diagram in Orange; (b) and (c) are processed and illustrated from different perspectives based on figure (a). (b) demonstrates the division of the different functional modules, and (c) illustrates the main stages of the workflow.

center. The two primary functions of GIS in this framework are to supply the spatial dimensional data required for Orange-based clustering analysis (Table 1), and to provide the Geodetector with the appropriate spatial analysis environment support, that is, to provide the objects on which the force of the measurement factor operates. More specifically, the GIS platform is used to collect the fundamental geographic element data required for spatial analysis (e.g., distance to prefecture-level city, distance to large and medium cities, distance to the coastline, elevation, and slope). This is then used to build a "city-town" system comprised of large and medium cities, prefecture-level cities, and characteristic towns. To further characterize the relative geographical linkages of characteristic towns, this research uses NND (Nearest Neighbor Distance) to calculate the spatial distance relationships between any place on the map and cities as well as their spatial elements (e.g., coastline). NND is a statistical inference strategy that exploits the similarity between points and their nearest neighbors (Bar-Hen *et al*., 2015); the strategy is often applied in space to compare the observed measure of closeness and a reference point procedure, resulting in a hierarchical clustering of spatial points (Bar-Hen *et al*., 2015). The elevation and slope of the study region are shown in Figures 3a and 3b; Figures 3c, 3d, 3e, and 3f show the distribution of distances between any position in the study area and large and medium cities, prefecture-level cities, characteristic towns, and coastline.

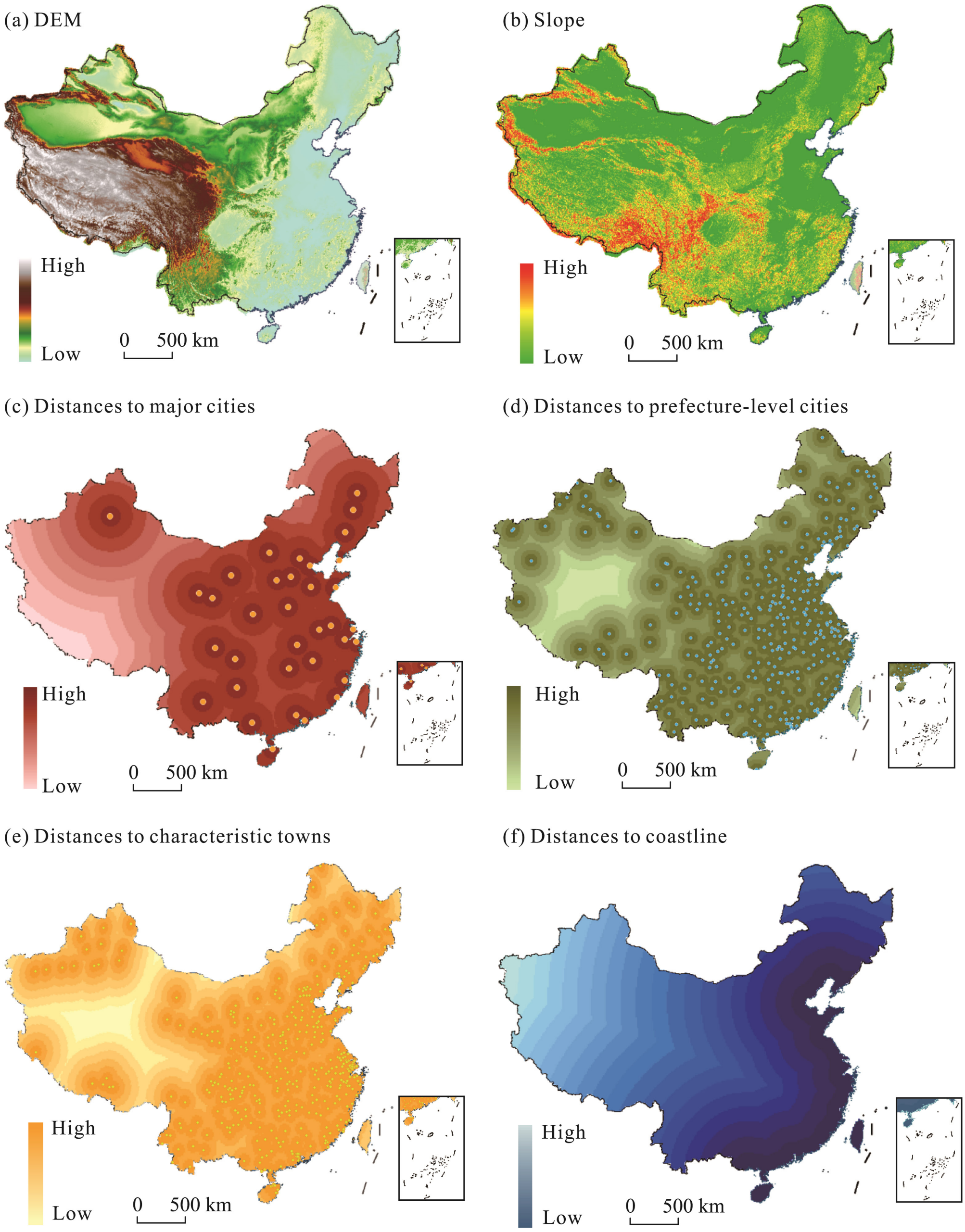

Figure 3 Data processing of basic geographic elements in GIS. Notes: DEM (a) is based on SRTM 30 m resolution elevation data provided by the Consortium for Spatial Information (http://srtm.csi.cgiar.org/Index.asp). Slope (b) is calculated by GIS based on DEM (a). (c), (d), (e), and (f) show the distances to large and medium cities, prefectural-level cities, characteristic towns, and coastline, respectively.

2.4 Geodetector-based analysis

Geodetector has been widely used in recent years to detect geographical divergence and to identify drivers (Zhu *et al.*, 2020; Yan *et al.*, 2021). The method assumes that if two independent factors have a significant effect on a dependent variable, then their spatial distributions should be similar (Wang and Hu, 2012). Geodetector has four main functions: factor detector, interaction detector, risk detector, and ecological detector (Wang *et al.*, 2016). This research explores the variables governing the spatial distribution of characteristic towns using two models within the "human-nature" context: the factor detector and interaction detector. In particular, the factor detector quantifies the effect of each independent variable on the dependent variable in terms of its q-value, and the interaction detector measures the interaction of two explanatory variables on a given target variable. This study uses the superposition to detect the joint effect of the two influencing factors on the spatial stratification heterogeneity of characteristic towns, to determine whether the effect is enhanced or weakened, or whether each is independent. This overcomes the problem that traditional regression models determine the interaction between two factors merely by using the multiplicative relationship.

3 Results

3.1 Spatial features

Figure 4 shows how the Orange-based widgets before the K-means are used to analyze spatial features. The Data module is used for the selection and statistics of data calculated by GIS, containing widgets of File, Select Rows, and Feature Statistics. The Select Rows widget (Figure 4c) filters the data by selecting characteristic towns, and the Feature Statistics widget (Figure 4d) provides a detailed count of all features of the characteristic towns. The remaining widgets, belonging to the Geo and Visualized modules, visually represent the important facts about the characteristic towns. The Geo map widget (Figure 4b) can display the spatial distribution of cities for different classes. Figures 4e, 4f, and 4g visualize the distinctive indicators about characteristic towns in the form of box plots, histograms, and scatter plots. The widgets specifically select different characteristic indicators to enable a comparative examination of the characteristic town's fundamental data.

In detail: (1) Figure 4e shows the average distribution of the characteristic town elevations in different regions of the country. The DEM distribution of characteristic towns across the country decreases steadily from west to east, largely corresponding to China's natural geography. (2) Figure 4f depicts the per capita GDP of characteristic towns, with a general pattern of uneven development. The characteristic towns in the east and central are expanding faster, with a bigger share of characteristic towns having higher per capita GDP levels. The characteristic towns in the west have a lower per capita GDP, with more concentrated values. (3) Figure 4g shows a scatter plot indicating a correlation between two indicators. When the DisC and GDPPC are chosen, the level of per capita GDP in western characteristic towns is not significantly affected by the distance to the coastline, since the western region is farther from the coastline. In contrast, there is a positive and weak correlation between DisC and GDPPC in the eastern and central regions, especially in the former. This highlights the need to further investigate the results using Geodetector.

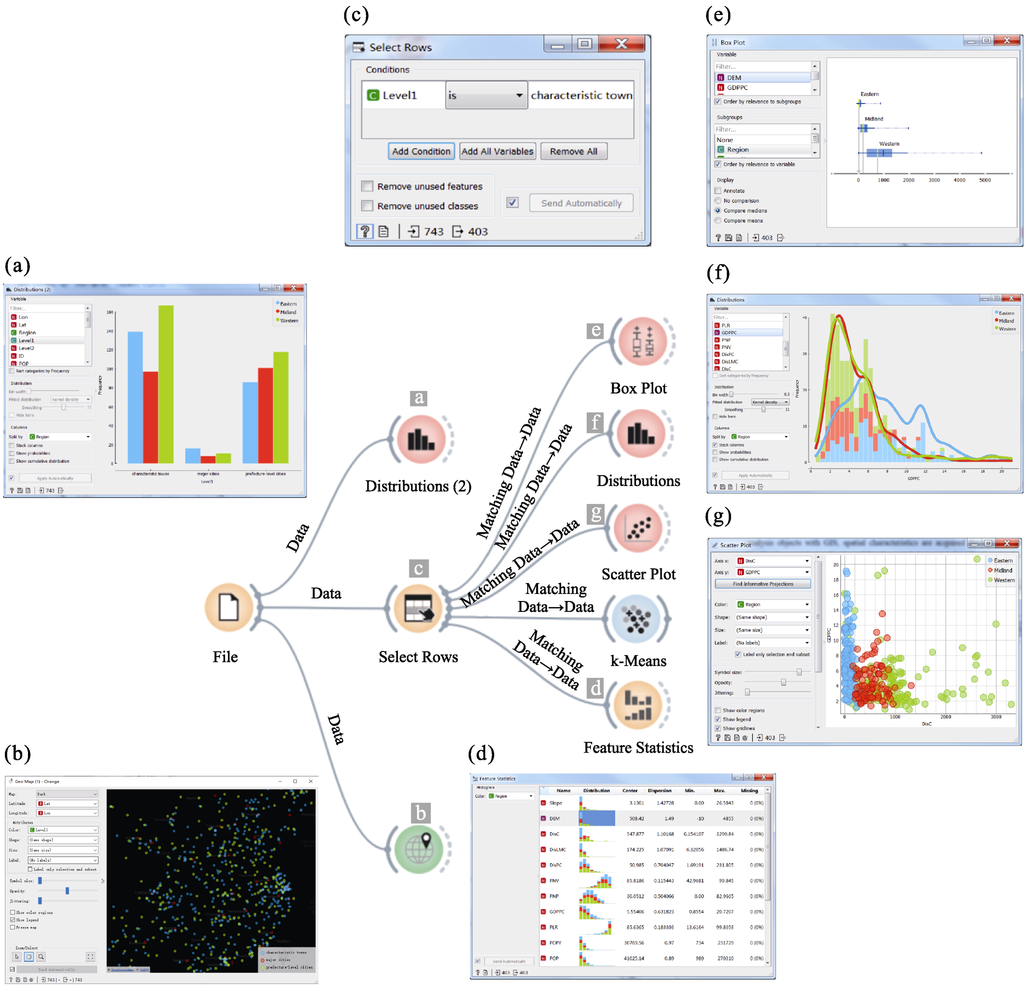

Figure 4 Analysis of the spatial distribution characteristics in Orange. Notes: (a) shows the histogram distribution about different levels of characteristic cities and towns in the eastern, central, and western regions. (b) depicts the distribution of the characteristic cities and towns on the map. (c) shows the screening procedure for characteristic towns. (d) shows the feature statistics for characteristic towns. (e) reveals a box plot of the DEM centered on characteristic towns in the eastern, central, and western regions. (f) shows a density graph and histogram distribution of GDPPC for characteristic towns in the eastern, central, and western regions. (g) shows a scatter plot of the association between GDPPC and DisC for characteristic towns in the eastern, central, and western regions.

3.2 Clustering results

The second half of the Orange-based analysis framework is dedicated to the clustering analysis (Figure 5), with the widgets of K-means, Distances, Hierarchical Clustering, Distance Map, and MDS (Multidimensional scaling) belonging to the Unsupervised module. This is the core computational part of this study. The Data module is used for data collation and output. The remaining widgets are part of the Visualized and Geo modules, used to visualize the clustering analysis thereafter (Figure 4). The Distance Map widget provides a broad overview of the clustering groupings but does not break them down further due to the sheer volume of data (Figure 5c). As a result, it is critical to fully describe the clustering outcomes within the Hierarchical Clustering widget. The study adopts silhouette scores ranging between [–1, 1] to indicate the quality of the clustering analysis. A higher value is associated with a more reasonable clustering. In addition, the ward linkage connects the clusters based on the degree of similarity between objects in the same cluster. Then, the internal sum of squares of each cluster is minimized, which means minimizing the intra-cluster variation and maximizing the inter-cluster variation. The clustering grouping outcomes are exported using different height ratios to display a more detailed classification of the characteristic towns (Figure 5e). MDS, like the Distance Map, is used to describe the dynamics of cluster grouping, except that MDS is based on the nearest distance of town indicators, and the Distance Map is based on the results of widget clustering (Figure 5d).

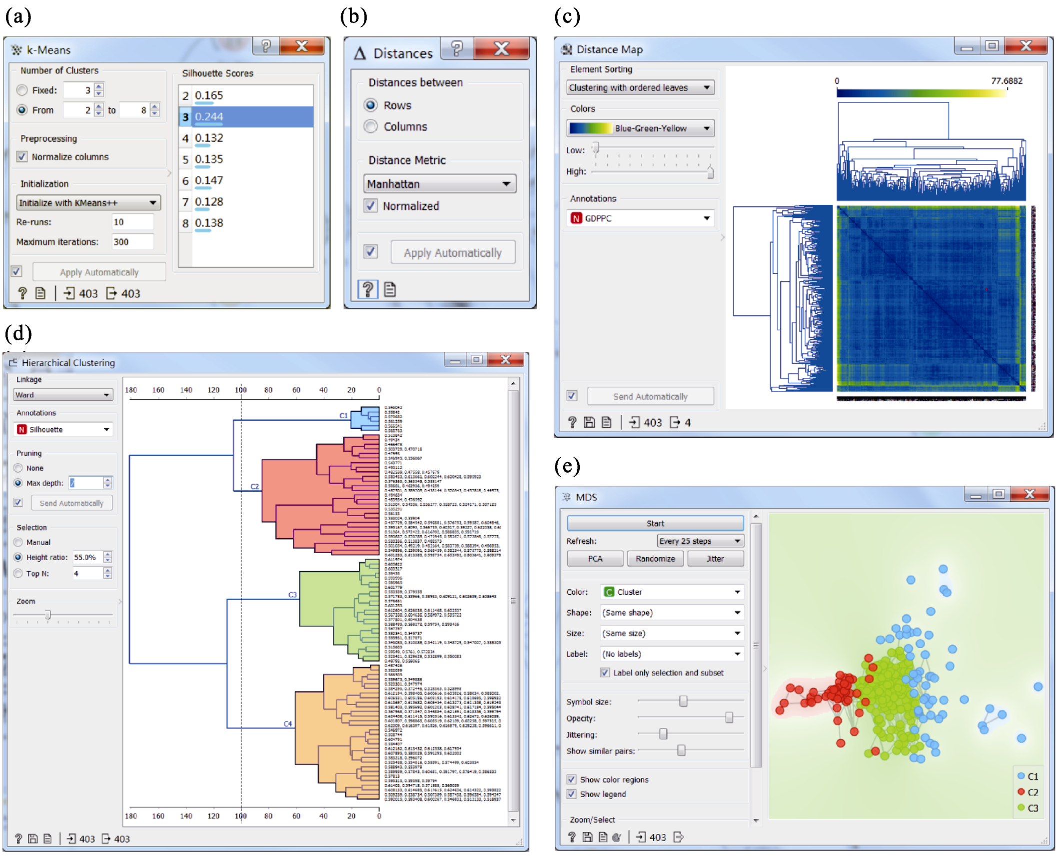

Figure 5 Clustering analysis in Orange. Notes: The K-means calculation is the core step of Orange (a). Based on the selection of manhattan distance (b), (c), (d), and (e) show the distance map, tree diagram, and MDS about the clustering results of the characteristic towns, respectively.

Figures 6 and 7 show the details of the clustering results, and the following points can be summarized.

(1) In the clustering results, the quantity and accuracy of clusters revealed vary depending on the height ratio used. Based on the dendrogram results of hierarchical classification about characteristic towns, this research selects extremely suitable height ratios of 65%, 55%, and 45% for linkage observation (Figure 6a). For example, when the height ratio is equal to 65%,

the hierarchical clustering results are divided into 2 groups: C1 and C2. Accordingly, the box plot, scatter plot, and spatial distribution map output by the clustering results produce corresponding changes (Figures 6b and 6c).

(2) The box plot ranks the attributes based on the degree of the selected clusters defined by the attributes. The example in Figure 6 illustrates the attribute of GDPPC and indicates that the data dispersion in different grouping scenarios consistently varies with the number of groups. When there are fewer groupings, the city-town data are more concentrated. The reverse is also true.

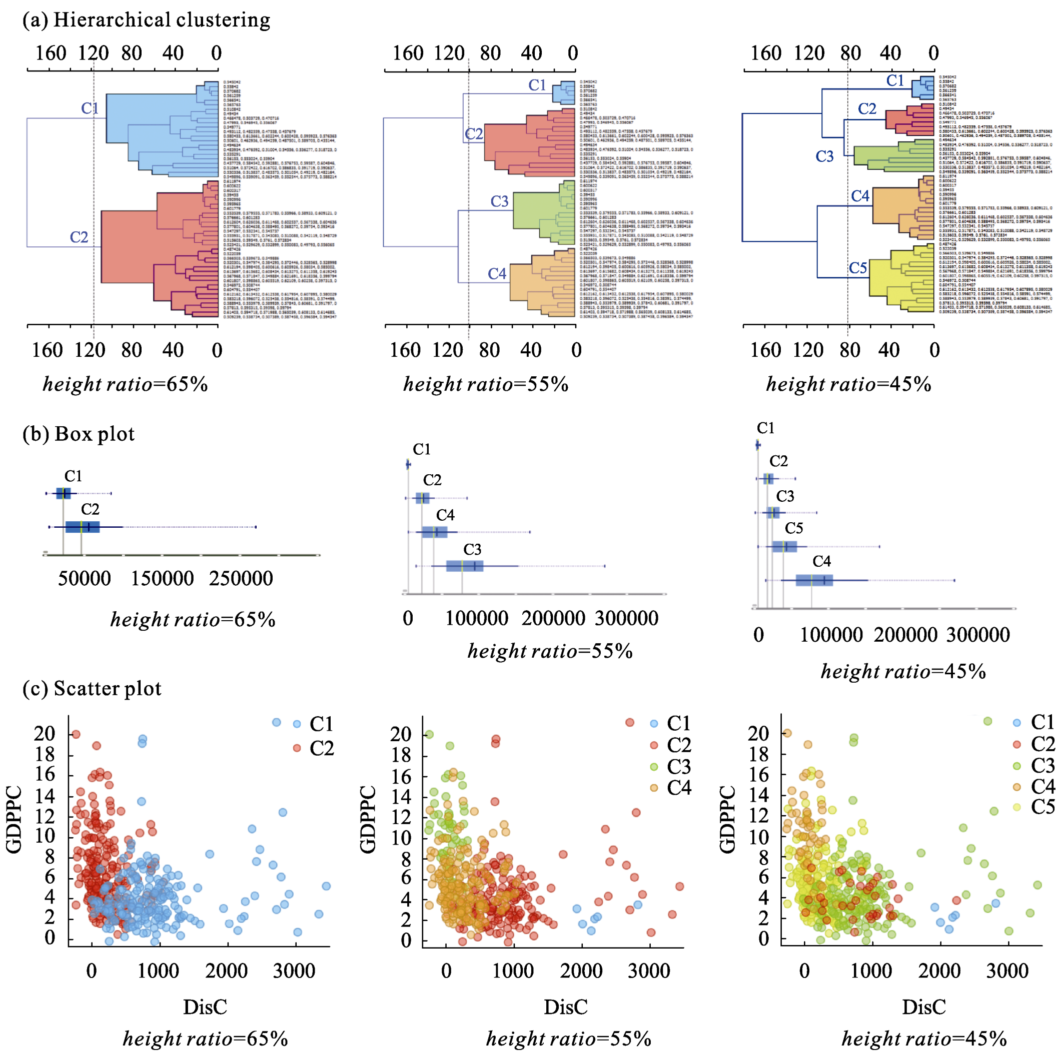

Figure 6 Visualization of clustering analysis in Orange. Notes: (a), (b), and (c) show the distribution about the clustering results of the characteristic towns in a tree plot, box plot, and scatter plot for the height ratio equal to 65%, 55%, and 45%, respectively, where the example variables selected for the box plot and scatter plot are GDPPC and DisC.

(3) The scatter plot is a ranking of different clusters based on the selected attributes. The example shows a roughly negative correlation between the distance to the coastline and the

per capita GDP. When the height ratio is 45%, the per capita GDP of the two clusters of C1 and C3 does not significantly differ with the change in the distance from the characteristic towns to the coastline. The distance from the characteristic towns to the coastline in C4 and C5 remains at approximately the same level, while a slight change in the distance leads to a significant difference in the per capita GDP.

(4) Figure 7 shows the spatial pattern for different clustering results. Combined with the scatter plot analysis, when the height ratio is equal to 45%, C1 and C3 are located in the northwest inland area, and the distance to the coastline has little effect on the per capita GDP of the characteristic towns. In contrast, C4 and C5 are located in the southeast coastal area, where the distance to the coastline plays an important role in economic development. The difference in coloring shows that the spatial distribution of characteristic towns essentially conforms to the Hu Line (Hu, 1935); there is a dense distribution in the southeast and a sparse distribution in the northwest. The characteristic towns located in the northwest are relatively few and concentrated. However, most characteristic towns rely on the resources of urban agglomerations to build using a batch approach. This results in many characteristic towns in the southeastern region, especially in the coastal areas, which are widely distributed and highly dispersed.

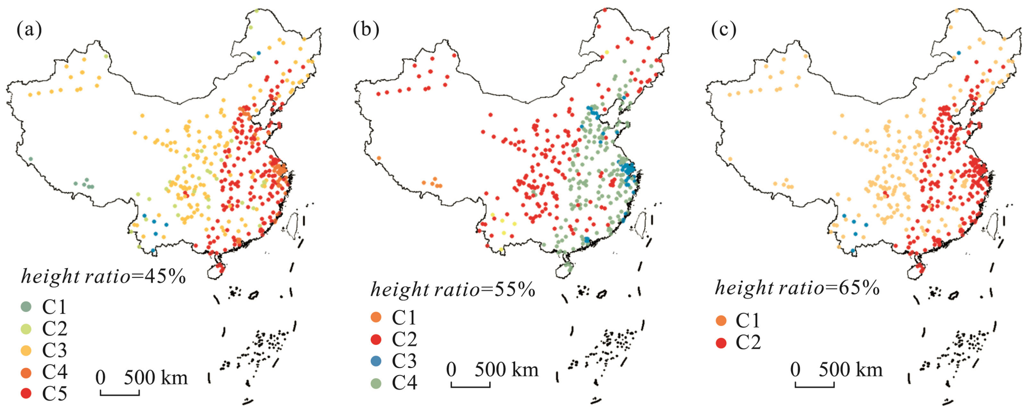

Figure 7 Spatial distribution of clustering results. Notes: (a), (b), and (c) show the spatial distribution for the height ratio equal to 65%, 55%, and 45%, respectively.

3.3 Results of Geodetector

Table 2 shows the factor detection results obtained using Geodetector, with the following results. (1) Given the detection results of the characteristic towns' impact factors in the eastern, central, and western regions in China, the key factors affecting the first column are Distance to the coastline, Mean elevation, per capita GDP, Total population aged 15 to 64, and Proportion of non-agricultural output value. (2) When the detection subject in the second column is the height ratio of 65%, the significant influences on the spatial patterns of characteristic towns are Mean elevation, Distance to the coastline, Total population aged 15–64, Total population, and er capita GDP. (3) When the height ratio is equal to 55% as the clustering criterion, the key factors found by factor detection in the third column are generally consistent with the height ratio of 65%; (4) When the height ratio is 45%, the principal elements affected in the fourth column are Mean elevation, Distance to the coastline, Mean slope, Distance to large and medium cities, and Total population aged 15–64. In general, the four factors that most influence the spatial distribution of characteristic towns are Mean elevation, Distance to the coastline, Total population aged 15–64, and Per capita GDP. These reflect the primary factors contributing to town construction: natural geographical advantages, abundant labor resources, and subsequent strong economic support. These results nearly satisfy the traditional construction concept of "favorable timing, geographical and human conditions," indicating that the current construction of characteristic towns should be rooted in real-world situations. At the same time, development should be tailored to the local conditions and advantages, so that the characteristic towns can be better integrated with the natural landscape and existing conditions.

Table 2 Results of the factor detector in Geodetector

Factors	Region	Height ratio=65%	Height ratio=55%	Height ratio=45%
POP	0.091***	0.217***	0.111***	0.123***
POPY	0.12***	0.246***	0.132***	0.138***
PLR	0.09***	0.034	0.011	0.008
GDPPC	0.151***	0.186***	0.11***	0.108***
PNP	0.064***	0.07	0.052	0.042
PNV	0.108***	0.105***	0.035	0.039
DisPC	0.027	0.058***	0.049	0.067*
DisLMC	0.082	0.083***	0.086***	0.139***
DisC	0.447***	0.443***	0.368***	0.371***
DEM	0.439***	0.696***	0.602***	0.616***
Slope	0.015	0.096***	0.087***	0.193***

Note: ***, **, and * mean that the variables passed the significance tests of 0.0001, 0.001 and 0.01 levels respectively.

Figure 8 shows the results of the interactive detection, delivering the following direct information. The two-factor interaction is either nonlinear enhancement or bi-enhancement. In other words, the influence of the interaction between factors on the spatial differentiation of the characteristic towns significantly exceeds the influence of a single factor, and there is no independent effect factor. This indicates that the spatial differentiation of characteristic towns comes from the combined effect of multiple factors, with a complex synthesis between spatial differentiation features and influencing factors. Further generalized, the top-ranked determinants with significant two-factor enhancement include Mean elevation, Distance to the coastline, and Per capita GDP. This is consistent with the previous results from the factor detection. This also indicates that the construction of characteristic towns should concentrate on areas where these three influencing factors produce collective efficiencies; the whole is greater than the sum of the parts. For example, it is important to site characteristic towns at locations with excellent development conditions, including being close to the coastline and having a moderately mean elevation. Moreover, the level of local economic development consistently impacts the healthy development of characteristic towns.

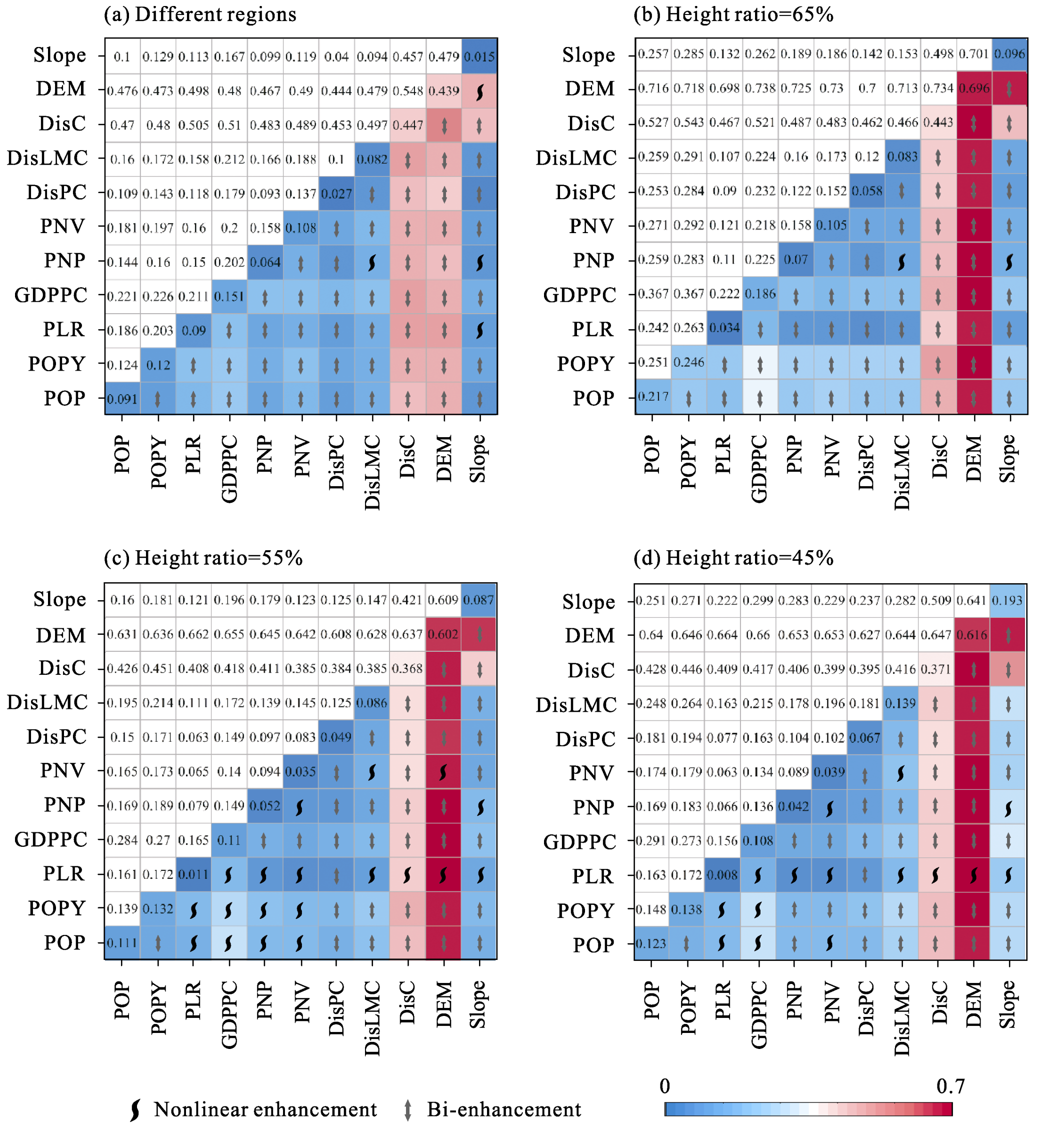

Figure 8 Results of the interaction detector in Geodetector. Notes: (a), (b), (c), and (d) show the results of interaction detectors for the characteristic towns under the eastern, central, and western regions as well as the height ratio equal to 65%, 55%, and 45%, respectively.

4 Discussion

Urbanization and rural revitalization are the two driving forces underlying China's urban and rural development (Chen *et al.*, 2021). However, bridging or narrowing the gap between urban and rural areas in China requires the organic blending of urbanization and rural revitalization. The development of characteristic towns is a "top-down" approach to small-town development within China's national innovation development policy. As a crucial node in the "city-town" network, characteristic towns play a role in mitigating the disadvantages of the "city-town" system when advancing China's urbanization. However, characteristic towns have a noticeable spatial radiation effect on the countryside; each is an essential junction and industrial spatial carrier linking the city and the countryside. Consequently, the characteristic towns, with their distinct spatial layout and development logic, may substantially affect the spatial organization of the region and its development trend.

The characteristic town scheme had only been effectively implemented in individual regions like Zhejiang province before being proposed at a national level (Zhao and Zou, 2021), leading to the question of whether a centralized government scheme is truly appropriate for China's diverse regional development. From the perspective of spatial layout, although there are few significant differences in the number of characteristic towns in each province, there are conspicuous regional disadvantages due to unequal regional resource allocation.

This study constructs a multiscale "city-town" network system of "large and medium cities – prefecture-level cities – characteristic towns" in a GIS environment using the NND method. A secondary "city-town" network system based on other spatial elements is then superimposed, to optimally support the "city-town-village" system of nearby urbanization and local urbanization. The findings indicate that multiple data applied to the domain in the spatial study using GIS could provide insights into the different stages and intrinsic mechanisms associated with urban and rural integration, while also optimizing the effectiveness and scientific nature of decision-making.

There are distinct levels of regional development in China with regional universality (Dedeoğlu Özkan and Beyazlı, 2018). To break the "top-down" growth of characteristic towns and truly realize the "one town, one policy" based on regional characteristics, it is important to discover the multi-dimensional geographic similarities of different characteristic towns in the "city-town" system and the inner correlation logic. In this setting, our study proposed a cross-platform analytical framework integrating data, GIS, unsupervised analysis, visualization, and Geodetector, with Orange at its heart. This framework is a systematic and effective approach that facilitates in-depth exploration of the intrinsic linkages and influencing factors between different regional characteristic towns, as well as precise spatial orientation and decision-making. Based on this analytical framework, we investigate the sustainable development paths of Chinese towns and cities from a complex spatial perspective, with a special focus on identifying the intrinsic links between the spatial characteristics of towns and providing a spatial decision-making foundation for optimizing city-town layout subsequently. To address multiple spatial decision scenarios, the Orange-based K-means computation and clustering algorithms provide a multi-dimensional spatial feature classification strategy for a variety of characteristic towns. During this research, the simplified operation and powerful functioning of Orange highlighted it as an important instrument to support academic research and improve future scientific decision-making. However, the framework also has some limitations in that it can only explain the main factors influencing the spatial location of characteristic towns under different scenarios and cannot yet achieve prediction. Additionally, the potential of Orange's application value in integrating with GIS has not yet been fully exploited.

Although it is an ideal solution to encourage the healthy growth of city-town systems and rural revitalization by means of characteristic towns, the process of implementation is difficult, requiring several "loops of learning by doing" cycles (Leutert, 2021). Achieving the sustainable growth of a unique community requires greater wisdom, further investigation, and more suitable industry chain. Drawing on the "human-nature" framework, this study investigates the drivers and influences underlying the spatial linkages of characteristic towns, considering more potential factors. To compare and analyze differences in the influence of factors and interactions among different spatial clusters (or regional divisions) in characteristic towns, Geodetector provides insight into the inner logic of the diverse non-linear spatial correlations implied in the "city-town" system. However, given data availability constraints, there is room for improvement in the quality, usability, and consistency of data about characteristic towns on a national scale. Further, the characteristic towns lack effective data integration despite rapidly expanding across province. The spatial dimension evaluation criteria for characteristic towns are not yet complete.

In summary, this paper selects an appropriate indicator system after a thorough investigation and based on references to other research findings. However, concerns remain about imprecise data and limited variables, highlighting the opportunity to broaden the study analysis.

5 Conclusions

Given the recent large-scale growth of characteristic towns in China, this study focuses on developing a precise spatial orientation and decision-making structure, to facilitate the customized development of characteristic towns based on regional features. The study completed the following three tasks: (1) investigating the multi-hierarchical spatial clustering features of characteristic towns from the perspective of China's "city-town" system; (2) evaluating the extent to which the main factors surrounding small towns affect the spatial heterogeneity about towns within the "human-nature" framework, and investigating the relationship between changes in physical geography, economic development, and the quality of construction of characteristic towns; and (3) using Orange as the core, proposing a cross-platform analysis framework that unifies data, GIS, unsupervised analysis, visualization, and Geodetector.

The results show that when applying Orange to cluster characteristic towns, the clustering results with height ratios equal to 65%, 55%, and 45% were selected, which provides a comparative perspective for studying different classes of characteristic towns and improves application scenarios for diverse decision-making options. The analysis presented by the Orange visualization widgets indicates that when there is a higher degree of clustering subdivision, the spatial distribution characteristics of the characteristic towns are also clearer. This is roughly distributed along the Hu Line, with plentiful and scattered characteristic towns in the southeast while rare and concentrated in the northwest. From a policymaking perspective, the main influencing factors measured by Geodetector include average elevation, distance to the coastline, and per capita GDP. These are the three decisive and enhancing factors, signaling the need to focus on these factors when selecting new characteristic towns. This means that small towns with a moderate or average elevation, a close distance to the coastline, and high per capita GDP should be prioritized list. Simultaneously, when constructing and developing a small town, it is important to focus attention on the number of people aged 15–64 in the labor force, and developing the per capita GDP in the town, as these two factors are the main contributors to its development. It is also important to attract a young and strong labor force and major metropolitan resources to support the development of characteristic towns.

Future research on characteristic towns should explore developing a territorial spatial charm as a crucial node to collapse the urban-rural dual structure, investigate the need for multiple data analysis tools, and enhance spatial visualization effects. Furthermore, it would be worthwhile for future studies to maximize Orange's potential, realize the cross-platform synergies of "Orange+GIS", and generate practical spatial analysis and decision-making applications.

References

Almond D, 2020. Everyday characteristics of American college towns: Identification and discussion. *Innovative Higher Education*, 45(4): 267–284.

Baimurzina G R, Kabashova E V, 2020. Features of socio-economic development of modern single-industry towns in the Republic of Bashkortostan. *Ekonomicheskie i Sotsialnye Peremeny*, 13(1): 106–124.

Bar-Hen A, Emily M, Picard N, 2015. Spatial cluster detection using nearest neighbor distance. *Spatial Statistics*, 14: 400–411.

Chapman R, Plummer P, Tonts M, 2015. The resource boom and socio-economic well-being in Australian resource towns: A temporal and spatial analysis. *Urban Geography*, 36(5): 629–653.

Chen M, Zhou Y, Huang X *et al.*, 2021. The integration of new-type urbanization and rural revitalization strategies in China: Origin, reality and future trends. *Land*, 10(2): 207.

Chiu W Y, Yen G G, Juan T K, 2016. Minimum manhattan distance approach to multiple criteria decision making in multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 20(6): 972–985.

Choi C, Jung H, Su L, 2019. Population structure and housing prices: Evidence from Chinese provincial panel data. *Emerging Markets Finance and Trade*, 55(1): 29–38.

Courtney P, Lépicier D, Schmitt B, 2008. Spatial patterns of production linkages in the context of Europe's small towns: How are rural firms linked to the local economy? *Regional Studies*, 42(3): 355–374.

Dedeoğlu Özkan S, Beyazlı D, 2018. Data set for the determination of regional development level: Criticism and recommendations. *Planlama-Planning*, 28.

Demšar J, Curk T, Erjavec A *et al.*, 2013. Orange: Data mining toolbox in Python. *The Journal of machine Learning Research*, 14(1): 2349–2353.

Dubnitskiy V, Lunina V Y, 2015. Development of single-industry towns on the basis of cluster approach. *Маркетинг і менеджмент інновацій*, (3): 140–148.

Filipovic M, Kokotovic K V, Drobnjakovic M, 2016. Small towns in Serbia: The "Bridge" between the urban and the rural. *European Countryside*, 8(4): 462.

Ghiassi N, Mahdavi A, 2016. A GIS-based framework for semi-automated urban-scale energy simulation. In: Hájek P, Tywoniak J, Lupíšek A *et al*. CESB16–Central Europe towards Sustainable Building 2016. Prague Czech: Grada Publishing, 999–1008.

Godec P, Pančur M, Ilenič N *et al.*, 2019. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. *Nature Communications*, 10(1): 1–7.

Gong H, Simwanda M, Murayama Y, 2017. An internet-based GIS platform providing data for visualization and spatial analysis of urbanization in major Asian and African cities. *ISPRS International Journal of Geo-Information*, 6(8): 257.

Grave P, Kealhofer L, Beavan N *et al.*, 2019. The Southeast Asian water frontier: Coastal trade and mid-fifteenth c. CE "hill tribe" burials, southeastern Cambodia. *Archaeological and Anthropological Sciences*, 11(9): 5023–5036.

Gu C, Li Y, Han S S, 2015. Development and transition of small towns in rural China. *Habitat International*, 50: 110–119.

Guan Z, Li P, 2020. Regional differentiation characteristics of the development ability of coastal villages and towns under the condition of market economy in the reform and transformation. *Journal of Coastal Research*, 103(SI): 120–124.

He Y, He Y, 2018. Urban Shanty Town recognition based on high-resolution remote sensing images and national geographical monitoring features: A case study of Nanning City. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(3).

Hou H, Liu Y, Liu Y *et al*., 2015. Using inter-town network analysis in city system planning: A case study of Hubei province in China. *Habitat International*, 49: 454–465.

Hu Huanyong, 1935. Distribution of China's population: Accompanying charts and density map. *Acta Geographica Sinica*, 2(2): 33–74. (in Chinese)

Jia X, Richards J A, 2005. Fast k-NN classification using the cluster-space approach. *IEEE Geoscience and Remote Sensing Letters*, 2(2): 225–228.

Kandasamy K, Kesavaperumal T, 2019. Holistic urban heritage management of an historic temple town: Kumbakonam, Tamil Nadu, India. *Journal of Cultural Heritage Management and Sustainable Development*, 10(2): 105–121.

Kwak Y, Park C, Deal B, 2020. Discerning the success of sustainable planning: A comparative analysis of urban heat island dynamics in Korean new towns. *Sustainable Cities and Society*, 61: 102341.

Leutert W, 2021. Innovation through iteration: Policy feedback loops in China's economic reform. *World Development*, 138: 105173.

Liu H, 2020. Spatio-temporal evolution mechanism and influence factors of the China's characteristic town policy and diffusion. *Open Journal of Social Sciences*, 8(7): 328–340.

Liu Y, Schen C, Li Y, 2015. Differentiation regularity of urban-rural equalized development at prefecture-level city in China. *Journal of Geographical Sciences*, 25(9): 1075–1088.

Lv Z, Li X, Wang W *et al*., 2018. Government affairs service platform for smart city. *Future Generation Computer Systems*, 81: 443–451.

Morehouse S, 2020. GIS-based Map Compilation and Generalization, GIS and Generalization. New York: CRC Press, 21–30.

Naik A, Samant L, 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, 85: 662–668.

Qian Z, 2017. Resettlement and adaptation in China's small town urbanization: Evidence from the villagers' perspective. *Habitat International*, 67: 33–43.

Shafabakhsh G A, Famili A, Bahadori M S, 2017. GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *Journal of Traffic and Transportation Engineering* (*English Edition*), 4(3): 290–299.

Shao M, Lin D, 2021. A study on how the five senses are affected when tourists experience towns with forest characteristics: An empirical analysis based on the data of Fujian, Guangdong and Sichuan in China. *Sustainability*, 13(15): 8283.

Song H, He T, 2022. Spatio-temporal dynamics of national characteristic towns in China using nighttime light data. *Remote Sensing*, 14(3): 598.

Stoica I V, Tulla A F, Zamfir D *et al*., 2020. Exploring the urban strength of small towns in Romania. *Social Indicators Research*, 152(3): 843–875.

Stojanović J, Kokotović Kanazir V, Stojanović M, 2017. Does small town with touristic function have demographic potential? *Journal of the Geographical Institute* "*Jovan Cvijic*", SASA, 67(2): 145–162.

Vaishar A, Zapletalová J, Nováková E, 2016. Between urban and rural: Sustainability of small towns in the Czech Republic. *European Countryside*, 8(4): 351.

Wang J F, Hu Y, 2012. Environmental health risk detection with GeogDetector. *Environmental Modelling & Software*, 33: 114–115.

Wang J F, Zhang T L, Fu B J, 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators*, 67: 250–256.

Wang Y, Zhu Y, Yu M, 2019. Evaluation and determinants of satisfaction with rural livability in China's less-developed eastern areas: A case study of Xianju county in Zhejiang province. *Ecological Indicators*, 104:

711–722.

Wu Y, Chen Y, Deng X *et al*., 2018. Development of characteristic towns in China. *Habitat International*, 77: 21–31.

Yan J W, Tao F, Zhang S Q *et al*., 2021. Spatiotemporal distribution characteristics and driving forces of PM2.5 in three urban agglomerations of the Yangtze River Economic Belt. *International Journal of Environmental Research and Public Health*, 18(5): 2222.

Ye C, Pan J, Liu Z, 2022. The historical logics and geographical patterns of rural-urban governance in China. *Journal of Geographical Sciences*, 32(7): 1225–1240.

Yu Y H, Xie W W, Dong Y S *et al*., 2018. The influence of factor endowment on financial industry agglomeration: An empirical study based on the financial characteristic town. *Journal of Interdisciplinary Mathematics*, 21(5): 1327–1332.

Zawadzka A K, 2021. Architectural and urban attractiveness of small towns: A case study of Polish coastal Cittaslow towns on the Pomeranian Way of St. James. *Land*, 10(7): 724.

Zhang X, Zhang M, He J *et al*., 2019. The spatial-temporal characteristics of cultivated land and its influential factors in the low hilly region: A case study of Lishan town, Hubei Province, China. *Sustainability*, 11(14): 3810.

Zhang Z, Zhan C, Li Z *et al*., 2022. Spatial patterns, dependencies, and disparities of characteristic towns and taobao towns in China. *Applied Spatial Analysis and Policy*, 1–26.

Zhao P, Bai Y, 2019. The gap between and determinants of growth in car ownership in urban and rural areas of China: A longitudinal data case study. *Journal of Transport Geography*, 79: 102487.

Zhao R, Zhan L, Yao M *et al*., 2020. A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. *Sustainable Cities and Society*, 56: 102106.

Zhao W, Zou Y, 2021. Creating a makerspace in a characteristic town: The case of Dream Town in Hangzhou. *Habitat International*, 114: 102399.

Zhao Y, Liu L, Kang S *et al*., 2021. Quantitative analysis of factors influencing spatial distribution of soil erosion based on geo-detector model under diverse geomorphological types. *Land*, 10(6): 604.

Zhou F, Zhao F, Xu Q *et al*., 2020. Evaluation and selection methods of tourism characteristic town: The case of Liaoning Province, China. *Sustainability*, 12(13): 5372.

Zhou L, Dang X, Mu H *et al*., 2021. Cities are going uphill: Slope gradient analysis of urban expansion and its driving factors in China. *Science of The Total Environment*, 775: 145836.

Zhu L, Meng J, Zhu L, 2020. Applying Geodetector to disentangle the contributions of natural and anthropogenic factors to NDVI variations in the middle reaches of the Heihe River Basin. *Ecological Indicators*, 117: 106545.

Zou Y, Zhao W, 2018. Searching for a new dynamic of industrialization and urbanization: Anatomy of China's characteristic town program. *Urban Geography*, 39(7): 1060–1069.