Prediction and Spatial Correlation Analysis of Standard Grain Theoretical Capacity for Land Improvement Based on GIS Technology

Hui Kong†, 2, 3, 4, a
1Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China.
2Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710021, China.
3Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710021, China.
4Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710021, China.

a283125485@qq.com

Abstract. China has a large population and a limited area of arable land. In recent years, society has developed rapidly, the area of arable land has gradually decreased, and the issue of food has become a highly regarded issue. With the promulgation of a national policy to adhere to the red line of 1.8 billion acres of arable land, the rational management and improvement of unused and degraded lands will be improved, and the problems will be solved in terms of both arable land area and arable land quality. To sum up, the issue of grain is the issue of yield. The fluctuation of grain yield is the most critical factor. Therefore, the use of GIS technology to analyze and monitor the change in grain yield from cultivated land resources has become a new management of cultivated land resources. And powerful means of optimal use. In this paper, the correlation coefficients of spatial autocorrelation are used to characterize, through the statistical yearbook data processing, the theoretical capacity of Shaanxi Province standard grain unit theoretical capacity is predicted and spatial autocorrelation analysis is performed.

1. Introduction

In GIS applications, the common problem faced by various fields is how to provide analysis techniques to mine data in the increasingly rich data environment, extract geographic knowledge to promote management innovation and scientific decision-making. Spatial analysis should be solved and answered. Question [1]. The development of spatial data analysis methods is of great significance both to the development of geography and the in-depth application of GIS technology [2]. The method and principle of spatial autocorrelation in GIS spatial data analysis technology can effectively show the position of the spatial unit and the relationship between it and other spatial units. Based on the theoretical capacity accounting of agricultural land in Shaanxi Province, a spatial autocorrelation...
analysis of standard grain yields can predict the theoretical production capacity of agricultural land resources to understand the agricultural land grain production and its spatial distribution characteristics in different regions of Shaanxi Province. [4]. It can reflect the agricultural science and technology level and agricultural science and technology utilization potential of the study area to a certain extent. In order to take reasonable and effective measures to effectively guarantee the safety of agricultural land resources, scientifically develop agricultural land resources, and promote the refined management of agricultural land resources to provide technical support and decision-making basis. It reflects the distribution of regional output, and provides a basis for macro-control of strategic land resources [5].

2. Principles and methods of agricultural land capacity accounting

The calculation of the theoretical production capacity of agricultural land is based on indexes such as the natural quality of agricultural land [6]. The basic idea is as follows: First, establish a linear equation for the theoretical yield of agricultural land and the natural quality of agricultural land through regression analysis. Substitute the natural quality index of the agricultural land for the graded unit into the linear equation to obtain the agricultural land for the graded unit. Theoretical yield [7]; Secondly, the theoretical yield of the standard grain of the crop is obtained by multiplying the theoretical yield of the grading unit by the yield ratio coefficient of the crop; finally, the theoretical yield of the standard grain of the county is determined in combination with the specific farming system of the county.

(1) Technical route

The calculation of the theoretical agricultural land capacity of each unit is based on the theoretical yield of the standard agricultural land and the sown area data of the corresponding graded unit, and multiplying them to calculate theoretical agricultural land capacity of each graded unit. The technical route is shown in Figure 1.

Figure 1. Technical Roadmap for Theoretical Capacity Accounting for Agricultural Land
(2) Theoretical yield accounting

Similar to the thinking of theoretical capacity accounting, the calculation of theoretical yields of agricultural land is based on the establishment of a linear regression model of the designated theoretical yields of crops and natural quality \[8\]. The linear regression relationship model of the theoretical yield point of the crop and the natural quality and other indices corresponding to the grading unit is calculated as follows:

\[
M_k = aN_k + b
\]  

(1)

In the formula, \(M_k\) is the theoretical yield point value of the k-th grading unit; \(N_k\) is the index of the natural quality of the k-th grading unit; \(a\) and \(b\) are linear regression coefficients. According to formula (1), the theoretical yield \(M_k\) of each graded unit can be calculated.

(3) Graduation unit theoretical capacity accounting

The theoretical capacity calculation of each grading unit of agricultural land is based on the theoretical yield calculation. The theoretical yield of each grading unit is multiplied by the planting area of the corresponding grading unit to obtain the theoretical capacity of the grading unit. Calculated as follows:

\[
W_k = M_k \times S_k
\]  

(2)

In the formula, \(W_k\) is the theoretical capacity of the k-th class unit; \(M_k\) is the theoretical yield of the k-th class unit; \(S_k\) is the planting area of the k-th class unit.

3. Inter-autocorrelation measure

According to the spatial proximity matrix \(W\) and the mathematical form describing the difference in the values of the neighboring attributes, a variety of spatial autocorrelation measures can be proposed \[9\]. The appropriate spatial autocorrelation statistics are Moran’s I and Geary’s C and generalized G statistics. These measures are regarded as “global measures of spatial autocorrelation or spatial association”. In order to describe the spatial autocorrelation under such heterogeneity, we must be able to detect the spatial autocorrelation measurement method on a local scale. LISA (spatial connection localization index) and local G statistics are designed for this purpose \[10\].

3.1. Global space autocorrelation

3.1.1. Moran’s I Statistics. Suppose there are n area units in the study area, the observed value on the i-th unit is \(y_i\), and the mean value of the observed variables in the n units is recorded as. Then Moran’s I is defined as:

\[
I = \frac{n}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (y_i - \bar{y})(y_j - \bar{y})
\]  

(3)

To simplify the formula, it can also be written in the form of a matrix:

\[
I = \frac{n}{\sum_{i} \sum_{j} W_{ij}} \frac{Y^T W Y}{Y^T Y}
\]  

(4)
In the formula, \( W \) is a matrix; and, is a column vector formed.

The range of Moran's I index is \((-1,1)\). If the spatial process is uncorrelated, the expectation of I is close to 0. When I takes a negative value, it generally indicates a negative correlation, and when I takes a positive value, it indicates a positive autocorrelation.

Suppose that the observed value of the random variable \( Y \) comes from the normal distribution and the sum is spatially dependent, then the distribution of I obtained by sampling is an approximate normal distribution. A statistic \( Z \) that obeys the normal distribution can be constructed to test the spatial autocorrelation Significance. Statistics are expressed as

\[
z = \frac{(I - E(I))}{\sqrt{\text{Var}(I)}}
\]

(5)

### 3.1.2. Generalized G statistic

For those who do not have the ability to recognize different types of spatial aggregation patterns, the generalized G statistic (A.Getis, JKOrd, 1992) is proposed. Its advantage as a new statistic pattern is that it can detect "hot" in the research area. "spots" or "cold spots", that is, hot and cold spots.

Generalized statistics also take the form of cross products. Cross product is also often used as a measure of spatial association. The generalized G statistic is generally defined as:

\[
G(d) = \sum_{i} \sum_{j} W_{ij} (d) x_i x_j / \sum_{i} \sum_{j} x_i x_j
\]

(6)

Also, calculate its normalized value:

\[
Z = \frac{G(d) - E(G)}{\sqrt{\text{Var}}}
\]

(7)

### 3.2. Local spatial autocorrelation

It is reasonable to assume that there are different spatial autocorrelation values in the study area, or that the values of the spatial autocorrelation in some areas may be high, and the values in other areas may be low, or even in the study area. A positive spatial autocorrelation was found in one part of the image and a negative spatial autocorrelation was found in other areas. The reason for this phenomenon is the existence of spatial heterogeneity. In order to obtain a measure of spatial heterogeneity, we must rely on other measures. Appropriate modifications to the global measure statistics (I, C, and generalized G) can be used to detect spatial autocorrelation on a local scale.

#### 3.2.1. Local index of spatial connection \( \text{LISA} \)

\( \text{LISA} \) is a localized version related to I and C (L. Anselin, 1995). In order to explain the level of spatial autocorrelation on a local scale, it is necessary to derive the value of the spatial autocorrelation on any area unit. For area units, the local Moran ’s I statistic is defined as.

\[
I_i = z_i \sum_j W_{ij} z_j
\]

(8)

In the formula, \( Q \) is the normalized variable for the mean and standard deviation respectively, and \( Q \) is the standard deviation of. Because there may be occasionally lower or higher individual values, the obtained local Moran ’s I cannot fully explain the problem, and it needs to be further explained with standardized statistics. The standardized value is calculated using the following formula:
3.2.2. Geary's C statistic. The local transformation formula is transformed into:

\[ C(i) = c_i \sum_j W_{ij} (z_i - z_j)^2 \]  

(10)

3.2.3. Local G statistic. Local space autocorrelation statistics and localized versions of generalized G statistics are defined as:

\[ G_i(d) = \frac{1}{\sum_i x_i} \sum_{i \neq j} W_{ij} dx_j \]  

(11)

The standardized statistics are:

\[ Z(G_i) = \frac{G_i - E[G_i]}{\sqrt{Var[G_i]}} \]  

(12)

4. Experimental treatment

4.1. Global space autocorrelation analysis

Global space autocorrelation statistics include Moran’s I, Geary’s C, and generalized statistics G. Moran’s I and Geary’s C have the same characteristics, but Moran’s I’s numerical distribution characteristics are more ideal than Geary’s C. This paper mainly uses Moran's I and generalized statistics G for global spatial autocorrelation analysis.

4.1.1. Moran’s I Statistical analysis. A variety of weight matrices are applicable to Moran’s I. Here, the center-of-gravity distance weight matrix is used to calculate the Moran’s I statistic of standard grain theoretical yield. According to formulas (3) and (4), the Moran’s I expectation, variance, and statistic Z are calculated respectively to analyze the global spatial autocorrelation of three yields.

| index         | Moran's I | E       | V      | Z       |
|---------------|-----------|---------|--------|---------|
| Theoretical yield | 0.5496    | -0.0101 | 0.0015 | 14.4322 |

The results of the global spatial autocorrelation analysis of the standard grain theoretical yields are shown in Table 1. The Moran’s I values of the standard grain theoretical yields calculated by this method are all greater than 0 and less than 1, within the allowable range, and are greater than expected. The value indicates that there is a certain positive spatial autocorrelation at the global level. However, it is impossible to determine whether the spatial autocorrelation is significant. This needs to be tested with significance statistics. From Table 1, it can be seen that the statistical values of the theoretical output value of the standard grain calculated by the weight matrix are also positive values, and they are all far greater than the threshold at the significance level of 0.05. The value is 1.96, which indicates that the theoretical yield of standard grain has a significant positive spatial autocorrelation in the global scope of Shaanxi Province. Through the above analysis, it can be shown that the spatial...
distribution of the standard grain theoretical yield of Shaanxi Province does have clustering characteristics. The high counties and low counties have clustering characteristics in space, and are not randomly distributed. That is to say, the counties with high yield per unit of standard grain are spatially contiguous, and the counties with low yields are spatially contiguous.

4.1.2. Generalized G statistic. In the study of spatial autocorrelation, spatial cluster analysis is often used. The local clusters with higher values are called hotspots, and the lower ones are cold spots when they are close to each other locally. Moran ’s I was unable to distinguish effectively between these two types of aggregation modes. The generalized G statistic is also a global spatial autocorrelation index. Compared with Moran ’s I, it has the advantage of being able to measure whether there are cold hot spots in the entire research area, and can distinguish between two different types of spatial autocorrelation, H-H and L-L. From the generalized G statistic method:

Table 2. Generalized G statistics results

| index           | \( G \)  | \( E \)  | \( V \)  | \( Z \)  |
|-----------------|---------|---------|---------|---------|
| Theoretical yield | 0.01295 | 0.0101  | 0.0000005 | 12.7519 |

The results of cluster analysis of the theoretical yield of standard grains in Shaanxi Province using generalized G statistics are shown in Table 2. The generalized G statistic analysis is performed on the theoretical yield of the standard grain. The expected values are 0.0101, and the G values are 0.01295, which are slightly larger than the expected values. This indicates that there is a certain degree of positive spatial autocorrelation in the study area, and the value is very low. This shows that there are cold spot LL aggregations in the study area, that is, low-value aggregations. As can be seen in Table 2, the significant statistic Z value of the theoretical yield of the standard grain is 12.7519, which is far greater than the critical value of 1.0.05 at the significance level of 1.96. For the generalized G statistic, a positive Z indicates a high value Aggregation, a negative Z indicates the presence of low value aggregation. The results show that there are significant hotspots H-H, L-L aggregation in the study area, and the Z-values of their significant statistics are shown in Figure 2.

Figure 2. Normal graph of theoretical yields of standard grains
4.2. Local spatial autocorrelation analysis

The global spatial autocorrelation index is an explanation of the autocorrelation of the entire research area. In order to show the level of autocorrelation at the local level, the spatial autocorrelation value of each area unit must be calculated. The indicators of autocorrelation analysis include local Moran's I and local G statistic and LISA aggregation map.

4.2.1. Local Moran’s I Statistics

For each area unit in the study area, the calculation of the local spatial autocorrelation index uses formula (9), and the standard deviation statistic Z value for the significance test calculation uses formula (11). The weight matrix is a row-normalized weight matrix. The local autocorrelation coefficients and standard statistic Z values for the theoretical yields of standard grains in Shaanxi Province are shown in Table 1 and Figure 3, Figure 4.

![Figure 3](image1.png)  ![Figure 4](image2.png)

**Figure 3.** The theoretical yield of the moran’s I coefficient.

**Figure 4.** The standardized yield of the theoretical yield.

It can be seen from Table 1 and Figure 3 that the theoretical yield of standard grains in Shaanxi Province presents two spatial structures in the three regions of northern Shaanxi, Guanzhong, and southern Shaanxi. There are different degrees of positive spatial autocorrelation. The silver gray area indicates that the local Moran’s I value of the counties involved is negative, showing a certain degree of negative spatial autocorrelation. In addition, the standardized Z value of the theoretical yield of standard grains in Shaanxi Province can be seen from Table 1 and Figure 4 that the standardized Z value is obviously greater than the critical value of 1.96 at a significant level of 0.05, so the correlation between these regions is significant. The normalized Z values of the silver-gray areas in the figure are all less than the critical value of -0.05 at the significance level of -0.05, and the correlation of these areas is also significant.

4.2.2. Local G statistic

The local G statistic is a statistic calculated for each area unit to indicate the correlation between the value of the area unit of interest and the value of neighboring units defined by
the distance around them. The local G statistic is preferably based on the standardized Z Value to explain. When similar and higher attribute values are clustered together, the standardized Z value will be higher; when similar but lower attribute values are clustered together, the standardized Z value will be lower; and if the standardized Z value is close to 0, it means that there are no significant spatial correlation patterns. The local G statistics and standardized Z values of the local G spatial autocorrelation analysis of the yield of the standard grain theory in Shaanxi Province can be seen in Table 2 and Figure 5 for the spatial visualization results of the standardized Z values on the map.

Figure 5. Normalized Z-value plot of local G theory yield

According to the previous explanation of the standardized values of local statistics, combined with Figure 5, it can be seen that the standardized values in the red area are significantly higher, and the standardized values in the purple area are also higher, indicating that the yield of the three standard grains is in red and purple. High-value and high-value aggregations exist in the two-color regions. Most of the northern Shaanxi region (shown as silver gray in the figure) has a smaller negative value, that is, the absolute value of the standardized value is larger, indicating that the three standard grain yields have low-value accumulation in these areas.

4.2.3. LISA Cluster graph. The LISA clustering map can quantitatively know the specific degree of these associations, can intuitively reflect the aggregation mode of each spatial unit, and indicate different types of spatial autocorrelation with different colors. The spatial association mode can be divided into four types, namely "High-High" association (HH), local high value aggregation type; "Low-Low" association (LL), local low value aggregation type; "High-Low" association (HL), local high value outlier type; "Low-High" association (LH), local low outlier type. "High-High" and "Low-Low" indicate strong spatial positive correlation, indicating the clustering and similarity of regions; "High-Low" and "Low-High" indicate strong spatial negative correlation, suggesting The existence of local heterogeneity. Figure 6 shows the LISA agglomeration chart of the theoretical yield of standard grains in Shaanxi Province. The yellow area is the high value area of Z the three standard grain yields,
which is shown as a high value aggregation, that is, HH aggregation; the blue area is the output of three standard grain yields Lower, the performance is low-value aggregation, that is, LL aggregation; the yield of the three standard grain yields in the pink area is lower than that of its adjacent area units, and it is expressed as LH aggregation. H-L aggregation is not shown in the figure.

Figure 6. Agglomeration chart of theoretical yield per standard grain

5. Conclusion analysis
Based on the above analysis, the global spatial autocorrelation statistic Moran ‘s I of the theoretical yield of standard grains in Shaanxi Province is calculated using the center-of-gravity distance weight matrix. The values I of I are 0.5496, 0.5669, and 0.5542, all of which are greater than 0 and greater than the expected value. The significance test statistic values Z are 14.44322, 14.6440, 14.3312. It seems that these values Z are far greater than the critical value of 1.96 at a significance level of 0.05. The above shows that the theoretical yield of standard grains in Shaanxi Province has significant positive spatial autocorrelation in the entire study area of Shaanxi Province. Whether there are hot and cold spots in the study area, that is, whether there are high-value aggregates and low-value aggregates, needs to be further explained by generalized G statistics. The analysis results show that the theoretical yield of standard grains is indeed true in the entire research area of Shaanxi Province. There is significant positive spatial autocorrelation, and there is significant LL aggregation and HH aggregation.

In addition, through the above three methods of local spatial autocorrelation analysis, the local autocorrelation characteristics of the standard grain yield of Shaanxi Province have been well explained. Local Moran's I explains the local spatial autocorrelation by its value, and the significance of the local spatial autocorrelation by the significance statistic Z value; the local statistic G explains the aggregation of variables in the study area by the standardized Z value; the LISA aggregation map similar to the expression of the local statistic G, the LISA aggregate graph visually indicates different
types of spatial autocorrelation with different colors. By comparing and analyzing the conclusions of the three methods of local space autocorrelation, the results are basically the same, indicating that the results of local space autocorrelation analysis are correct.

Acknowledgements
Thanks for Research on Quick Calculation of Earthwork in Land Consolidation Based on Digital Elevation Model Data (2018-TD02) and Research on Quick Calculation of Earthwork in Land Consolidation Based on Digital Elevation Model Data (DJNY2019-29).

References
[1] Shaanxi Provincial Department of Land and Resources. Implementation Plan of Shaanxi Provincial Agricultural Land Capacity Accounting Work [R]. Xi'an: Department of Land and Resources, 2010.
[2] Xiao Guoping. Application of Agricultural Land Productivity Accounting Results in Ningxia [J]. China High-tech Enterprises, 2011, (24): 11-12.
[3] Zhang Chunlai, Zhao Hongliang, Liu Wenqiong. Application of Achievements in Agricultural Land Capacity Accounting in Heilongjiang Province—Determining Priority Areas for Agricultural Land Capacity [J]. Forestry Science and Technology Information, 2011, 43 (4): 120-122.
[4] Po-Chi CHEN, Ming-Miin YU. Total factor productivity growth in China's agricultural sector [J], China Economic Review 19 (2008): 580-593.
[5] Yingfeng Xu, Agricultural productivity in China [J], China Economic Review 10 (1999): 108–121.
[6] Zhang Leina, Tong Wenju. Research on Capacity Accounting Based on Agricultural Land Classification Results [J]. Journal of Agricultural Engineering, 2008, 24 (1): 133-136.
[7] BoumaJ, DroogersP. A procedure to derive land quality indicators for sustainable arable production [J]. Geoderma, 1998, 85 (1): 103-110.
[8] Jia Shuhai, Ma Yue, Pan Jinhua, Wang Jin, Shen Yuyu. Study on Agricultural Land Productivity Accounting in Liaoning Province [J]. Hubei Agricultural Sciences, 2010, 49 (7): 1747-1751.
[9] Colin A. CARTER. Agricultural productivity growth in China: farm level versus aggregate measurement [J], China Economic Review 14 (2003): 53-71.
[10] Chao Lumen. Application of GIS / RS Technology in the Second National Land Survey [D]. Hohhot: Inner Mongolia Normal University, 2010.