Optimized agricultural site selection based on geographic similarity

Zilong Qin [1], Yi Chen [1], Zongyao Sha [1,2]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, Hubei, China – zilongqin@whu.edu.cn; 1124549946@qq.com
[2] Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical area of South China, Ministry of Natural Resources, Guangzhou, 510631, Guangdong, China – zongyaosha@whu.edu.cn

Abstract: In this study, we proposed an optimal agricultural site selection method based on the third law of geography and geographic similarity, and took the vegetable moss growing area in Hongshan District, Wuhan, China, as an example. The geography of the source region (Hongshan District) and the target region (Hubei Province) was examined and the temperature, precipitation, topographical factors and soil were selected as the environmental covariates. The specific design of the experiment included the steps of creating sample points and points to be inferred, classifying the environmental factors, calculating and ranking geographic environmental similarity, and modifying in missing values. The results showed that the areas with high similarity between the source and target regions were mainly concentrated in the southeastern part of Hubei, i.e., the area suitable for growing Hongshan vegetable moss, while the western part of Hubei was mostly a low-similarity area. The methodology of this study provides an important theoretical reference for crop site selection and is of practical application.

Keywords: Optimized Site Selection; Third Law of Geography; Geographic Similarity; Agricultural Site Selection.

1. Introduction

The Third Law of Geography indicates that "the more similar the geographical environment, the more similar the geographical characteristics," also known as the law of geographical similarity. This law is characterized by its universality, independence, and applicability, focusing on the similarity of geographical environmental factors. Zhu et al. (2022) proposed that spatial prediction based on geographical similarity does not require sample size or spatial distribution, allowing the representativeness of a single sample in spatial prediction, making it suitable for spatial predictions in large and complex geographical areas. Site selection is an important step in agricultural production. According to the characteristics of crops, selecting areas with fertile soil, suitable climate, and gentle terrain is an important condition for successful crop cultivation. It can be seen that determining site selection factors based on research objectives and selecting appropriate indicators for quantification is an important prerequisite for site selection based on the Third Law of Geography. Meanwhile, Geographical information system (GIS) technology also provides support for data analysis, calculation of geographical environment similarity, and visualization (Andresen et al., 2009).

In recent years, GIS technology has developed rapidly and has been widely applied in many fields, among which the application in agriculture is also one of the popular fields. With the technical support of GIS, traditional agriculture is gradually transformed into modern agriculture, which plays a role in the decision-making management of arable land resources, agricultural siting and planning, sustainable development of agriculture, agro-ecological monitoring and so on. China's application of GIS technology for site selection and planning in the field of agriculture is relatively extensive, and GIS provides processing and analyzing functions to deal with spatial data in various formats, such as vector data and raster data, etc., and at the same time, it has a strong ability to deal with large-scale datasets. Some researches combined GIS technology to establish Ding'an County cropland resource database, and calculated cropland fertility index based on fuzzy mathematical theory using various methods for evaluation. There are also studies applying GIS technology to vineyard site selection, combining with hierarchical analysis method to carry out scientific evaluation, and dividing vineyard site selection into five levels to improve the efficiency of site selection.

Vegetable moss is a specialty of Wuhan, named for its best quality produced in Hongshan District, with a cultivation history of more than 1700 years. It is rich in nutrition, rich in various vitamins, and has the value of promotion. However, with the acceleration of urbanization, at the beginning of this century, the planting area of authentic vegetable moss has been sharply reduced, with a production of less than 60,000 pounds, and due to poor management, it is on the verge of being lost. Optimizing the cultivation area of vegetable moss in Hongshan District of Whuhan is conducive to the recovery and promotion of vegetable moss in Hongshan District, and also helps to better build agricultural brands and create characteristic industries.

In conclusion, we take vegetable moss in Hongshan District as an example, according to its cultivation conditions, determine three geographical environmental factors of climate, terrain, and soil, standardize the data and carry out similarity calculation, comprehensively obtain the similarity ranking of the source area and the target area, and recommend the most similar target area, to explore the application of optimization agricultural site selection based on geographical environment similarity. In addition, with the development of agricultural modernization and the rise of characteristic industries, the method of optimization site selection based on geographical environment similarity has a wide application prospect. It can not only be applied in agricultural site selection, but also play an important role in other fields such as urban planning and resource development. Therefore, the significance of this study is not only to solve the problems in the development of specific agricultural industries, but also to provide practical cases and method guidance for the application of geographical environment similarity theory.

2. Methods and Experimental Design

2.1 Study area and data sources

Tianxingzhou is one of the areas of Hongshan vegetable moss, located in Hongshan District, Wuhan City, Hubei Province. The aim of this study is to find areas in Hubei Province that are suitable for growing Hongshan vegetable moss, i.e. areas that are most similar to the environment of Tianxingzhou.

In order to quantify and describe the conditions associated with the growing environment of Hongshan vegetable moss, climate, topographic factors, and soil conditions (McBratney et al., 2003) (i.e., soil PH, downloaded from http://loess.geodata.cn) are selected as environmental covariates in this study. Climate factors include multi-year average temperature and multi-year average precipitation during 2012-2022, which spatial resolution is 1 km (downloaded from http://www.geodata.cn). Climate data are synthesized by annual accumulation based on the 1km resolution month-by-month mean temperature dataset of China from 1901 to 2022 and the 1km resolution month-by-month precipitation dataset of China from 1901 to 2022, respectively, and the time range of this paper is 2012-2022, and then the multi-year mean temperature and multi-year mean precipitation data are obtained by using the ArcGIS software based on the annual meteorological data, and extracted using the vector data of Hubei Province. precipitation data, and extracted using Hubei Province vector data.

Topographic factors include digital elevation model (DEM), slope and aspect, slope and aspect are computed with DEM using ArcGIS 10.6, which spatial resolution is 30m (downloaded from http://www.gscloud.cn).

calculate the similarity of different types of environmental factors separately.

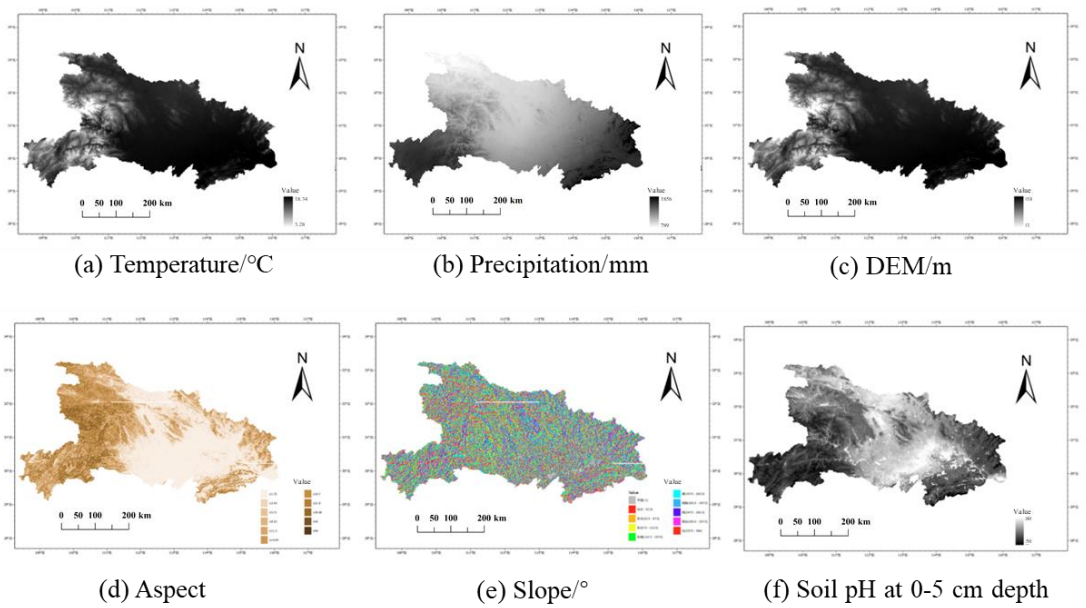
The spatial resolution of soil conditions data is 250m. Soil conditions are characterized using pH, and a 250m resolution soil pH dataset for Hubei Province was downloaded from the National Earth System Science Data Center, including different depths, namely 0-5cm, 5-15cm, 15-30cm, 30-60cm, 60-100cm, and 100-200cm.Based on the object of the study, as well as the objectives of the study, it was decided to select the data for the topsoil layer, the i.e. 0-5cm and 5-15cm data. Finally, all data are resampled using ArcGIS to a spatial resolution of 250m (Figure. 1).

2.2 Methods and experimental design

The optimal site selection method based on environmental similarity mainly includes three steps: 1) classification of environmental factors, 2) normalization of environmental factor data, and 3) calculation and ranking of geographic environment similarity. The specific steps are as follows:

1) Classification of environmental factors

We divide the environmental factors into numerical factors and type factors according to the data type, and then



(a) Temperature/°C (b) Precipitation/mm (c) DEM/m

(d) Aspect (e) Slope/° (f) Soil pH at 0-5 cm depth

Figure 1. Environmental factors of modelling.

Specifically, in this experiment the land use type is classified as a type factor, and the climate, topography and soil factors are classified as numerical type factors.

2) Normalization of environmental factor data

Differences in data sources and ranges of data values can lead to disparities in the magnitude of the data, and the standardization of data solves this problem. Data standardization is most typically the normalization of data, mapping the data to the interval [0,1]. The min-max normalization method is a way of linear transformation, also called deviation normalization (Liao et al., 2019). The transformation function is as follows:

$$y_i = \frac{x_i - \min\limits_{1 \le j \le n} x_j}{\max\limits_{1 \le j \le n} x_j - \min\limits_{1 \le j \le n} x_j} \quad (1)$$

where x_i denotes raw data; y_i denotes normalized data; max and min denote the maximum and minimum values of the sample, respectively; and *n* denotes the number of samples.

3) Calculation and ranking of geographic environment similarity

The calculation of geographic environment similarity refers to the calculation of the similarity between the sample point (source point) and the point to be inferred (target point) (Zhu et al., 2022). The formula for calculating the geographic similarity between the source point *i* and the target point *j* is as follows:

$$E(e_{vi}, e_{vj}) = \begin{cases} 0, & e_v \text{ is nominal and } e_{vi} \neq e_{vj} \\ 1, & e_v \text{ is nominal and } e_{vi} = e_{vj} \\ exp\left(-\frac{(e_{vi}-e_{vj})^2}{2\left(SD_{e_v} \times SD_{e_v}/SD_{e_{vj}}\right)^2}\right), & e_v \text{ is numerical} \end{cases} \quad (2)$$

where $E(e_{vi}, e_{vj})$ denotes the geographic environment similarity between source point *i* and target point *j* for *e*-th environmental factor; e_{vi} and e_{vj} denote the values of source point *i* and target point *j* for *e*-th environmental factor, respectively; SD_{e_v} denotes the standard deviation of *e*-th environmental factor; $SD_{e_{vj}}$ denote the root-mean-square error of the *v*-th environmental variable at source point *i* with respect to target point *j*.

$$SD_{e_{vj}} = \frac{\sqrt{\sum_{i=1}^{k}(e_{vi}-e_{vj})^2}}{k} \quad (3)$$

where *k* denotes the number of target points.

$$E_{ij} = \begin{cases} 0, & e_v \text{ is nominal and } e_{vi} \neq e_{vj} \\ 1, & e_v \text{ is nominal and } e_{vi} = e_{vj} \\ exp\left(-\frac{(e_{vi}-e_{vj})^2}{2\left(SD_{e_v} \times SD_{e_v}/SD_{e_{vj}}\right)^2}\right), & e_v \text{ is numerical} \end{cases} \quad (4)$$

Then, the integrated similarity can be calculated based on the similarity between the source point *i* and the target point *j* for each environmental factor.

$$S_{ij} = \frac{\sum_{e=1}^{m} w_e \times E(e_{vi}, e_{vj})}{\sum_{e=1}^{m} w_e} \quad (5)$$

where S_{ij} denotes the integrated environmental similarity of environmental factors; w_e denotes the weight of *e*-th environmental factor; *m* denotes the number of environmental factors.

Finally, the results of the environmental similarity calculated between the source points and target points are ranked to obtain the results of the agricultural site selection, i.e., the higher the similarity, the more suitable it is to be planted with the crop.

This section focuses on the design of an optimal site selection method based on the similarity of geographic environment. First, the geographic location, climatic characteristics, and the relevant conditions of Tianxingzhou in Hubei Province were described in detail by providing an overview of the study area. Then, for the cultivation area speculation of Hongshan vegetable moss, multi-year average temperature, multi-year average precipitation, topographic factors and soil conditions were selected as environmental covariates, and data preparation was carried out, including data sources and processing methods. In the experimental design section, the inferred method based on geographic similarity is elaborated in detail, including the steps of point set creation, data extraction, standardization, and similarity calculation, as well as the process of how to fill in the missing values and generate heat maps. In the section of findings and analysis, the influence of three key geo-environmental factors, namely climate, soil and topography, on the cultivation of vegetable moss in Hongshan was explored in depth by analyzing the one-way similarity degree of these factors. The geo-environmental similarity between Hongshan vegetable moss cultivation areas and sample sites was demonstrated through heat maps, and the attribute values of highly similar and dissimilar areas were analyzed and compared. Finally, the effectiveness of spatial speculation based on geo-environmental similarity with respect to the cultivation conditions of Hongshan vegetable moss was emphasized through the comprehensive analysis results, and provided an important basis for further research and decision-making.

3. Results

3.1 Spatial distribution of similarity for environmental factors

The results of similarity for environmental factors are shown in Figure. 2. The regions with different climate similarity are mostly distributed in blocks, showing spatial clustering, and there are large differences in climate conditions between different regions, so it can be hypothesized that climate have a greater impact on the cultivation of Hongshan vegetable moss, and the weights should be set larger in the comprehensive similarity calculation. Soil pH within the target area is generally similar, with only a few areas showing a high degree of difference, and almost all showing a high degree of similarity. It suggests that soil conditions in Hubei Province are relatively stable and may be more suitable for the promotion of the same crop varieties. The results of the topographic similarity suggest that the topographic characteristics of the eastern Hubei area may be closer to the source region, including the flatter terrain, which may be more suitable for the promotion of crop varieties from the source region.

Climatic factors are one of the important influences on crop growth. The regions with different degrees of similarity in the results of climate similarity are mostly distributed in blocks, showing spatial aggregation, and the climatic conditions between different regions differ greatly, which can be inferred that the climatic conditions have a greater influence on the cultivation of Hongshan vegetable moss, and the weights should be set larger in the calculation of the comprehensive degree of similarity. Southeast E region is generally more similar to the source area in terms of climate, while West E region shows obvious dissimilarity. The climatic conditions in Southeast E may be closer to the source area, including factors such as temperature, precipitation and humidity. This similarity may make Southeast E region more suitable for promoting crop varieties from the source region. In contrast, the climatic conditions in West E region are more different from those in the source region, and it may be necessary to adjust to the local climatic characteristics or select other suitable crop varieties.

Soil factors also have an important influence on the growth and development of crops. Soil acidity and alkalinity are generally similar in Hubei Province, with only a few regions showing large differences, and almost all showing a high degree of similarity. This suggests that soil conditions in Hubei Province are relatively stable and may be more suitable for promoting the same crop varieties. The high degree of soil similarity provides favourable conditions for crop location and planting and reduces the impact of soil factors on agricultural production.

The results of the degree of similarity in climate, soil, and topography suggest that, on the whole, the eastern part of Hubei is more suitable for the promotion of crop varieties from the source area, while the western part of Hubei may need to be adapted to the local environmental characteristics or to find other suitable crop varieties.

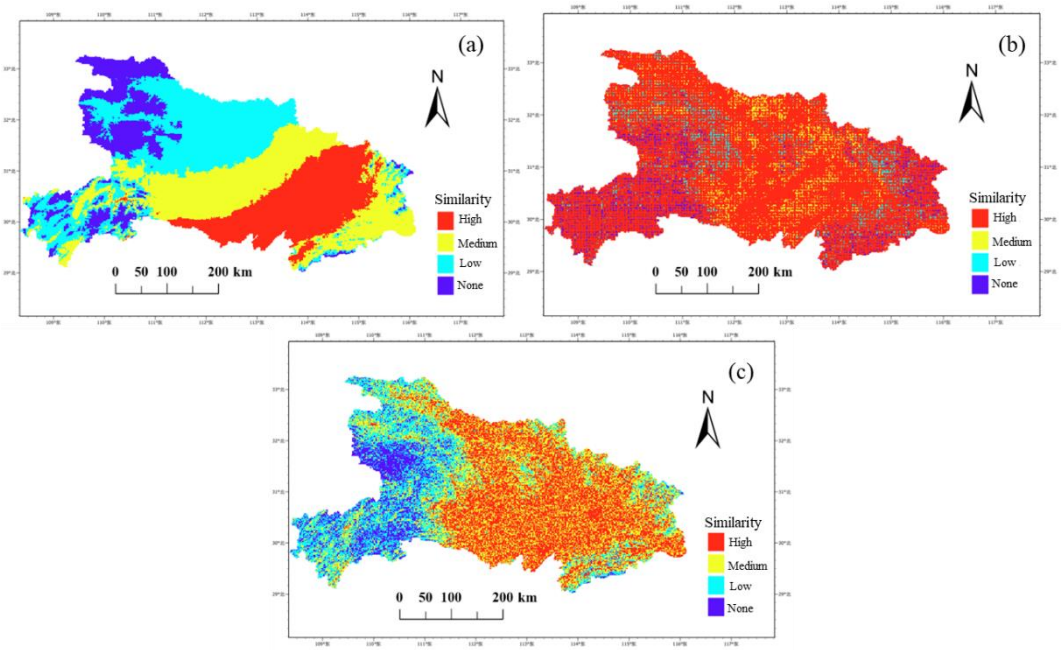


Figure 2. Spatial distribution of similarity for (a) climate, (b) soil and (c) topography.

Topographic factors are also one of the important factors in agricultural production and have an impact on soil conditions, such as soil nutrients. The results of the study on the degree of topographic similarity showed that the similarity considering only topographic factors showed that the west E region was not similar, while the east E region was generally similar, and the percentage of similar areas was higher. This suggests that the topographic features of East E region may be closer to the source region, including the feature of flatter terrain, which may be more suitable for promoting crop varieties in the source region. On the other hand, the topographic characteristics of West E region are significantly different from the source region, and it may be necessary to adjust to the local topographic conditions or select other suitable crop varieties.

The East E region is generally similar in terms of topography and climatic factors, while soil factors have less influence on the overall similarity; in contrast, the West E region is quite different from the source area in terms of topography and climate, but soil factors are more similar to the source area. Further analysis found that the topography of East E region is relatively flat, the climate is humid, and the soil conditions are relatively consistent, which is more similar to the cultivation environment of Hongshan vegetable moss. While the west E region is mostly mountainous and has a drier climate, but the soil conditions are more similar to the source area, which may be related to the soil's heat preservation and moisturizing properties.

3.2 Spatial distribution of integrated similarity of environmental factors

In this study, the difference method was applied to measure the similarity between environmental variables, where the smaller the difference, the smaller the distance, representing a higher degree of similarity. According to the results of similarity calculation between the sample points and the points to be inferred, this study carries out further categorization, and the data are classified into four categories according to the similarity value based on the natural discontinuities, which are the four categories of highly similar, relatively similar, not too similar and not similar. And the attributes of highly similar and dissimilar regions were statistically analyzed, and the average value was taken as the characteristic value to speculate the geo-environmental attributes of the two regions.

The results of integrated similarity of environmental factors are shown in Figure. 3 and Figure. 4. The red area in Figure. 3 represents a high degree of environmental similarity between the target region and the source region of Hongshan vegetable moss, i.e., the red area is suitable for cultivation of Hongshan vegetable moss. The blue area indicates a low degree of environmental similarity and a large difference in geographic environmental conditions between source region and target region.

In the region of high similarity, the average multi-year average precipitation is 1152.6 mm, the average multi-year average temperature is 17.5 ℃, the average slope direction is 173.6 °, the average elevation is 70.2 m, the average slope is 1.54 °, and the average pH of the topsoil layer is 67.5. From the data, it can be deduced that the geography of the region of high similarity is characterized by a relatively flat topography, low elevation, and it may be located in the plains or low-lying areas, and the slope is small. plains or low-lying areas, with smaller slopes, relatively gentle terrain, neutral pH, more adequate precipitation, and possibly a humid and warm climate, which is consistent with the growing conditions of Hongshan vegetable moss.

In contrast, in the dissimilar region, the average multi-year mean precipitation is 1284.1 mm, the average multi-year mean temperature is 8.0°C, the average slope direction is 204.3°, the average elevation is 2056.5 m, the average slope is 19.2°, and the average acidity is 55.9. From the data, it can be concluded that the dissimilar region is located in the high altitude area, although the precipitation is also more abundant, but the average elevation is larger. The high altitude areas usually have lower temperatures, as well as larger slopes, which may affect soil retention and water

infiltration, and neutral to acidic soil pH, characteristics that differ significantly from the growing conditions required for Hongshan vegetable moss. Based on the above statistics it is possible to roughly determine the geography of highly similar areas (areas suitable for Hongshan vegetable moss cultivation). In the generated heat map, the change of color shows the change of the degree of similarity, where red represents high similarity and blue represents low similarity. In relation to the study objectives, the red area represents a high degree of similarity in geographic features between the area and Tianxingzhou, one of the origins of Hongshan cauliflower moss, i.e., a high degree of similarity in geo-environmental conditions, while the blue area indicates a low degree of similarity in geographic features and a high degree of difference in geo-environmental conditions.

The areas with high similarity are mainly concentrated in the southeast of E, while the west of E presents more blue areas with low similarity. This coincides with the geographical environment and climatic conditions of the region. The relatively flat terrain and humid climate in the south-eastern part of E, may be more similar to the cultivation environment of Hongshan vegetable moss, and at the same time, the geographic environment conditions derived from some of the similarity points fit, so it scored higher in the similarity calculation; while the western part of E is mountainous and the climatic conditions may be drier, which is more different from the cultivation environment of Hongshan vegetable moss, and thus scored lower in the similarity calculation. According to the results of the study, it can be inferred that the climatic conditions of the red region are similar to those of the sample area, including factors such as temperature, precipitation, and humidity; it has similar characteristics in terms of topography, mostly mountainous, hilly, or plains; and the soil characteristics are also relatively similar, such as soil type, soil acidity and alkalinity, and soil water content, which have similar growing conditions. At the same time, the impacts of human activities, such as fertilization and irrigation, are also more similar due to the similar patterns of human activities and agricultural production methods in neighboring regions.

In the blue low-similarity region, on the contrary, the geographic environment characteristics between this region and the sample sites are more different, in many aspects, the topography of the blue region and the source region may be different, such as high mountainous areas and plains; in terms of climate, the temperature of this region is significantly lower than the average temperature of the source region, and the amount of precipitation may be unevenly distributed; in terms of soil conditions, the differences in characteristics may be more obvious, in terms of soil type and texture, as well as pH, which leads to poorer growing conditions. The differences in soil conditions may be more pronounced in terms of soil type, texture, and pH, leading to poorer growing conditions. At the same time, the blue zone is at a higher altitude, mostly mountainous, with a lower level of agricultural development, and has a different pattern of human activities and impacts compared to the source region. The results of the study show that the red areas (high similarity areas) are mainly concentrated in the south-eastern part of E. E., while the western part of E. E. is mostly blue areas (low similarity areas).

The result coincides with the geographical environment and climatic conditions of the region. The relatively flat terrain and humid climate in the southeast of E is suitable for the growth of Hongshan vegetable moss, which corresponds to its higher score in the similarity calculation; while the west of E is mostly characterized by high elevation, mountainous terrain and low temperature, which is different from the growing environment of Hongshan vegetable moss and unsuitable for the growth of Hongshan vegetable moss, which corresponds to the result of the low score in its similarity calculation. Moreover, according to the histogram statistics, the proportion of highly similar and relatively similar regions is much more than that of less similar and dissimilar regions, which shows that most of the regions in Hubei Province are suitable for the cultivation of Hongshan vegetable moss, which makes it possible to build a specialty industry of Hongshan vegetable moss in Hubei Province.

The similarity within the cropland of Hubei Province is further extracted based on the results in Figure. 3, as shown in Figure. 4, the areas of higher environmental similarity are mainly concentrated in the southeast of Hubei, while the west of Hubei presents more blue areas of low similarity. The site selection method based on environmental similarity is an important reference for evaluating the site selection of Hongshan vegetable moss planting areas and developing crop-planting policies.

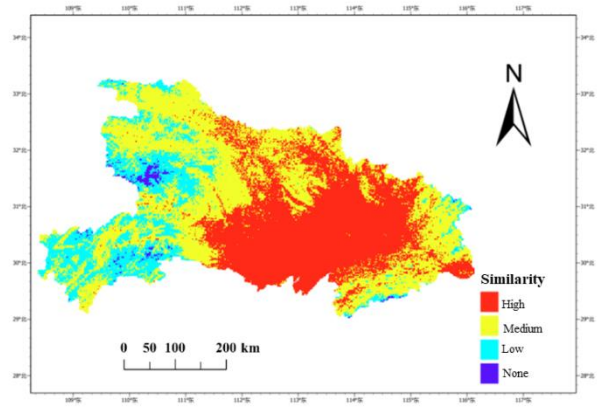


Fig. 3. Spatial distribution of integrated similarity of environmental factors.

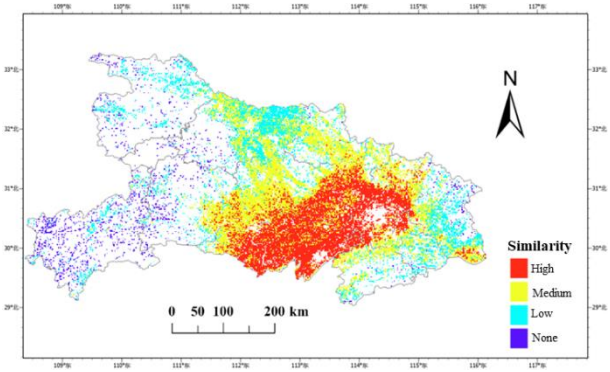


Fig. 4. Integrated similarity of environmental factors within cropland.

4. Conclusion

In this study, we first investigated and analyzed the origin of Hongshan vegetable moss and determined Tianxingzhou as the source area, then we deeply analyzed the geographic environment characteristics of Tianxingzhou, and at the same time, we selected the corresponding geographic environment factors in combination with the growth environment of Hongshan vegetable moss, and we used a variety of environmental variables to quantitatively describe the combination of geographic elements of the target variables. In the experimental part, the data were pre-processed with format conversion and data cleaning, and after in-depth discussion on the calculation method of geographic environment similarity, a suitable function was selected to calculate the similarity, and the final result was converted into a heat map, which visualized the geographic environment similarity in different areas, and combined with the comparative analysis of the attribute values of the high similarity

area and the non-similarity area, the method of spatial speculation based on the geographic environment similarity was verified. The effectiveness of the spatial speculation method based on geographic environment similarity is verified.

Geographic similarity has an important impact on agricultural site selection. According to the results of similarity ranking, the south-eastern part of E is more suitable for planting Hongshan vegetable moss, while the western part of E is the opposite, and the southern part of E is more suitable for the growth of Hongshan vegetable moss than the northern part of E. This is closely related to the topography, climate and soil conditions of the region. By measuring the similarity between the selection method based on the similarity of the geographic environment. The results of the study provide an important reference for evaluating the site selection of Hongshan vegetable moss planting area, meanwhile, it is of guiding significance for the policy development of crop planting, which helps to optimize the allocation of agricultural resources and improve the efficiency of agricultural production.

environmental variables, the results showed that the geographic environment factors had a significant influence on the planting areas, and the areas with higher similarity had higher similarity to the geographic characteristics of Tianxingzhou, one of the origins of Hongshan vegetable moss, while the geographic environment characteristics of the areas with lower similarity differed greatly from that of the planting environments of Hongshan vegetable moss.

In this study, we analyzed the geographic environment characteristics of the cultivation area of vegetable moss in Hongshan, Hubei Province, and proposed an optimal site

References

Andresen, M.A. Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach. Applied Geography 2009, 29, 333-345.

Liao, W.; Hou, D.; Jiang, W. An approach for a spatial data attribute similarity measure based on granular computing closeness. Applied Sciences 2019, 9, 2628.

McBratney, A.B.; Santos, M.M.; Minasny, B. On digital soil mapping. Geoderma 2003, 117, 3-52.

Zhu, A.X.; Miao, Y.; Liu, J.; Bai, S.; Zeng, C.; Ma, T.; Hong, H. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. Catena 2019, 183, doi:10.1016/j.catena.2019.104188.

Zhu, A.X.; Turner, M. How is the Third Law of Geography different? Annals of GIS 2022, 28, 57-67, doi:10.1080/19475683.2022.2026467.