

Type of the Paper (Article, Review, Communication, etc.)

Using restaurant POI data to explore regional structure of food culture based on cuisine preference

Shangjing Jiang ^{1,2,3}, Haiping Zhang ^{1,2,3*}, Haoran Wang ^{1,2,3}, Zhou Lei ⁴ and Guoan Tang ^{1,2,3}

¹ School of Geographic Science, Nanjing Normal University; Nanjing, 210023, China; 181302072@njnu.edu.cn

² Key laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University; Nanjing 210023, China; 181302072@njnu.edu.cn

³ State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China; 181302072@njnu.edu.cn

⁴ School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023; zhoulai@njupt.edu.cn

* Correspondence: 90775@njnu.edu.cn.

Abstract: As a result of the influence of geographical environment and historical heritage, food preference has significant regional differentiation characteristics. However, the spatial structure of food culture represented by the cuisine culture at the regional level has not yet been explored from the perspective of geography. This study aims to explore such patterns by focusing on the restaurants of the eight most famous cuisines in Mainland China. Initially, the density based geospatial hotspot detector method is proposed to analyze and mapping the spatial quantitative characteristics of the eight major cuisines. A heuristic method for geographical regionalization based on machine learning was used to analyze spatial distribution patterns in accordance with the proportion of these cuisines in each prefecture-level city. Results show that some types of single-category cuisines have a stronger spatial concentration effect in the present, whereas others have a strong diffusion trend. In the comprehensive analysis of multicategory cuisines, the eight major cuisines formed a new structure of geographical regionalization of Chinese cuisine culture. This study is helpful to understand regional structure characteristics of food preference, and the density based hotspot detector proposed in this paper can also be used in the analysis of other type of POI data.

Keywords: food culture; cultural regionalization; Chinese cuisines; machine learning; spatial pattern

1. Introduction

Diet is not only a basic material element to meet human physiological needs, but it is also an important carrier of human cultural elements. The cultural phenomena related to the demand, production, and consumption of food were investigated in food culture studies [1,2]. Food culture has the characteristics of multi-disciplinary research, and the study of food culture from the perspective of geography emphasizes the human–environment relationship [3] and the geographical differentiation law reflected by the preference of food culture [4,5]. Geographical regionalization and mapping of food culture are important analytical and presentation methods to reveal regional differences of food culture, which are significant for mining local food culture resources and understanding local food culture characteristics [6]. At the same time, the regional structural characteristics presented by the food culture regionalization provide references for studying human–environment relationship in the geography of food culture [7].

Research on regional food culture is a trending topic in geography and related disciplines. Research topic includes two aspects, namely, what influences regional food cultures and what regional food cultures have influence on. The former focuses on how economy, politics, history, religion, and geographical environment affect regional food culture, whereas the latter focuses on the impact of regional food culture on people's health,

tourism, and other aspects, namely, impact of regional food cultures on the formation of sense of place and the construction of local cultural brands and symbols. In terms of the influence of economy on food culture, certain scholars reported that fast food chains in developed regions will be the first to enter and then affect the local food culture [8,9]. Several researchers also claimed that groups with different incomes will produce various food consumption cultures [10]. In the research on the relationship between regional food culture and politics, Tellstrom et al. studied how governments interpret and present their national image by shaping local food culture [11], that is, how the differences in political culture affect the food culture and daily food consumption patterns of people in different regions [12,13]. Historical and cultural customs and population migration in the history also profoundly impact local food cultures [14,15]. The relationship between religions and regional food culture is also a hot topic; for example, religious beliefs affect structures of food cultures, and the degree of attention to the green food differs between religious and nonreligious people [15-17]. In studies on geographical environment and food culture, climate, food materials and their proximity to coastal areas are the core influencing factors for the formation of regional food culture [18-20]. These studies discussed food culture as affected objects, that is, which factors affect the production and reproduction of food cultures. In the studies of food cultures, the relationship between regional food culture and health attracted the attention of scholars as the major influencing factor. For example, certain studies suggested that the food cultures of different regions always tend to develop in the direction that is beneficial to people's health. The food cultures of tropical regions help people keep cool, whereas those of cold regions help people stay warm [21,22]. Certain studies also suggested that the existing food culture in certain places is inconducive to people's health [23,24]. Another widely followed research topic is how regional food culture affects the development of tourism. Several studies reported that people's pursuit of local characteristic cuisine stimulated the interest of tourism developers and tourists in food tourism. In terms of academic research, food tourism is also an expanding field [25]. For example, Robinson et al. conducted a data-driven empirical study on gastronomic tourism among tourists [26]. In recent years, several scholars also focused on the production of cross-regional food culture and the construction of local sense [27,28].

Most studies regarded food as a culture and emphasized regional characteristics as inherent geographical attributes of food cultures. These studies also highlighted the human-environment relationship reflected in food cultures. However, the regional characteristics of people's dietary preference differences emphasized by geography are rarely mentioned in the literature. Only several studies are available on the establishment of boundaries of food culture and homogeneous-culture regions by geographical zoning method when the same region involves multiple food cultures. China puts importance on food culture. However, it has vast territory, diverse food culture, and evident regional differences in food culture. Hence, conducting geographical regionalization based on various food cultures in the Chinese context is necessary to have a deep understanding of the characteristics of regional food cultures.

In terms of research methods, geographical regionalization is usually adopted in the following ways: semi-manual cartography method based on experience and data [29,30], traditional GIS overlay analysis and cartographic synthesis method [31,32], and clustering method based on machine learning [33,34]. The semi-manual cartography method has low quantification degree and large workload. Although methods based on overlay analysis and cartographic synthesis have a high degree of automation, handling the weight distribution of geographical objects with complex features is difficult. At present, the multi-source clustering method based on machine learning has been introduced into the research of geographical regionalization and extended into the multisource clustering method with spatial constraint. In addition to its ability to handle the complex geographical regionalization with multi-source characteristics, the main advantage of this method is that it is a clustering method based on object model. The multisource clustering method with spatial constraints is adopted to conduct geographical regionalization of food cultures with multiple characteristics, which not only give play to the advantages of the

method but also solve two problems, namely, the complex features of cultural phenomenon and the difficulty of using traditional method for a reasonable regionalization result.

This study takes the entire China as research area, adopts the restaurants of the eight most famous great cuisines as the main data source, and takes prefecture-level cities as the main research units. It uses the preference index of each prefecture-level cities for each cuisine as the regional analysis variable and adopts regionalization method with spatial constraints for geographical regionalization of food cultures. The principle is to maintain the similarity of food preference structure within the same area and maximize the difference in different partitions. Finally, China is divided into several continuous food culture regions. Thus, the quantification of the regional structure of Chinese food culture based on the eight traditional cuisines is conducive to deep cognition of the spatial differentiation rules of Chinese food culture and excavation of local characteristic food culture.

2. Study area and data description

The research area is Mainland China. Taiwan, Hong Kong, and Macao are excluded due to the lack of data on catering facilities in these regions. The research area comprises 31 secondary administrative regions (provinces, autonomous regions, and municipalities). In the regional classification and geographical regionalization of food cultures, the basic geographical unit adopted in this study is prefecture-level city, which is the third administrative regionalization after the province or autonomous region. Prefecture-level cities are adopted as the analysis unit because their population reached a certain scale, and the number of catering service facilities can guarantee the diversity to a certain extent to avoid the situation where the number of certain cuisines is zero in certain analysis units. The whole research area is shown in Figure 1. The yellow highlights in Figure 1 are the provinces where the eight major cuisines are located and the birthplace of their respective cuisine culture.

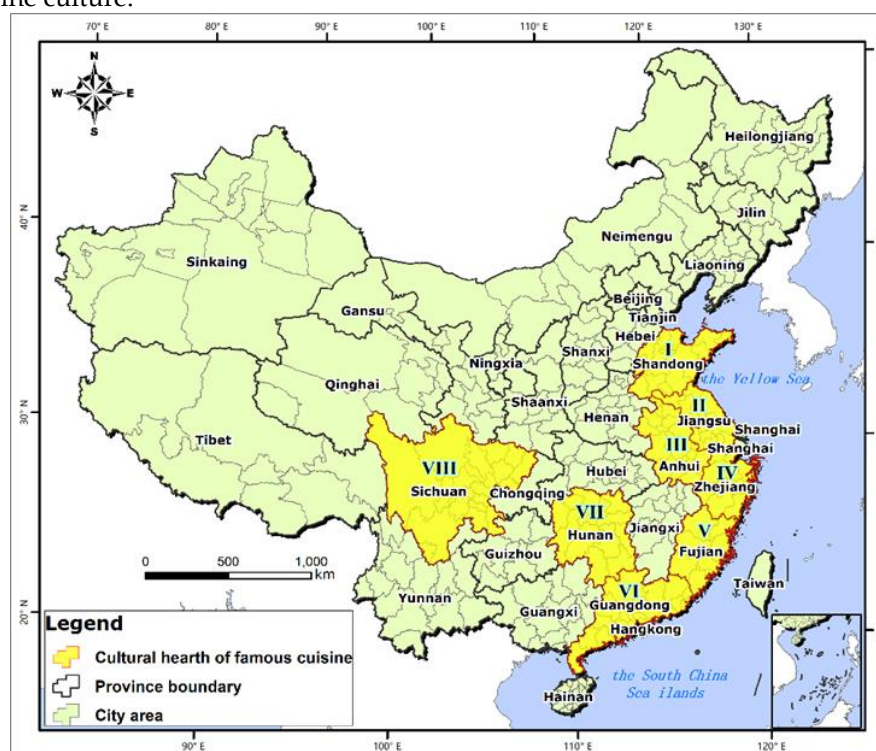


Figure 1. Study area and cultural hearth of eight famous Chinese cuisines

As the representative of Chinese food culture, the eight major cuisines have a long history of inheritance. These specialty restaurants can now be found in most Chinese cities, with some expanding rapidly and others slowly. For example, Sichuan and Hunan cuisines have the fastest diffusion, whereas Jiangsu and Shandong cuisines have the

slowest (Table 1). The data came from the online map service of AMap, China’s most famous online map service provider. In AMap map, the POI of catering type contains the classification field of specific catering type, which contains the classification value of traditional eight cuisines. Eight major cuisines and their geographical location information can be extracted using this field. Table 1 lists the names of the eight major cuisines, major origins, total number of restaurants nationwide, and percentage of restaurants in each cuisine.

Table 1. Quantity and proportion of eight famous Chinese cuisines.

Cuisine categories	Major origin place	Restaurants number	Restaurants number
Shandong cuisine	Shandong	I	4996
Jiangsu cuisine	Jiangsu	II	2273
Anhui cuisine	Anhui	III	9706
Zhejiang cuisine	Zhejiang	IV	18310
Fujian cuisine	Fujian	V	1495
Guangdong cuisine	Guangdong	VI	16771
Hunan cuisine	Hunan	VII	41585
Sichuan cuisine	Sichuan	VIII	129650

3. Research framework and methodology

3.1. Research framework

The whole research framework of this study is shown in Figure. 2. Initially, restaurants of eight cuisines are extracted from 7.5 million restaurants in China (classified labels in POI data of each catering service are provided). The analysis is then performed from the aspects of quantity and ratio. At the level of quantitative features, the hotspot detector is used to detect the spatial hotspots of Chinese restaurants of eight cuisines. The natural breaks method is then used to grade the hotspots detected. Finally, the spatial quantitative distribution characteristics of various cuisines are analyzed on the basis of hotspots with hierarchical structure. In, restaurants of various cuisines tend to concentrate in cities with large and high-density population and developed economy. To eliminate this effect, the proportion of the number of restaurants corresponding to each cuisine in the total restaurants in the prefecture-level city is calculated. At the ratio level, the method in Section 3.2.2 is adopted to classify all prefecture-level cities by taking the ratio of eight cuisines in each prefecture-level city as a regional variable. The result of type regionalization can reflect the structural characteristics of eight cuisines in each prefecture-level city and identify which eight cuisines are homogeneous in ratio structure. If a homogeneous prefecture-level city has spatial agglomeration, then comprehensive regionalization can be conducted on the basis of the ratio of eight cuisines as a regional variable to obtain a zoning map reflecting the regionalization and differentiation mode.

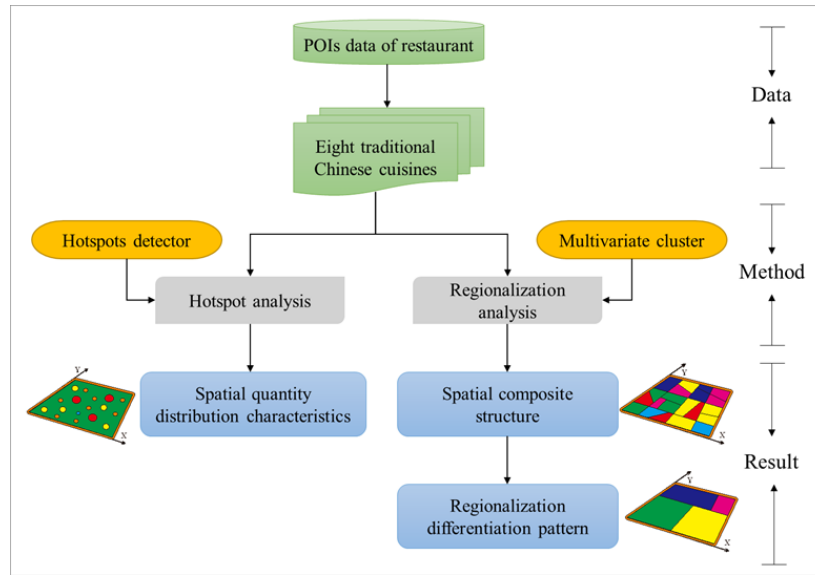


Figure 2. Research framework (analysis data, method, and result)

3.2. Methodology

3.2.1. Hotspot detecting method

An analytical method that can detect the maximum local density of all restaurants in the spatial distribution, where these local maximum values are called hotspots, is constructed to detect the maximum local density of restaurants. Hotspots can be classified with the clustering algorithm because of different sizes in the local maximum values. The hotspot ranking structure in the whole analysis region can then be obtained. A probability density surface is constructed via kernel density method for each cuisine. The obtained probability density is Matrix A with m rows and n columns, which can be expressed as

$$A = (a_{ij})_{mn} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} (i = 1, 2, \dots, m; j = 1, 2, \dots, n.) \quad (1)$$

For any element a_{ij} in Matrix A, the eight adjacent neighborhood elements, which constitute A matrix subset W_{ij} , are obtained, and $W_{ij} \in A$ is satisfied. The element, where the maximum value in W_{ij} , is obtained. All the element values in W_{ij} are set as the maximum value, and the new matrix subset $W_{ij_{max}}$ is finally obtained. These steps are performed for all the elements in matrix A to obtain a new matrix B. For good understanding, the above logic is illustrated with an example. Here, the sample matrix is expressed as

$$A' = (a_{ij})_{44} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \mathbf{a_{14}} \\ a_{21} & \mathbf{a_{22}} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ \mathbf{a_{41}} & a_{42} & \mathbf{a_{43}} & a_{44} \end{bmatrix} (\mathbf{a_{14}} > \mathbf{a_{22}} > \mathbf{a_{41}} > \mathbf{a_{43}}) \quad (2)$$

The elements in bold are assumed to be locally maximized.

$$W_{11} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a_{11} & a_{12} \\ 0 & a_{21} & \mathbf{a_{22}} \end{bmatrix} \quad (3)$$

Certain neighboring elements do not exist because a_{11} is at the edge of the matrix. Hence, it is zero. $W_{11_{max}}$ is obtained and satisfied as

$$W_{11_{max}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{a_{22}} & \mathbf{a_{22}} \\ 0 & \mathbf{a_{22}} & \mathbf{a_{22}} \end{bmatrix} \quad (4)$$

Similarly,

$$W_{12_{max}} = \begin{bmatrix} 0 & 0 & 0 \\ a_{22} & a_{22} & a_{22} \end{bmatrix}, W_{13_{max}} = \begin{bmatrix} 0 & 0 & 0 \\ a_{14} & a_{14} & a_{14} \end{bmatrix}, W_{14_{max}} = \begin{bmatrix} 0 & 0 & 0 \\ a_{14} & a_{14} & 0 \\ a_{14} & a_{14} & 0 \end{bmatrix} \quad (5)$$

For an element without boundary effect, such as a_{23} , because of

$$a_{23} = \begin{bmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{34} \end{bmatrix} (a_{14} > a_{22}) \quad (6)$$

The calculation result of $W_{23_{max}}$ is

$$W_{23_{max}} = \begin{bmatrix} a_{14} & a_{14} & a_{14} \\ a_{14} & a_{14} & a_{14} \\ a_{14} & a_{14} & a_{14} \end{bmatrix} \quad (7)$$

When all the elements are computed, the local maximum matrix B' of the entire example matrix A' is computed as follows:

$$B' = \begin{bmatrix} a_{22} & a_{22} & a_{14} & a_{14} \\ a_{22} & a_{22} & a_{22} & a_{43} \\ a_{41} & a_{41} & a_{22} & a_{43} \\ a_{41} & a_{41} & a_{43} & a_{43} \end{bmatrix} \quad (8)$$

Finally, the difference value between the local maximum matrix B' and the original probability density matrix A' is calculated, and each element is reassigned as follows: when $a_{ij} < 0$, $a_{ij} = 0$; when $a_{ij} \geq 0$, $a_{ij} = 1$. Then another new matrix C' can be obtained as follows:

$$C' = B' - A' = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (9)$$

In C' , elements with a value of 1 are at the hotspot. If C' is multiplied to the prime, then the matrix R' with the hotspot value is obtained.

$$R' = A' C' = A' (B' - A') = \begin{bmatrix} 0 & a_{12} & 0 & a_{14} \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a_{41} & 0 & a_{43} & 0 \end{bmatrix} \quad (10)$$

The method of detecting hotspots based on a restaurant service facility is shown above, and the results are obtained by detecting the hotspots of each cuisine of restaurants and be divided into different levels by natural break clustering method.

3.2.2. Regional preference index of cuisine

Regional preference index refers to the popularity of a cuisine in a certain region. In this study, R_i represents the i th geographical unit; N_i represents the total number of restaurants in the geographic unit R_i ; and M_{ij} represents the number of restaurants in cuisine j in the geographical unit i . Regional preference index can then be calculated by

$$RPI_{ij} = \frac{M_{ij}}{N_i} \quad (11)$$

In this study, the value of j is $j \in (1, 2, 3, 4, 5, 6, 7, 8)$, and each element represents every cuisine. The geographical unit is prefecture-level city. Shanghai is the geographical unit $i (i = 1)$, whereas Sichuan is the cuisine $j (j = 1)$. In this study, $N_1 = 18000$, and $M_{11} = 4000$. Hence, $RPI_{11} = \frac{4000}{18000} = 0.011$. On this basis, the preference index of Sichuan cuisine is 0.011 in Shanghai. Similarly, we can calculate the other preference index of the other cuisines in Shanghai and calculate their preference index using the same methods.

3.2.3. Comprehensive regionalization method

Geographic regionalization is a process for classification of spatial geographical objects, and the result is the merging of adjacent regions with homogeneity in multiple attributes into the same geographical partition in the form of surface elements to form a regionalized map composed of multiple regions. This process is significant to understand the similarity and heterogeneity between regions deeply. In this study, a heuristic method

of regionalization in machine learning that can effectively realize the classification of regions is adopted. The classification without region constraint only needs K clustering method based on multiple attribute information. The methods with region constraint include the construction of adjacency matrix, generation of minimum span tree based on adjacency matrix, and segmentation of minimum span tree. The relationship between regional classification without regional constraints and regional integration with regional constraints can be expressed as follows: regional integration is meaningful only when the results of regional classification have significant homogeneous regional agglomeration effect. Therefore, we classify types using this method and judge the clustering degree of homogeneous regions. If the clustering effect is evident, then the geographic regionalization will be further developed. The regionalization process is briefly introduced below.

A set of more than 300 local surface sources in China is defined as O . The surface source set O contains the attribute set $X = \{A_1, A_2, \dots, A_n\}$, which in turn contains the vector $x = \{a_1, a_2, \dots, a_n\}$. The attribute values in x correspond to the attribute columns in the attribute set X . If the region constraint is disregarded and only the type regionalization is performed, then the region classification in O is implemented directly via K clustering. If the region constraint is considered, then the adjacency matrix among surface sources in O needs to be constructed. The adjacency topological relation between surfaces in set O can be expressed as $G = (V, L)$, where V and L are the set of nodes and edges of the topological tree, respectively. For the two nodes v_i and v_j with adjacent relations, the attribute vectors of the plane object in O corresponding to them are x_i and x_j , respectively. The cost distance between the objects represented by v_i and v_j is defined as $d(v_i, v_j)$, and the formula is as follows [33]:

$$d(v_i, v_j) = \sum_{l=1}^n (x_{il} - x_{jl})^2 \quad (12)$$

where x_{il} and x_{jl} are the attribute of node v_i and v_j , respectively. The sum of all the differences between the attributes of the two nodes is the cost distance defined in this article. The whole tree G is divided into c node clusters G_1, G_2, \dots, G_c by using the construction and segmentation strategy of the minimum span tree.

This method also has the perplexity of NP problem because it belongs to the unsupervised space classification algorithm. Therefore, no optimal solution exists. The number of types or regions can be customized, or the validity of the number of different categories can be determined by the Calinski–Harabasz pseudo-F test. Calinski–Harabasz pseudo-F statistic provides the optimal classification scheme by calculating the ratio d of inter- and intra-cluster variance. In other words, it is the ratio reflecting intragroup similarity and intergroup difference. The evaluation formula is as follows [35]:

$$F = \frac{\left(\frac{R^2}{n_c - 1}\right)}{\left(\frac{1 - R^2}{n - n_c}\right)} \quad (13)$$

whereas:

$$R^2 = \frac{SST - SSE}{SST}, \quad (14)$$

where SST reflects the differences between different groups, whereas SSE reflects the similarity of attributes within a group. The formula for the two variables can be expressed as

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (v_{ij}^k - \bar{v}^k)^2 \quad (15)$$

$$SSE = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (v_{ij}^k - \bar{v}_i^k)^2, \quad (16)$$

where n is the number of surface elements; n_i is the number of elements in the first group i ; n_c is total number of regions; n_v is variables involved in the regionalization; v_{ij}^k represents the value of variable k of the key element j in the group i ; \bar{v}^k is the mean value of variable k ; and \bar{v}_i^k is the mean value of variable k in group i . In this study, Calinski–Harabasz pseudo-F test is used to evaluate the optimal group number for regionalization. When F test value reaches the maximum value at a certain classification value, this

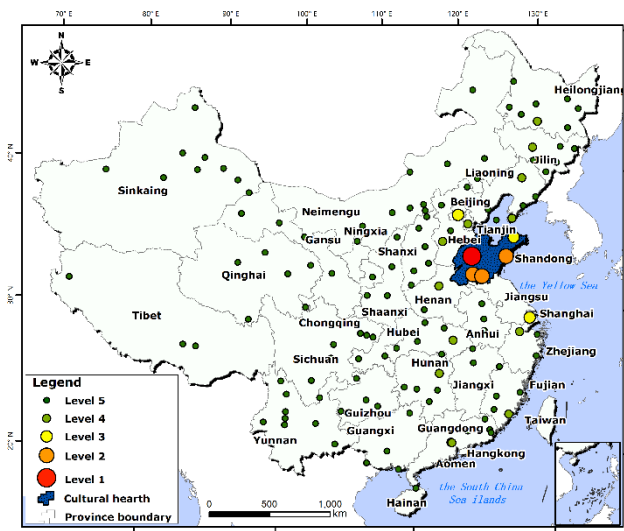
classification is the best grouping value with the smallest difference within the group and the largest difference between groups.

4. Result

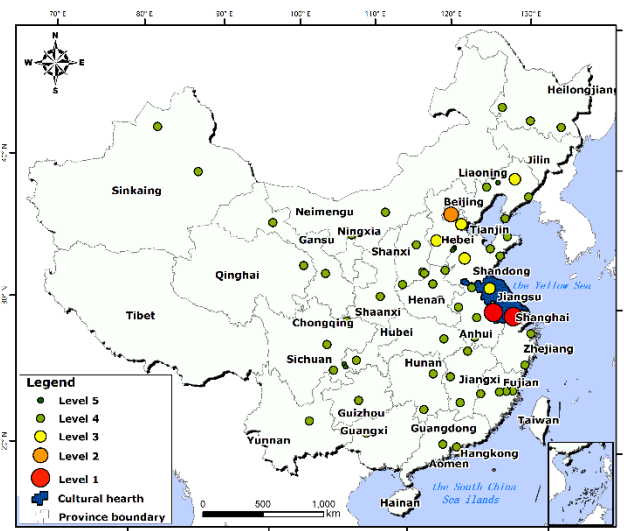
4.1. Spatial quantitative characteristics

The first-level hotspots of Shandong and Jiangsu cuisines were located in their originally cultural region, whereas the second- and third-level hotspots were relatively fewer (Figure 3a and b). Shandong cuisine had no evident trend of outward diffusion, whereas Jiangsu cuisine diffused five-level northward. The diffusion trend of Shandong cuisine was insignificant, but the Jiangsu cuisine diffused to the north. The hotspots of Shandong cuisine covered the whole country. Conversely, no five-level hotspots of Jiangsu cuisine existed. In the distributing map of hotspots of Anhui cuisine, the only hotspot of the first level was located in Shanghai city outside the originally cultural region (Figure 3c). Hunan cuisine also showed a similar phenomenon (Figure 3g). Its only hotspot was located in Guangdong Province rather than in its own originally cultural region. Anhui, Shandong, and Jiangsu cuisines had the same features wherein the hotspots in the first three level were fewer than in other levels. In addition, Anhui cuisine also had a significant trend of diffusion to the north. By contrast, the middle and high-level hotspots of Zhejiang and Fujian cuisines were widely distributed throughout the Middle East and East China (Figure 3d and h). The Zhejiang cuisine diffused significantly to the north, especially along the coast (Figure 3d). Fujian cuisine formed three second-level hotspots and two first-level hotspots within the province and formed a second-level hotspot in Guangdong and Shanghai, which were economically developed areas. High-level hotspots were concentrated for Guangdong cuisine. Only one first-level hotspot was located in the originally cultural region, and only one second-level hotspot was located in Shanghai. The most hotspots were in Hunan and Sichuan cuisines in eight major restaurants. The number of high-level hotspots of Sichuan cuisine was large and distributed all over the country (Figure 3g and h).

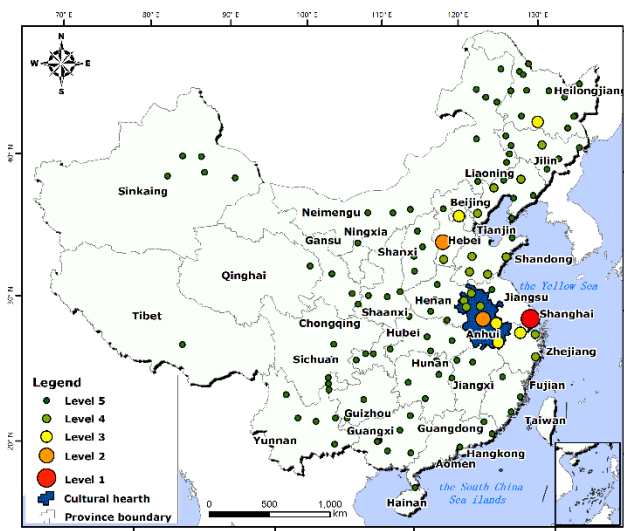
The spatial diffusion ability of northern cuisines (such as Shandong or Jiangsu cuisines; Figure 3a and b) was weak, whereas southern cuisines had strong spatial diffusion ability (Figure 3g). Shandong cuisine had the weakest diffusion ability among the eight major cuisines, whereas Sichuan cuisine had the strongest diffusion ability. Among the eight major cuisine provinces, Anhui Province was the weaker one in the north, and Hunan Province was the weaker one in the south. The cuisine of these two provinces had only one first-level hotspot and was located outside the originally cultural region. The first-level hotspot of Anhui cuisine was located in Shanghai, which was the most developed city in its neighborhood. Meanwhile, the first-level hotspot of Hunan cuisine was located in Guangzhou, which was the most developed city in its neighborhood. In short, the spatial distribution of quantity and the characteristics of hierarchical structure varied greatly among different cuisines.



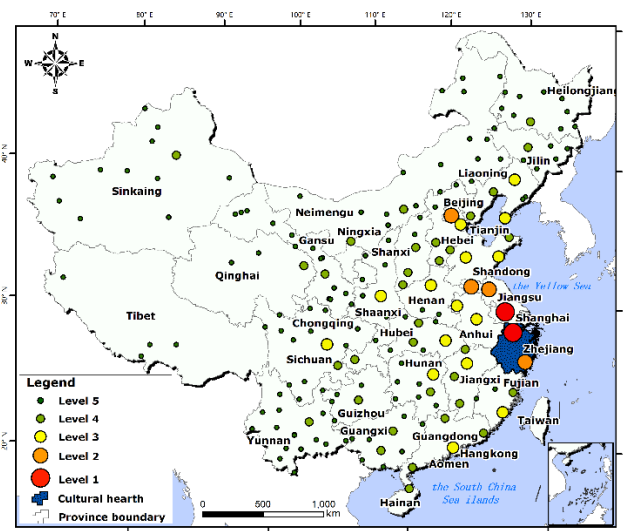
(a) Shandong cuisine



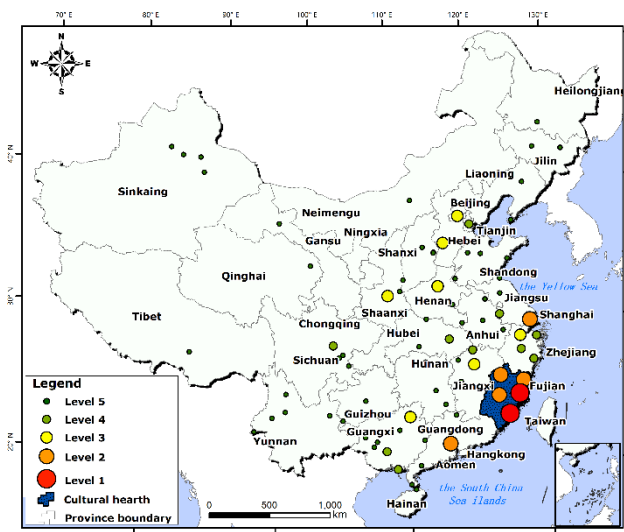
(b) Jiangsu cuisine



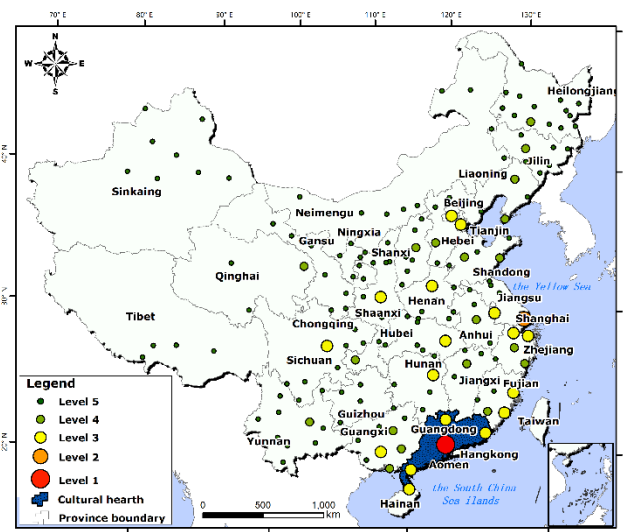
(c) Anhui cuisine



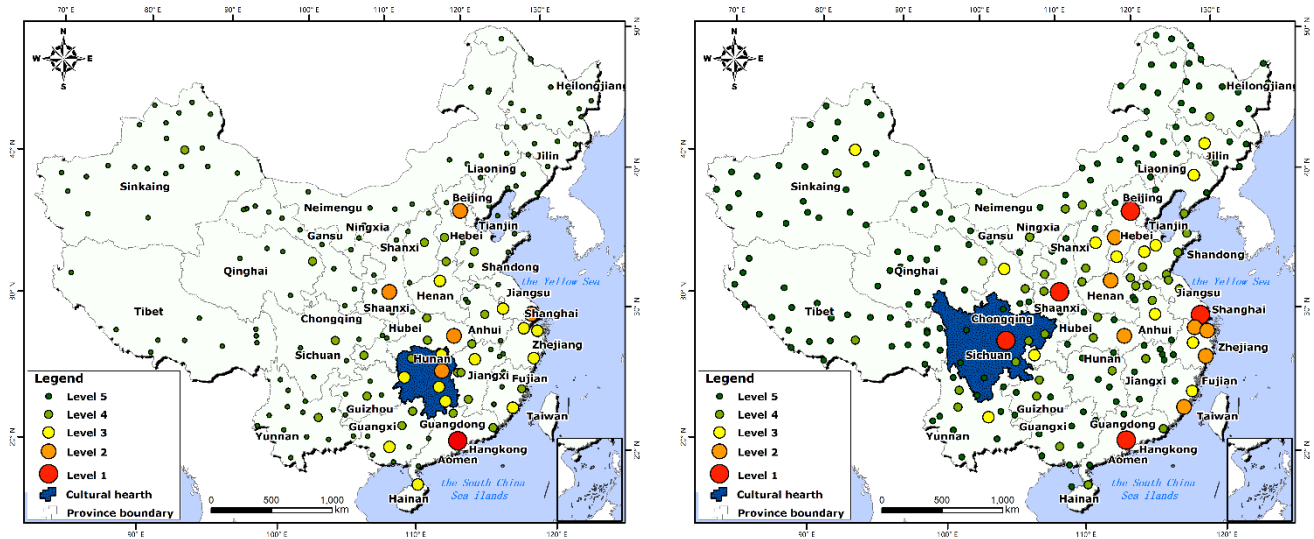
(d) Zhejiang cuisine



(e) Fujian cuisine



(f) Guangdong cuisine



(g) Hunan cuisine

(h) Sichuan cuisine

Figure 3. Results of hotspots detecting and divide rank of Eight Chinese cuisine.

4.2. Regionalization differentiation structure

From a regional perspective, exploring the spatial distribution of the proportion of eight major cuisines will help further exploration of people's preferences in different regions for various cuisines. The ratio of restaurants corresponding to different cuisines in each prefecture-level city to the total number of restaurants in that prefecture-level city was calculated. The method introduced in Section 3.2.3 was used to model all prefecture-level cities. The pseudo-F value was the highest when the number of classifications was 20. Hence, the threshold of the number of classifications was set to 20. The result of region type regionalization was obtained (Figure. 4). The prefecture-level cities with Type Code 9 were widely distributed all over the country, and the dietary preferences of the eight major cuisines in these prefecture-level cities were similar. In addition, many prefecture-level cities with the same classification number showed spatial agglomeration effect, which means that the local cuisine preferences were similar. This phenomenon occurs especially in the densely populated and economically developed eastern and southern coastal areas. For example, prefecture-level cities with Type 4 were primarily concentrated in Shandong Peninsula, and prefecture-level cities with type code 16 were primarily concentrated in Fujian Province. The ubiquitous existence of this spatial autocorrelation phenomenon created a prerequisite for the synthesis and geographic analysis of homogeneous regions.

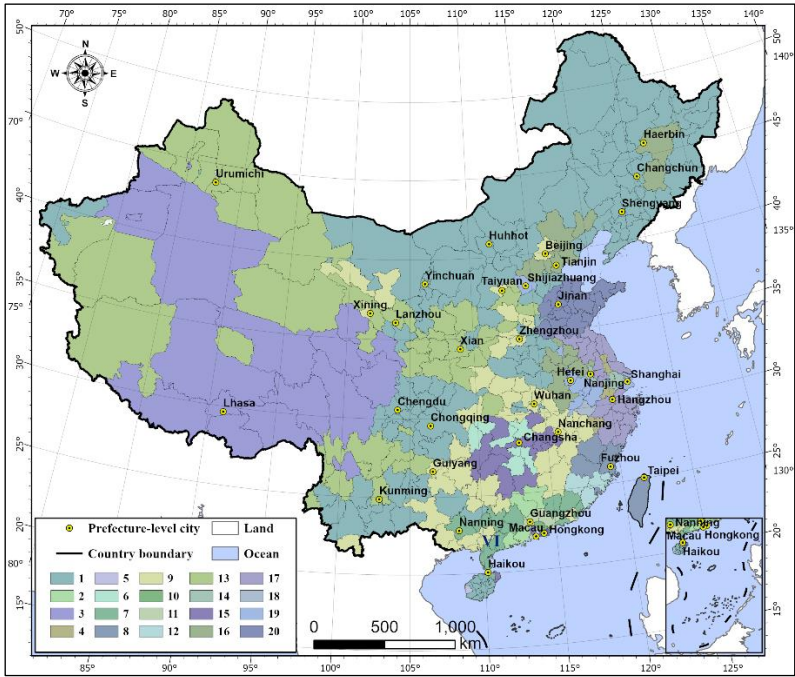


Figure 4. Result of region classification.

The ratio of restaurants to the total number of restaurants in each city served as regionalized variables (Figure 5). The results were analyzed by using 3.2 geographic regionalization method with spatial constraints. The number of partitions corresponding to the maximum value obtained by pseudo-F test was 8. Thus, the threshold 8 was used as the number threshold of geographical partitions. The Roman numeral I-VIII was used to number eight regions. From the overall structure of the regionalization results, the thinner the coastal area was, the closer the inland area was, the coarser the partition size was. For example, the whole northern region was divided into two parts, namely, northwest (VIII) and northeast (I). In turn, the coastline was divided into four regions, namely, II, III, V, and VI. The southwest region was divided into two major regions, namely, areas IV and VII. These cultural regions reflected the structural characteristics of the popularity of the eight cuisines in the same cultural area and the homogeneity of the development of these cuisines in these areas. The spatial differentiation structure of the eight major cuisine cultures were reflected among regions.

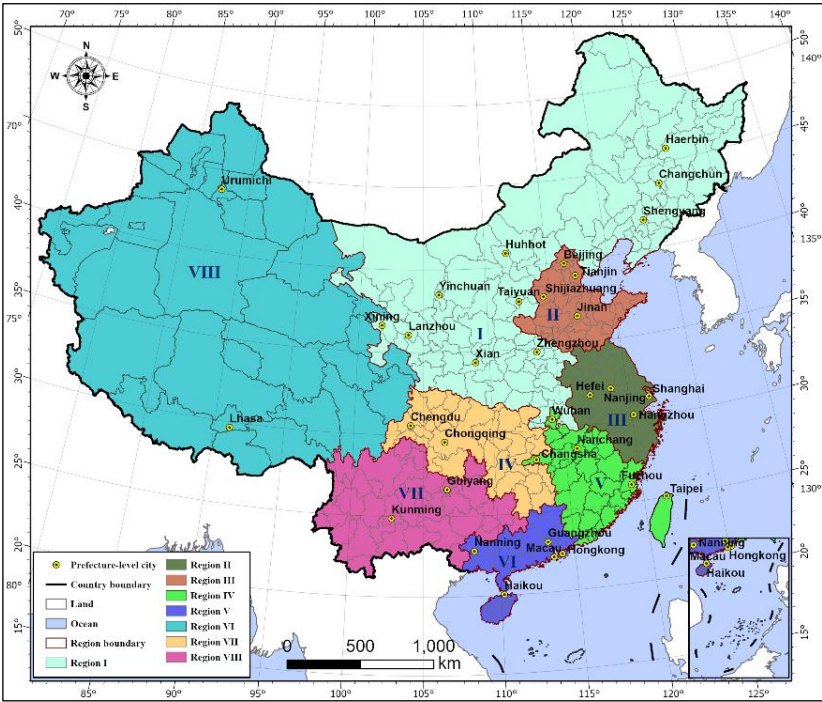


Figure 5. Result of geographical regionalization.

To illustrate the structural characteristics of the proportion these cuisines in each district further, the ratio boxes and their mean lines of each kind of cuisine at prefecture level and city level were drawn, as shown in Figure. 6 and table 2. The proportion of restaurants corresponding to Sichuan, Hunan, Guangdong, and Zhejiang cuisines was relatively stable in more than 300 prefecture-level cities, whereas the proportion of other cuisines in different prefecture-level cities was unstable. Combining with the mean line, the proportion of characteristic restaurants of these cuisines in Region I remained at a medium level. On this basis, people in this area had a neutral preference for eight cuisines and had no particular preference of any cuisine. In Region II, Shandong and Anhui cuisines were the main preferences, whereas the preferences for other cuisines were at a moderate level. Region III had a strong preference for Zhejiang and Anhui cuisines. In addition, Region III had a high preference, which was close to the quartile threshold, for Guangdong cuisine. Other cuisines were on average or low level. Region IV, which basically belonged to Hunan cuisine's catering culture area, had a strong preference for Hunan cuisine. Region V had the strongest inclusiveness to all major cuisines, with a high preference for Fujian, Guangdong, Hunan, and Zhejiang cuisines. The preferred degree of zoning VI to Guangdong cuisine, Hunan cuisine was relatively high, Shandong cuisine, Zhejiang cuisine was relatively low, and the preferred degree to other cuisine was in the middle level. Region VII had a low preference for Anhui cuisine and a medium preference for other cuisines. Region VIII had a high preference for Sichuan, Jiangsu, and Guangdong cuisines, a low preference for Anhui cuisine, and a medium preference for other cuisines. Generally, Anhui, Guangdong, Hunan, and Zhejiang cuisines were popular in many districts. Particularly, Hunan and Sichuan cuisines, which were above moderately preferred in almost every district, were the most popular.

Table 2. Characteristic variables of the box-plot of eight famous Chinese cuisines.

Cuisine categories	minimum	first quartile	median	third quartile	maximum	IQR
Sichuan cuisine	0	0.00867	0.01296	0.02122	0.33333	0.01255
Jiangsu cuisine	0.00042	0.00042	0.00042	0.00042	0.04293	0
Anhui cuisine	0.00088	0.00088	0.00095	0.0012	0.05375	0.00033

Zhejiang cuisine	0.00052	0.00163	0.00228	0.00333	0.05617	0.0017
Fujian cuisine	0.00086	0.00086	0.00086	0.00103	0.02713	0.00018
Hunan cuisine	-0.00033	-0.00025	-0.00013	0.00008	0.00515	0.00033
Guangdong cuisine	-0.00022	-0.00014	-0.00006	0.00009	0.00489	0.00023
Shandong cuisine	0	0	0.00009	0.00022	0.01111	0.00022

2

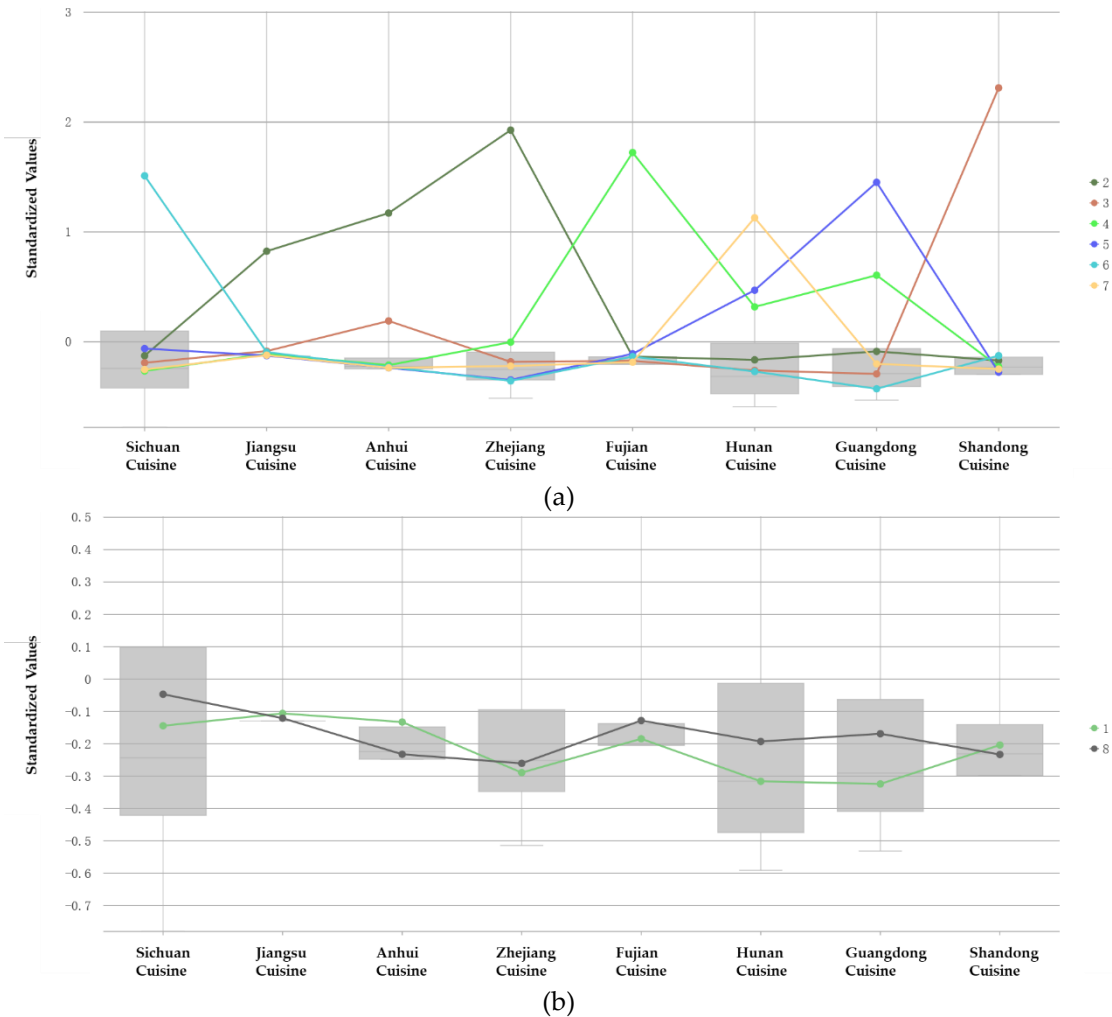


Figure 6. Box-plot of cuisines in different regions. (a)box-plot with high standardized values,(b) box-plot wit low standardized values.

5. Conclusions and future direction

The study of food cultures from the perspective of geography emphasizes the human–environment relationship and the law of regional differentiation reflected by food cultures. Various characteristic cuisines are the carrier of local food cultures and the symbol of regional food cultures. The formation of various cuisines (such as the eight major cuisines in this study) is the objective presentation of regional differences in food cultures. Starting from the research paradigm of cultural geography, this study analyzes the spatial dissemination trend and regional differentiation mode of food culture reflected by the eight most distinctive cuisines in China from the perspectives of quantitative and regional characteristics. The spread of the eight traditional Chinese cuisines is primarily manifested in two aspects, that is, spatial distribution and spatial hierarchy. Regional differentiation is primarily reflected in the formation of food culture regions and the emergence of new cultural boundaries.

From the perspective of the quantitative characteristics of these cuisines, the farther south the regional cuisines are, the greater the spatial diffusion scope of cuisine food culture is in terms of spatial distribution. By contrast, the farther north the cuisines are, the more limited the scope is. Among these cuisines, Sichuan cuisine presents the most vigorous space diffusion, whereas Shandong cuisine presents the weakest space diffusion. For most types of cuisines, urban agglomerations, such as the Yangtze River Delta, the Pearl River Delta, and the Beijing–Tianjin–Hebei Regions, do not form the hotspots of absolute superiority due to their population aggregation, developed economy, and large population mobility. Instead, these agglomerations show a trend of outward diffusion from the origin of each cuisine. In terms of hierarchical structure, the hotspots of different levels of Shandong, Jiangsu, and Anhui cuisines are uneven, and polarization in these hotspots is evident. Other southern cuisines have a relatively balanced hotspots distribution in five grades. To sum up, regardless of spatial distribution or hierarchical structure, the overall trend of north–south differentiation is shown. The former is reflected in the diffusion range, whereas the latter is reflected in the equalization of hotspots at different levels.

This study uses prefecture-level cities as the basic research unit and takes the proportion of each cuisine as the variable to make comprehensive regionalization. The geographical regionalization results of the preference of the eight Chinese cuisines are actually the construction process of food culture regions. Each food culture region not only maintains the spatial continuity but also ensures the similarity of people's preferences for these cuisines in the cultural region. Cultural districts ensure the diversity of their preferences for these cuisines. Cultural regions and their boundaries are in dynamic change, deeply influenced by the daily production and reproduction of cuisine culture. This study constructs new cultural districts using the restaurant data for the understanding of the characteristics of food culture in different food culture regions. This study also provides clues for a deep understanding of the regional preference of cuisine culture structure.

References

1. Floyd, V.D. Food, Culture and Politics. *Journal of the Association for the Study of Food and Society* **1996**, *1*, 5-8, doi:10.2752/152897996786623408.
2. Bissell, D. Social & Cultural Geography at 20 years: looking back, thinking forward. *Social & Cultural Geography* **2019**, *20*, 1-3, doi:10.1080/14649365.2018.1554162.
3. Goodman, M.K. Food geographies I: Relational foodscapes and the busy-ness of being more-than-food. *Progress in Human Geography* **2015**, *40*, 257-266, doi:10.1177/0309132515570192.
4. Winter, M. Geographies of food: agro-food geographies making reconnections. *Progress in Human Geography* **2003**, *27*, 505-513, doi:10.1191/0309132503ph446pr.
5. Chang, R.C.Y.; Kivela, J.; Mak, A.H.N. Food preferences of Chinese tourists. *Annals of Tourism Research* **2010**, *37*, 989-1011, doi:<https://doi.org/10.1016/j.annals.2010.03.007>.
6. Slocum, R. Race in the study of food. *Progress in Human Geography* **2010**, *35*, 303-327, doi:10.1177/0309132510378335.
7. Cwierotka, K.J. EATING THE WORLD: RESTAURANT CULTURE IN EARLY TWENTIETH CENTURY JAPAN. *European Journal of East Asian Studies* **2003**, *2*, 89-116.
8. Gordon-Larsen, P.; Guilkey, D.K.; Popkin, B.M. An economic analysis of community-level fast food prices and individual-level fast food intake: A longitudinal study. *Health & Place* **2011**, *17*, 1235-1241, doi:<https://doi.org/10.1016/j.healthplace.2011.07.011>.
9. Thornton, L.E.; Lamb, K.E.; Ball, K. Fast food restaurant locations according to socioeconomic disadvantage, urban–regional locality, and schools within Victoria, Australia. *SSM - Population Health* **2016**, *2*, 1-9, doi:<https://doi.org/10.1016/j.ssmph.2015.12.001>.
10. Barrett, C.B. Urban bias in price risk: The geography of food price distributions in low-income economies. *The Journal of Development Studies* **1996**, *32*, 830-849, doi:10.1080/00220389608422442.

11. Tellström, R.; Gustafsson, I.-B.; Fjellström, C. Food culture as a political tool – meal construction during the Swedish EU-chairmanship 2001†. *Food Service Technology* **2003**, *3*, 89-96, doi:<https://doi.org/10.1046/j.1471-5740.2003.00069.x>.
12. Johnston, J.; Goodman, M.K. Spectacular Foodscapes. *Food, Culture & Society* **2015**, *18*, 205-222, doi:10.2752/175174415X14180391604369.
13. Wohlers, T.E. The Role of Risk Perception and Political Culture: A Comparative Study of Regulating Genetically Modified Food. In *Risk and Cognition*, Mercantini, J.-M., Faucher, C., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; 10.1007/978-3-662-45704-7_2pp. 21-59.
14. Shortridge, B.G. Apple Stack Cake for Dessert: Appalachian Regional Foods. *Journal of Geography* **2005**, *104*, 65-73, doi:10.1080/00221340508978618.
15. Giovine, R.J.S.o.C. The Shifting Border of Food Perceptions and Cultural Identity in Maghrebi Muslim Migrants. History and the Contemporary Experience of Cultural Mediators in Northern Italy. **2013**, *5*.
16. Bailey, E. Religion, Food, and Eating in North America. *Food, Culture & Society* **2016**, *19*, 195-196, doi:10.1080/15528014.2016.1145404.
17. Mumuni, A.G.; Veeck, A.; Luqmani, M.; Quraeshi, Z.A.; Kamarulzaman, Y. Religious identity, community and religious minorities' search efforts for religiously sanctioned food: The case of halal food in non-Muslim majority markets. *International Journal of Consumer Studies* **2018**, *42*, 586-598, doi:<https://doi.org/10.1111/ijcs.12423>.
18. Opel, A.; Johnston, J.; Wilk, R. Food, Culture and the Environment: Communicating About What We Eat. *Environmental Communication* **2010**, *4*, 251-254, doi:10.1080/17524032.2010.500448.
19. Geoghegan, H.; Leyson, C. On climate change and cultural geography: farming on the Lizard Peninsula, Cornwall, UK. *Climatic Change* **2012**, *113*, 55-66, doi:10.1007/s10584-012-0417-5.
20. Sterckx, R. Food and Environment in Early and Medieval China. *Journal of Historical Geography* **2016**, *53*, 132, doi:<https://doi.org/10.1016/j.jhg.2015.08.010>.
21. Goto, K.; Vue, W.M.; Xiong, T.; Wolff, C. Divergent Perspectives on Food, Culture, and Health among Hmong Mothers with Middle-School Children. *Food, Culture & Society* **2010**, *13*, 181-200, doi:10.2752/175174410X12633934463079.
22. Zaman, S.; Selim, N.; Joarder, T. McDonaldization without a McDonald's. *Food, Culture & Society* **2013**, *16*, 551-568, doi:10.2752/175174413X13758634982010.
23. Parkinson, A. Resourcing the food crisis: geographies of food. *Geography* **2010**, *95*, 88-93, doi:10.1080/00167487.2010.12094289.
24. Jarosz, L. Comparing food security and food sovereignty discourses. *Dialogues in Human Geography* **2014**, *4*, 168-181, doi:10.1177/2043820614537161.
25. Andersson, T.D.; Mossberg, L.; Therkelsen, A. Food and tourism synergies: perspectives on consumption, production and destination development. *Scandinavian Journal of Hospitality and Tourism* **2017**, *17*, 1-8, doi:10.1080/15022250.2016.1275290.
26. Robinson, R.N.S.; Getz, D.; Dolnicar, S. Food tourism subsegments: A data-driven analysis. *International Journal of Tourism Research* **2018**, *20*, 367-377, doi:<https://doi.org/10.1002/jtr.2188>.
27. Cai, X.; Zhu, H.; Liu, C. Characteristics and causes of situational theme restaurant staffs sense of place: A case study of the Taste-Yunnan restaurant in Guangzhou. *Acta Geographica Sinica* **2012**, *67*, 239-252.
28. ZENG Guojun, L.L., LIU Bo, LU Rurui. Reconstructed authenticity in translocal cultural production of food: A case study of Northwest China Cuisine in Guangdong. **2014**, *69*, 1871-1886, doi:10.11821/dlxb201412012.
29. Lankford, P.M. Regionalization: Theory and Alternative Algorithms. *Geographical Analysis* **2010**, *1*, 196-212, doi:10.1111/j.1538-4632.1969.tb00615.x.
30. Niesterowicz, J.; Stepinski, T.F.; Jasiewicz, J. Unsupervised regionalization of the United States into landscape pattern types. *International Journal of Geographical Information Science* **2016**, *30*, 1450-1468, doi:10.1080/13658816.2015.1134796.
31. Qing, L.; Guozhao, P.; Jun, R.; Yanqiu, C.; Peng, F.; Dazhong, L.; Armuzhong; Doumin, H.; Qiaojuan, H.; Yuanzhi, C. GIS-based climatic regionalization of potato late blight in mountain areas of Southwest Sichuan. In Proceedings of Proc.SPIE.

-
32. Zhao, J.; Wang, J.; Lei, Y.; Huang, X. Risk assessment and regionalization of natural disaster of corn in China based on GIS. In Proceedings of 2011 19th International Conference on Geoinformatics, 24-26 June 2011; pp. 1-5.
 33. Assunção, R.M.; Neves, M.C.; Câmara, G.; Da Costa Freitas, C. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science* **2006**, *20*, 797-811, doi:10.1080/13658810600665111.
 34. Gu, Q.; Zhang, H.; Chen, M.; Chen, C. Regionalization Analysis and Mapping for the Source and Sink of Tourist Flows. *ISPRS International Journal of Geo-Information* **2019**, *8*, doi:10.3390/ijgi8070314.
 35. Maulik, U.; Bandyopadhyay, S. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2002**, *24*, 1650-1654, doi:10.1109/TPAMI.2002.1114856.