

*Original Research*

# Comparative Spatial Distribution Simulation of Plateau Mountain Cultivated Land Based on Spatial Multi-Scale Model, Yunnan Central Urban Agglomeration Area, China

**Guoping Chen\***

Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming, China

*Received: 11 September 2022*

*Accepted: 17 February 2023*

## Abstract

Land use is highly dependent on scale and heterogeneous, serving as an important tool to examine land use change by comparing spatial multi-scale cultivated land models through an appropriate simulation model. In the present study, such indicators as the goodness of fit, the spatial auto-correlation of residuals, the number of factors and spatial scale are adopted to compare the variability and accuracy of four spatial multi-scale models: OLS (Ordinary Least Squares), SL (Spatial Lag), SE (Spatial Error) and GWR (Geographically Weighted Regression). According to the global characteristics, there is a positive autocorrelation between the distribution of cropland and the driving factors at multiple spatial scales and they are relatively sensitive to the sampling scale. According to the local characteristics, the smaller the spatial scale, the more accurate the spatial location of regional cropland aggregation. The order of merit of models characterizing the local spatial pattern and distribution pattern of cultivated land at the scale of plateau urban clusters is as follows: GWR, spatial error model, spatial lag model, and OLS. The GWR performs better in characterizing local spatial features, simulation accuracy and driving factor coefficients. Besides, the CLUE-S simulation model improved by GWR Logistic is more accurate in characterizing the spatial pattern of local land use and its distribution pattern at the scale of urban clusters in the plateau region.

**Keywords:** spatial multi-scale, spatial autocorrelation, Geographically Weighted Regression model (GWR), Plateau mountain cultivated land, Yunnan central urban agglomeration area

---

\*e-mail: 115432640@qq.com

## Introduction

LUCC (Land use/land cover change) research plays a key role in global change research whether in China or abroad [1]. Currently, one of the main tasks of the LUCC science program and the focus of land use science research is to "develop new LUCC modeling methods and simulation models" to make better use of LUCC models and to establish links with other models. The study of land use pattern change and dynamic simulation is the focus of research on land science, geography and other disciplines, providing an important solution to exploring land use patterns and processes [2].

Currently, land use simulation models are studied worldwide [3]. One is based on the combination of econometric statistics and models [4, 5], and the other is to set rules and scenario objectives (models) in advance [6]. On this basis, multi-scenario models are constructed for multi-scenario simulations [7]. In terms of theory, methodology and practical applications, a series of progresses have been made, such as the evolution from a single non-spatial model to the integration of non-spatial and spatial models and the diversification of functions and roles. The commonly used models include empirical statistical model [8], CA (Cellular Automata) model [9], CLUE(the Conversion of Land Use and its Effects)/CLUE-S (the Conversion of Land Use and its Effects at Small regional) model [10], FLUS (Future Land Use Simulation) model [11], SD (System Dynamics) model [12], and MAS (Multi-Agent System) model [13]. The principles followed by these models are mostly based on multiple regression and simulation. With the spatial pattern of land use is simulated through empirical probability, the focus is placed on global spatial characteristics. Despite some scholars having realized the importance of scale issues to land use study, there has been little breakthrough made in methodology and case study so far [14].

As for the formation of land use, it is influenced by various factors such as natural, economic, ecological, and social factors. Also, it is a dynamic evolutionary process with highly scale-dependent, scale-coupled, and non-smooth characteristics, the morphology and state of which can change at multiple spatial and temporal scales [15]. The difference in scale selection can have a significant impact on the results of simulation performed by the established model. In addition, the model is subjected to some limitations in expressing the local pattern and the local characteristics of the driving factors. As a result, there are changes to the driving factors affecting the pattern, which is ignored by many prior studies. Therefore, the construction of a multi-scale land use pattern model requires that consideration is given to its local spatial heterogeneity and the complexity of the spatial drivers [16]. Moreover, the correlations based on spatial and temporal scales are essential for the existence of order, pattern and

diversity in nature [17]. Most of the current studies on land use simulation adopt mathematical models but fail to quantify the scale effects and the magnitude of impact caused by land use patterns. In addition, the spatial analysis ignores the local and non-stationary characteristics of land use and drivers, which is adverse to revealing the characteristics of autocorrelation within the data.

For quite long, cultivated land conservation has been playing an important part in the development of agricultural economy for China [18]. In view of the special status and importance of cultivated land, the cultivated land in Yunnan central urban agglomeration area is taken as the research object in this paper to construct cultivated land prediction models based on spatial autoregression and geographically weighted regression for analysis on the overall and local clustering characteristics of the spatial pattern of cultivated land at multiple scales of space. Besides, a comparative analysis is conducted on their accuracy to identify a more accurate cultivated land simulation model, and to improve the existing models. On this basis, the scale applicability of the models and the accuracy of land spatial layout simulation are improved, thus providing theoretical reference and technical support for cultivated land protection.

## Material and Methods

### Overview of the Study Area

As one of the 19 urban agglomerations designated by the state, Yunnan central urban agglomeration area is the most well-developed region in Yunnan Province of China both economically and socially. It features the most superior resource endowment and the most significant location advantages [19], attracting the main production factors and economic scale from across the province. It is also the area that can best demonstrate the prosperity and strength of Yunnan, providing a key driving force for the growth of provincial economy. The GDP generated by the urban agglomeration in 2020 amounts to 12320.66 billion yuan, accounting for 53.05% of the provincial total. It has a population of 18,139,400, accounting for 37.56% of the provincial total.

### Data Source

In this paper, the land use classification data for the three years of 2000, 2010 and 2020 are sourced from the Resource and Environment Data Center of the Chinese Academy of Sciences (<https://www.resdc.cn/>). The elevation data are obtained by DEM processing with SRTM 30 m resolution [20]. The data of population and socio-economic development are collected from the statistical yearbook of Yunnan Province of China in previous years.

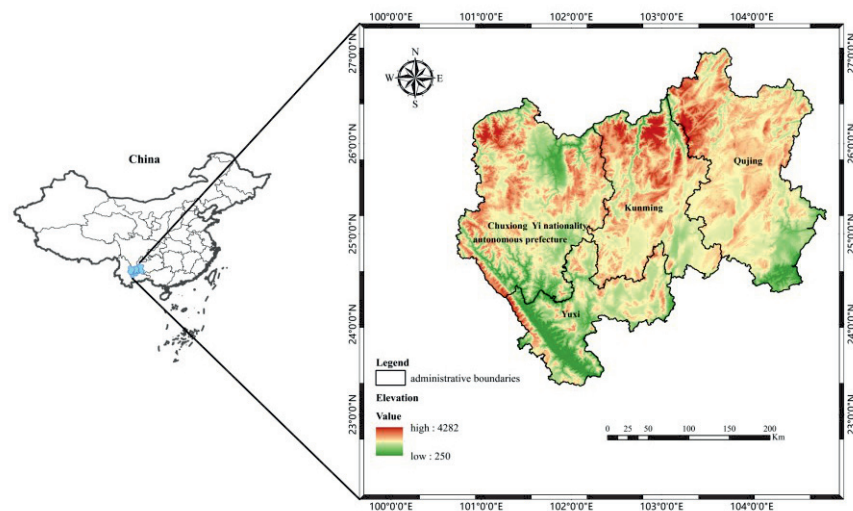


Fig. 1. Location of study area.

Due to the wide range of data sources, there are variations in data format, coordinate system and scale, etc. Based on the 2020 land use classification data, data assimilation was performed to process other multi-source data [20], such as the distance data obtained by GIS spatial analysis, slope and slope direction as calculated by DEM data. Besides, each socioeconomic factor was rasterized.

Driving forces refer to the main influencing factors for changes in land use patterns and purposes. They are the driving factors in the evolution of land use. By combining the characteristics of the study area and data accessibility, a total of 16 influencing factors were identified as drivers from five perspectives: natural environment and topography, traffic accessibility, water resources, scale indicators, and intensity indicators. They are shown in Table 1. (1) natural and topographic indicators include elevation, slope and slope direction; (2) transportation accessibility indicators include the closest distance to roads, the closest distance to railroads, the closest distance to rural roads, the closest distance to towns, and the closest distance to villages; (3) water resources indicators include the distance to ditches and the distance to rivers system; (4) scale indicators include total population and agricultural population; and (5) intensity indicators include GDP per capita, population density, net income per farmer and cultivated land per capita.

On this basis, the logistic function of SPSS stepwise regression analysis was applied to analyze the drivers of land use. Given the sensitivity of logistic regression model to the presence of multivariate covariance in the independent variables, it is necessary to diagnose covariance among the independent variables before regression analysis, so as to eliminate the factors with significant covariance while improving the accuracy of model construction. In general, the covariance diagnosis is assessed against the tolerance (TOL) and the variance inflation factor (VIF, Variance Inflation Factor),

which are correlated with each other. The smaller the tolerance value, the stronger the covariance. Usually, it is considered to have covariance when the value is less than 0.2, and have serious covariance problem when the value is less than 0.1. The larger the VIF, the stronger the covariance. It is considered to have serious covariance problem when the value is greater than 10. In this study, covariance diagnosis was performed on 16 independent variables initially selected according to the different needs of the model.

Table 1. Indicators of Drivers of Land Use Pattern Evolution in the Study Area.

Categories	Factors
Natural and topographic	Elevation(E)
	Slope(S)
	Aspect(A)
Traffic accessibility	Distance from railroads(D-RR)
	Distance to roads(D-RO)
	Distance from country roads(D-CR)
	Distance to towns(D-T)
	Distance to villages(D-V)
Water Resources	Distance to ditches(D-D)
	Distance to rivers(D-RI)
Scale indicators	Total population(TP)
	Agricultural population(AP)
Intensity indicators	GDP per capita(GDP/pc)
	Population Density(PD)
	GDP per farmer(GDP/pf)
	Cultivated land per capita(CL/pc)

The difference in scale of research can lead to variations in the level of detail in the study population, which leads to different findings [21]. Herein, reference was made to the study of Wang Yuanfei et al [22] for

the optimal sample size formula  $Q = \frac{2A}{n}$  (where  $Q$

represents the sample area,  $A$  denotes the study area, and  $n$  stands for the number of study area points), and to the proportional relationship between the study area and the scale selection by previous authors [23, 24]. On this basis, 500 m×500 m was set as the basic study scale and GIS software was applied to convert the land use data into the five raster scales as required for the study: 1 km, 5 km, 10 km, 20 km, and 30 km.

## Research Methods

### *Spatial Autocorrelation*

Spatial correlation aims to analyze the characteristics of spatial distribution of spatial units based on the matching of location similarity and attribute similarity [25]. If the values at the neighboring locations are similar, positive spatial autocorrelation is observable; otherwise, negative spatial autocorrelation can be observed. The spatial autocorrelation analysis conducted in this paper involves the global Moran's I index, local Moran's I index, local G index and Moran scatter plot [17].

#### *Global Spatial Autocorrelation*

The global spatial autocorrelation indicates the extent of spatial dependence within a total spatial extent [26]. It is expressed as follows.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n \sum_{j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j \neq i}^n w_{ij}} \quad (1)$$

Where  $x_i$  and  $x_j$  represent the attribute value of the element  $i$  and  $j$ ,  $\bar{x}$  denotes its average value,  $w_{ij}$  indicates the spatial weight between the elements  $i$  and  $j$ , and  $n$

refers to the total number of elements,  $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ .

#### *Local Spatial Autocorrelation*

The global assessment fails to indicate the specific spatial location where aggregation or anomalies occur. This is because different spatial units and neighborhoods within the study area vary to some extent in the level of spatial autocorrelation [27, 28]. To address this problem, a local spatial autocorrelation analysis is required. The three main methods to achieve

this purpose are G-statistics for spatial linkage, local indicators and Moran scatter plots.

#### (1) LISA (Local indicators of spatial association)

LISA is used to measure the degree of similarity (positive correlation) or difference (negative correlation) between the value of observed unit attribute and that of the surrounding unit attribute. LISA includes the local Moran index and the local Geary index. The local Moran index is expressed as follows:

$$I_i = \frac{n(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{nz_i \sum_j w_{ij} z_j}{z^T z} = z_i' \sum_j w_{ij} z_j' \quad (2)$$

where  $z_i'$  and  $z_j'$  represent the standard deviation standardized observations.

#### (2) Moran scatter plot

Moran scatter plots are the statistical graphical methods that can be used to reflect the local autocorrelation of spatial location attributes [29] and to express the presence of concentrated aggregation or anomalous features within a local area. Moran scatter diagrams are presented in the form of coordinates in four quadrants, which correspond to four types of local spatial forms of connection between the regional units and their neighbors, respectively: high-high aggregation (HH) in the first quadrant; high-low anomaly (HL) in the second quadrant; low-low aggregation (LL) in the third quadrant; and low-high anomaly (LH) in the fourth quadrant.

#### *Spatial Autoregressive Model*

Based on the spatial correlation between the explanatory and dependent variables, Anselin L was used to obtain a generic sequential spatial linear regression equation [16]:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \quad (3)$$

Where  $y_i$ ,  $x_{ik}$ , and  $\varepsilon_i$  represent the dependent variable, independent variable, and random error components, respectively,  $\beta_0$  denotes the constant of the model, and  $\beta_0$  indicates the regression coefficient of the  $k$ -th independent variable.

#### *Geographically Weighted Regression Model*

GWR model was first proposed by British geostatistician A.S Fortheringham. Taking into account the more significant spatial correlation characteristics as observed and the spatial variations in the effects of different influencing factors on the dependent variable, the model can achieve higher accuracy and better goodness of fit through the GWR model [30]. It is expressed as follows.

$$y_i = \beta_0(\mu_i, \nu_i) + \sum_k \beta_k(\mu_i, \nu_i) x_{ik} + \varepsilon_i \quad (4)$$

Where  $(\mu_i, v_i)$  represents the spatial location coordinate of the  $i$ -th sample point,  $\beta_k$  denotes the regression coefficient changing with the spatial location, and  $\varepsilon_i$  refers to the standard deviation of the error estimation term.

## Results and Discussion

### Spatial Autocorrelation Analysis of Multi-Scale Model of Cultivated Land

#### *Overall Clustering Analysis of Global Cultivated Land Global Characteristics*

With the assistance of GeoDa software and ArcGIS spatial analysis tools, calculation was performed for the spatial autocorrelation coefficients Moran's I

values of cultivated land and 16 drivers at five scales of 1 km, 5 km, 10 km, 20 km and 30 km (Table 2). On this basis, it was found out that all of them tended to decrease with the increase in weight distance. Besides, all of the 13 driving factors showed strong spatial autocorrelation except for three of them, including slope direction, the distance from road and the distance from water system. Its implications are as follows. Firstly, the spatial pattern of cultivated land and its driving factors show spatial non-stationarity, and the spatial geographical location has a significant impact on various factors. Secondly, the smaller the weight distance, the more significant the spatial dependence, and spatial heterogeneity with the increase of the weight distance. Besides, the relationship is highly consistent at five spatial scales.

By comparing the values of the coefficients of different drivers at different spatial homogeneous scales,

Table 2. Multi-scale Spatial Auto-correlation Coefficients Moran's I of Cultivated Land and Driving Factors.

Scale/Weight distance		30	50	70	90	110	130	150	170	190	210	230	250
Cultivated land	1	0.297	0.246	0.233	0.202	0.167	0.163	0.155	0.153	0.142	0.131	0.124	0.117
	5	0.331	0.273	0.240	0.217	0.200	0.186	0.174	0.161	0.149	0.139	0.130	0.123
	10	0.448	0.372	0.328	0.296	0.273	0.253	0.235	0.217	0.199	0.184	0.171	0.160
	20	0.536	0.456	0.405	0.354	0.327	0.298	0.275	0.249	0.220	0.202	0.183	0.171
	30	0.606	0.485	0.408	0.379	0.340	0.289	0.289	0.252	0.222	0.204	0.177	0.159
E	1	0.473	0.366	0.323	0.276	0.251	0.228	0.200	0.181	0.168	0.157	0.146	0.135
	5	0.484	0.389	0.330	0.288	0.257	0.231	0.211	0.193	0.177	0.163	0.152	0.144
	10	0.495	0.402	0.341	0.295	0.264	0.238	0.216	0.197	0.179	0.164	0.151	0.143
	20	0.561	0.482	0.423	0.359	0.326	0.292	0.269	0.245	0.217	0.198	0.180	0.169
	30	0.539	0.479	0.376	0.346	0.298	0.271	0.249	0.220	0.196	0.179	0.158	0.147
S	1	0.482	0.433	0.389	0.332	0.287	0.269	0.238	0.201	0.188	0.176	0.165	0.143
	5	0.550	0.465	0.401	0.350	0.308	0.273	0.244	0.221	0.204	0.190	0.180	0.172
	10	0.626	0.533	0.459	0.397	0.348	0.308	0.273	0.246	0.225	0.209	0.196	0.186
	20	0.671	0.587	0.518	0.435	0.387	0.339	0.302	0.271	0.240	0.223	0.206	0.194
	30	0.704	0.623	0.519	0.470	0.404	0.359	0.321	0.282	0.250	0.232	0.208	0.191
A	1	0.098	0.066	0.051	0.038	0.031	0.025	0.027	0.026	0.025	0.024	0.032	0.020
	5	0.106	0.074	0.054	0.043	0.036	0.031	0.028	0.027	0.026	0.025	0.023	0.022
	10	0.130	0.093	0.068	0.052	0.043	0.037	0.035	0.033	0.031	0.029	0.027	0.025
	20	0.106	0.082	0.055	0.032	0.027	0.024	0.026	0.025	0.024	0.021	0.019	0.017
	30	0.148	0.083	0.071	0.054	0.046	0.045	0.043	0.039	0.036	0.034	0.028	0.022
D-RR	1	0.886	0.765	0.667	0.545	0.475	0.434	0.397	0.365	0.321	0.292	0.274	0.265
	5	0.839	0.721	0.624	0.525	0.454	0.401	0.360	0.326	0.300	0.279	0.263	0.252
	10	0.831	0.719	0.616	0.523	0.453	0.399	0.355	0.322	0.293	0.271	0.254	0.242
	20	0.777	0.672	0.583	0.478	0.419	0.364	0.325	0.292	0.258	0.239	0.221	0.209
	30	0.745	0.658	0.525	0.472	0.396	0.344	0.309	0.268	0.235	0.218	0.195	0.181



Table 2. Continued.

D-RO	1	0.144	0.092	0.082	0.071	0.061	0.055	0.051	0.046	0.044	0.041	0.038	0.033
	5	0.157	0.108	0.090	0.075	0.065	0.059	0.054	0.050	0.046	0.043	0.041	0.039
	10	0.134	0.097	0.083	0.069	0.060	0.054	0.049	0.045	0.041	0.037	0.034	0.032
	20	0.121	0.099	0.086	0.054	0.048	0.045	0.040	0.030	0.035	0.030	0.024	0.022
	30	0.083	0.111	0.081	0.075	0.063	0.057	0.050	0.043	0.038	0.031	0.022	0.018
D-CR	1	0.892	0.783	0.691	0.643	0.552	0.486	0.435	0.397	0.364	0.52	0.338	0.325
	5	0.872	0.776	0.685	0.605	0.532	0.470	0.421	0.383	0.357	0.340	0.327	0.317
	10	0.871	0.772	0.682	0.597	0.525	0.464	0.412	0.376	0.348	0.331	0.318	0.308
	20	0.832	0.747	0.670	0.570	0.510	0.446	0.400	0.362	0.329	0.312	0.299	0.289
	30	0.781	0.734	0.616	0.568	0.485	0.427	0.388	0.341	0.306	0.294	0.275	0.262
D-T	1	0.687	0.543	0.452	0.442	0.338	0.309	0.266	0.254	0.220	0.196	0.188	0.176
	5	0.659	0.486	0.378	0.309	0.264	0.235	0.208	0.187	0.173	0.164	0.156	0.151
	10	0.641	0.473	0.366	0.293	0.251	0.222	0.196	0.176	0.160	0.151	0.143	0.137
	20	0.532	0.400	0.316	0.239	0.208	0.180	0.159	0.139	0.123	0.114	0.107	0.101
	30	0.433	0.418	0.287	0.253	0.204	0.181	0.158	0.135	0.118	0.110	0.097	0.089
D-V	1	0.867	0.801	0.745	0.695	0.632	0.544	0.499	0.422	0.401	0.388	0.365	0.336
	5	0.852	0.753	0.664	0.585	0.520	0.468	0.428	0.395	0.370	0.349	0.332	0.317
	10	0.844	0.751	0.664	0.582	0.518	0.466	0.422	0.390	0.362	0.341	0.322	0.306
	20	0.795	0.708	0.634	0.540	0.486	0.432	0.394	0.362	0.329	0.310	0.289	0.275
	30	0.781	0.698	0.584	0.537	0.463	0.413	0.377	0.338	0.309	0.291	0.268	0.250
D-D	1	0.242	0.178	0.143	0.132	0.119	0.102	0.095	0.086	0.080	0.074	0.069	0.061
	5	0.225	0.165	0.139	0.121	0.105	0.093	0.084	0.076	0.070	0.065	0.061	0.059
	10	0.149	0.106	0.091	0.078	0.068	0.059	0.052	0.048	0.042	0.039	0.037	0.035
	20	0.123	0.089	0.077	0.062	0.055	0.049	0.044	0.039	0.034	0.031	0.027	0.026
	30	0.165	0.132	0.097	0.083	0.067	0.058	0.048	0.037	0.033	0.029	0.023	0.017
D-RI	1	0.961	0.923	0.883	0.834	0.771	0.723	0.678	0.623	0.570	0.537	0.492	0.465
	5	0.956	0.917	0.871	0.821	0.768	0.716	0.665	0.616	0.569	0.525	0.488	0.456
	10	0.952	0.915	0.870	0.817	0.765	0.713	0.659	0.611	0.562	0.519	0.479	0.447
	20	0.934	0.896	0.857	0.792	0.746	0.689	0.641	0.591	0.532	0.492	0.450	0.419
	30	0.904	0.894	0.830	0.792	0.733	0.682	0.639	0.580	0.524	0.491	0.438	0.399
TP	1	0.773	0.684	0.611	0.544	0.501	0.477	0.403	0.397	0.374	0.350	0.321	0.304
	5	0.707	0.617	0.547	0.492	0.449	0.413	0.383	0.357	0.335	0.315	0.299	0.284
	10	0.654	0.570	0.506	0.447	0.405	0.370	0.340	0.316	0.294	0.276	0.260	0.247
	20	0.513	0.440	0.384	0.326	0.300	0.272	0.253	0.234	0.212	0.197	0.183	0.173
	30	0.478	0.399	0.296	0.262	0.233	0.217	0.196	0.182	0.161	0.151	0.136	0.124
AP	1	0.754	0.698	0.654	0.605	0.532	0.499	0.473	0.421	0.398	0.387	0.364	0.352
	5	0.739	0.658	0.591	0.538	0.495	0.459	0.427	0.398	0.373	0.350	0.330	0.313
	10	0.687	0.609	0.548	0.491	0.449	0.414	0.382	0.355	0.330	0.309	0.290	0.274
	20	0.548	0.478	0.424	0.368	0.341	0.313	0.291	0.270	0.245	0.228	0.211	0.199
	30	0.508	0.432	0.329	0.295	0.266	0.247	0.224	0.208	0.187	0.175	0.157	0.143

Table 2. Continued.

TP	1	0.773	0.684	0.611	0.544	0.501	0.477	0.403	0.397	0.374	0.350	0.321	0.304
	5	0.707	0.617	0.547	0.492	0.449	0.413	0.383	0.357	0.335	0.315	0.299	0.284
	10	0.654	0.570	0.506	0.447	0.405	0.370	0.340	0.316	0.294	0.276	0.260	0.247
	20	0.513	0.440	0.384	0.326	0.300	0.272	0.253	0.234	0.212	0.197	0.183	0.173
	30	0.478	0.399	0.296	0.262	0.233	0.217	0.196	0.182	0.161	0.151	0.136	0.124
GDP/pc	1	0.601	0.553	0.475	0.400	0.339	0.321	0.287	0.251	0.221	0.197	0.165	0.146
	5	0.569	0.436	0.351	0.292	0.250	0.219	0.193	0.173	0.158	0.147	0.139	0.133
	10	0.525	0.405	0.332	0.272	0.234	0.204	0.178	0.159	0.143	0.131	0.122	0.116
	20	0.468	0.373	0.311	0.244	0.213	0.181	0.157	0.137	0.118	0.107	0.098	0.093
	30	0.496	0.433	0.329	0.295	0.249	0.221	0.196	0.178	0.159	0.145	0.129	0.118
PD	1	0.652	0.500	0.432	0.365	0.301	0.275	0.231	0.198	0.165	0.154	0.146	0.139
	5	0.611	0.453	0.351	0.288	0.245	0.213	0.188	0.169	0.154	0.143	0.136	0.132
	10	0.576	0.422	0.330	0.265	0.225	0.195	0.171	0.152	0.136	0.125	0.118	0.113
	20	0.504	0.390	0.314	0.239	0.207	0.174	0.151	0.132	0.113	0.101	0.093	0.089
	30	0.401	0.308	0.205	0.177	0.136	0.122	0.101	0.088	0.072	0.064	0.055	0.051
GDP/pf	1	0.566	0.500	0.455	0.421	0.355	0.288	0.276	0.242	0.220	0.211	0.188	0.174
	5	0.524	0.460	0.405	0.356	0.313	0.276	0.244	0.217	0.196	0.180	0.168	0.159
	10	0.465	0.403	0.355	0.310	0.275	0.243	0.214	0.189	0.168	0.153	0.140	0.131
	20	0.435	0.384	0.342	0.285	0.252	0.218	0.192	0.169	0.143	0.129	0.115	0.107
	30	0.423	0.381	0.322	0.291	0.254	0.226	0.197	0.173	0.145	0.131	0.111	0.099
CL/pc	1	0.550	0.443	0.356	0.308	0.242	0.200	0.156	0.112	0.101	0.097	0.092	0.090
	5	0.450	0.349	0.271	0.209	0.161	0.130	0.110	0.098	0.092	0.090	0.089	0.089
	10	0.372	0.281	0.216	0.158	0.119	0.094	0.077	0.068	0.064	0.064	0.064	0.064
	20	0.291	0.223	0.180	0.128	0.101	0.059	0.059	0.050	0.044	0.043	0.043	0.042
	30	0.240	0.196	0.138	0.112	0.082	0.064	0.045	0.038	0.036	0.032	0.032	0.030

it was discovered that the three topographic drivers, including cultivated land and elevation, slope and slope orientation, increased with larger spatial scales. This is because the spatial autocorrelation coefficient has a similar variable at a location in the spatial field to the variable at its neighboring location. For some specific spatial information, there is a certain level of spatial dependence in itself, and different spatial scales are adopted to smooth them. When the spatial scale increases, the more information it incorporates relatively, the greater the similarity to neighboring variables, and the higher the similarity coefficient. In addition, these 13 driving factors, such as the distance from roads, railroads, water, and total population, show a decreasing trend with the increase of scale. There are two reasons for this. On the one hand, the spatial information about these influencing factors is largely determined by the difference between the location information and another variable. On the other hand, the increase of scale has no impact on the change to location information

about another variable, but the difference to the real information can become increasingly significant, which reduces the spatial correlation coefficient reflected by it continuously. In the present study, it is demonstrated that weight values and spatial scales are the significant factors affecting the clustering results during overall clustering analysis. Therefore, it is necessary to conduct local cluster analysis on the regional characteristics of spatial structure and the drivers of cultivated land.

#### *Local Clustering Analysis of Cultivated Land*

According to the LISA clustering map shown in Fig. 2, the clustering areas of cropland at each scale are similar, with high-high clustering areas concentrating in the central, southern and eastern parts of the study area, but less cropland in the western and northwestern parts. Besides, there are a few “islands” observed at the edge of the study area. Notably, the 1 km grid is too dense. Also, due to insufficient resolution, the grid boundary

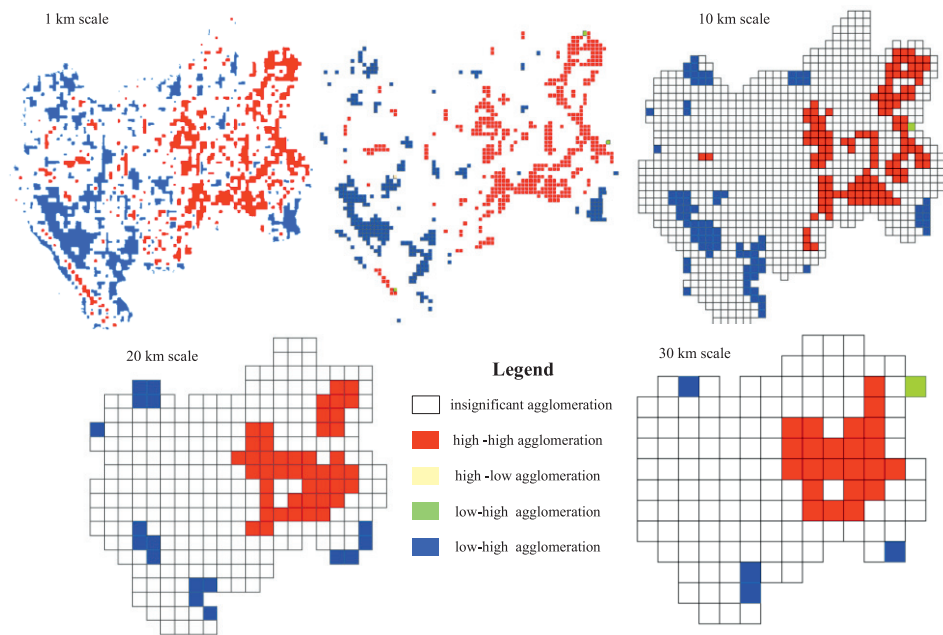


Fig. 2. Local clustering of multi-scale cultivated land spatial features.

and the study results are not clearly observable. Therefore, the scale boundary is hidden. Despite the high consistency shown by the overall characteristics of the cultivated land agglomeration areas with spatial scale changes, there are certain differences in the specific spatial location and the degree of detailed expression. In general, the larger the scale, the more insignificant the figure, the smaller the number of significant agglomerations, and the less obvious they are. Conversely, the smaller the scale, the more obvious the agglomeration, the larger the number of significant

agglomeration, and the more significant its local spatial differences.

In addition, slope was taken as the study object to illustrate the characteristics of local clustering map for the driving factors of cultivated land (Fig. 3). The study area is a highland mountainous region, the northern, western and northwestern parts of which are dominated by mountains with steeper slopes. They are shown as red agglomeration areas in the figure. Specifically, the central and eastern parts are dominated by alluvial plains with relatively flat terrain and gentle

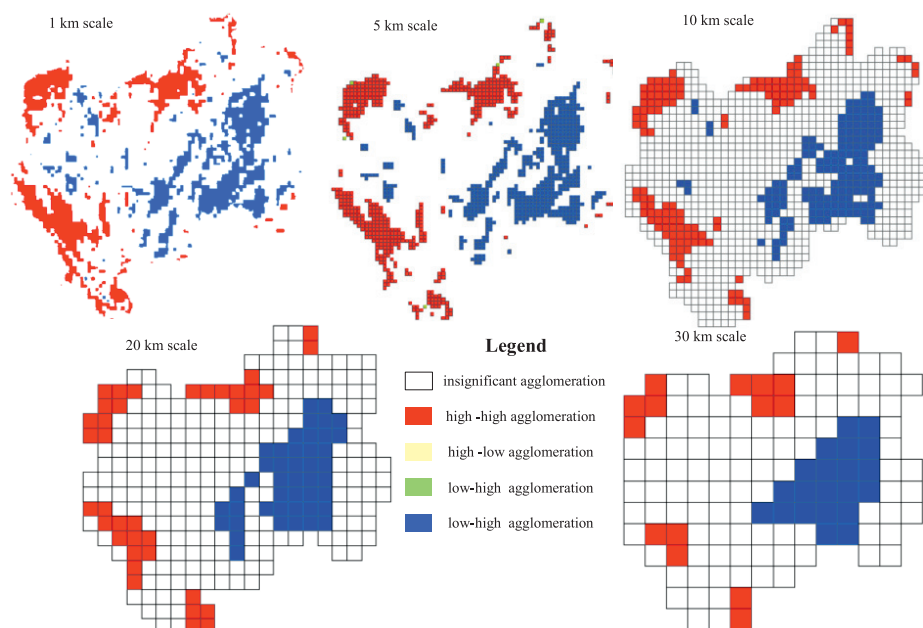


Fig. 3. Local clustering of multi-scales slope spatial features.



slopes. They are shown as blue agglomeration areas in the figure. By comparing the spatial distribution pattern with that of cultivated land, it can be found out that the spatial distribution pattern and scale characteristics of slope are closely correlated and similar to those of cultivated land. There are different spatial characteristics in different geographic aggregation areas, indicating a certain spatial heterogeneity in the spatial distribution characteristics of slope. As the spatial scale varies, change occurs accordingly to the resolution (granularity) of slope clustering maps at different spatial scales, which is similar to that of spatial multi-scale cultivated land spatial characteristics clustering. Scale is inversely related to resolution. That is to say, the smaller the scale, the more detailed the representation, and the more accurate the information provided for a specific spatial geographic location.

### Spatial Autoregressive Model Analysis

The aforementioned analysis fails to clarify the spatial autocorrelation between slope direction, the distance from the road and the distance from the water system. With these three factors excluded, the standard error, T-value (t-Statistic), P-value (Probability), and parameters such as LIK (Log likelihood), AIC (Akaike info criterion), and SC (Schwarz criterion) were used to construct the classical linear regression model of spatial and temporal evolution of cultivated land in the study area as well as the spatial autoregressive-based spatial lag model and spatial error model for comparison of the analytical results.

### OLS Regression Model

In the least squares model, GDP per capita, population density and cultivated land per capita were identified as insignificant ( $p > 0.05$ ). Also, by discounting those variables, an OLS model was constructed for cultivated land (Table 3).

According to the t-values, the most significant factors affecting cultivated land pattern in the study area include slope, the distance from villages, and the distance from railroads. By contrast, the less significant influencing factors include elevation, the distance from towns, and GDP per farmer.

The same method was adopted to construct the OLS models of cultivated land at 5 km, 10 km, 20 km, and 30 km scales, the results of which are as follows:  $R^2 = 0.424526$ ,  $LIK = 661.335$  at the 5 km scale;  $R^2 = 0.482097$ ,  $LIK = 510.390$  at the 10km scale;  $R^2 = 0.576697$ ,  $LIK = 203.302$  at the 20 km scale; and  $R^2 = 0.550092$ ,  $LIK = 94.276$  at the 30km scale.  $R^2$  was treated as a goodness-of-fit indicator to evaluate the OLS regression model. It was discovered that the  $R^2$  value is the largest at the 20 km scale. When LIK is used as the evaluation index, the OLS model at the 1 km scale is significantly better than at the other scales.

### Spatial Lag Model

It was verified that these five factor variables, including the distance from town, total population, agricultural population, GDP per capita, and population density, were insignificant in the spatial lag model.

Table 3. Ordinary Least Squares Model of 1km Scale Cultivated Land.

Impact factor	Regression coefficients	Standard error	T-value	P-value
Constant	0.670765	0.04577	14.65610	0.00000
E	-2.84E-05	9.57E-06	-2.96705	0.00303
S	-0.02203	0.00084	-26.29450	0.00000
A	0.00071	0.00016	4.57700	0.00000
D-RO	-1.24E-05	1.28E-06	-9.68755	0.00000
D-RR	-2.12E-06	2.04E-07	-10.41440	0.00000
D-CR	6.80E-07	1.20E-07	5.68816	0.00000
D-D	-7.61E-07	8.45E-08	-9.00235	0.00000
D-RI	-9.34E-06	1.46E-06	-6.39757	0.00000
D-T	-1.20E-06	3.08E-07	-3.89488	0.00010
D-V	3.60E-06	2.24E-07	16.04930	0.00000
TP	-0.00218	0.00047	-4.68501	0.00000
AP	0.00440	0.00053	8.37750	0.00000
GDP/pf	-1.82E-05	4.58E-06	-3.97625	0.00007

Note:  $R^2 = 0.395104$ ,  $LIK = 922.781$ .

Table 4. Lag Model of 1km Scale Cultivated Land.

Impact factor	Regression coefficients	Standard error	Z-value	P-value
$\lambda$	0.584616	0.0154711	37.7877	0.00000
Constant	0.613218	0.04795	12.78970	0.00000
E	-4.26E-05	8.55E-06	-4.98803	0.00000
S	-0.01575	0.00078	-20.15570	0.00000
A	0.00054	0.00013	4.08816	0.00004
D-RO	-8.87E-06	1.09E-06	-8.11068	0.00000
D-RR	-7.96E-07	1.73E-07	-4.60806	0.00000
D-CR	2.71E-07	1.01E-07	2.69312	0.00708
D-D	-4.26E-07	7.52E-08	-5.67091	0.00000
D-RI	-4.26E-07	7.52E-08	-5.67091	0.00000
D-V	1.93E-06	2.02E-07	9.58138	0.00000
GDP/pf	-3.54E-05	3.81E-06	-9.29544	0.00000
AP	0.00440	0.00053	8.37750	0.00000
CL/pc	-0.02455	0.00530	-4.62782	0.00000

Where  $R^2 = 0.567574$ ,  $LIK = 1363.54$ .

Besides, with these five variables discounted, a spatial lag model was established for cultivated land in the study area (Table 4).

The same method was applied to construct a spatial lag model for cultivated land at 5 km, 10 km, 20 km and 30 km scales, respectively. The results are as follows:  $R^2 = 0.587329$  and  $LIK = 744.321$  at the 5 km scale;  $R^2 = 0.616045$  and  $LIK = 604.616$  at the 10 km scale;  $R^2 = 0.663959$  and  $LIK = 220.018$  at the 20km scale; and  $R^2 = 0.671084$  and  $LIK = 104.564$  at the 30 km scale. It was discovered that the 30 km scale is the optimal if  $R^2$  is used for evaluation and the 1km scale is the optimal if  $LIK$  is used for evaluation.

#### *Spatial Error Model*

Similarly, after the removal of the distance to town, total population, agricultural population, GDP per capita, population density, cultivated land per capita as six insignificant factor variables, a spatial error model of cultivated land in the study area was constructed (Table 5).

The same method was adopted to respectively construct the spatial error models of cultivated land at 5 km, 10 km, 20 km and 30 km scales, the results of which are as follows:  $R^2 = 0.626523$ ,  $LIK = 860.007658$  at the 5 km scale;  $R^2 = 0.649617$ ,  $LIK = 627.590531$  at the 10 km scale;  $R^2 = 0.693935$ ,  $LIK = 224.320452$  at the 20 km scale; and  $R^2 = 0.649589$ ,  $LIK = 100.760601$  at the 30 km scale. It can be seen from above that the goodness of fit evaluation of the spatial error model is similar to that of the spatial lag model. It is the

optimal at the 20 km scale if  $R^2$  is used for evaluation and the optimal at the 1 km scale if  $LIK$  is used for evaluation.

Tables 3 and 5 show an analysis on the regression coefficients, standard errors, T-values (Z-values) and P-values of the driving factors for the relevant parameters of the constructed models at multiple spatial scales. It can be found out that 13 of the 16 influencing factors pass the significance test in the OLS model, 11 in the spatial lag model, and 10 in the spatial error model. It indicates that the influencing factors used in this paper have impact on the distribution of spatial pattern of cultivated land in the study area. Overall, the spatial error model is more sensitive to the variability of the driving factors, followed by the spatial lag model and the OLS model. By analyzing the regression coefficients of each driver, it was found out that all the factors in the 2 models were negatively correlated except for 4 factors, showing positive correlation: the distance from villages, average slope direction, agricultural population, and the distance to rural roads. Meanwhile, except for the constant term, the absolute magnitude of the drivers of the three models shows that topographic factors and human activities are more influential on the pattern formation of cultivated land.

It was also discovered that the average slope was negative in all three models and the absolute value was the largest in both the OLS and error models. This is because the cultivated land was distributed in the areas with less slope, the gentler the more suitable for cultivated crops, but also susceptible to human exploitation.

Table 5. Spatial Error Model of 1km Scale Cultivated Land.

Impact factor	Regression coefficients	Standard error	Z-value	P-value
Constant	1.14819	0.05938	19.33550	0.00000
E	-0.000129	1.50E-05	-8.60044	0.00000
S	-0.02630	0.00105	-25.01700	0.00000
A	0.00053	0.00016	3.38410	0.00071
D-RO	-1.16E-05	1.46E-06	-7.92372	0.00000
D-RR	-1.87E-06	4.63E-07	-4.04799	0.00005
D-CR	5.98E-07	2.78E-07	2.15094	0.03148
D-D	-8.98E-07	1.92E-07	-4.67352	0.00000
D-RI	-9.43E-06	1.64E-06	-5.75370	0.00000
D-V	2.84E-06	4.89E-07	5.79491	0.00000
GDP/pf	-3.98E-05	6.07E-06	-6.54361	0.00000
LAMBDA	0.674875	0.014626	46.1409	0.00000

Where  $R^2 = 0.607962$ ,  $LIK = 1473.349989$ .

### Analysis of Geographically Weighted Regression Model

Through the GWR, there is a better fit of the land class in different geographical locations and the weight of each influencing factor in different locations can be known [11]. Table 6 and Fig. 3. show the spatial distribution of cropland as simulated using the spatial multiscale GWR model. The probability of spatial distribution of cultivated land mainly ranges between 0 and 30%.

From Table 6 and Fig. 4, it can be seen clearly that the cultivated land is distributed mostly in the east of central China, not the southwest and northwest. When the probability distribution ranges between 31% and 45%, the simulated probability exceeds the statistical probability and shows an increasing trend with the increase of scale, indicating that the simulated cultivated land in this probability interval increases the probability of cultivated land distribution to a certain extent and expands the occupied area of cultivated land; when the statistical probability ranges between 46% and 60% or 61% and 100%, the simulated probability is lower than

the statistical probability and shows a decreasing trend with the increase of scale, indicating that the simulated cultivated land reduces the probability of cultivated land distribution, thus reducing the occupied area of cultivated land. To sum up, the bandwidth increases with spatial scale and the simulation probability of the GWR model rises, while the spatial resolution of the relative cultivated land patch declines significantly.

### Model Accuracy Comparison Analysis

#### *Comparison of Goodness-of-Fit*

From the calculation results in Tables 3 and 6, it can be seen that the largest regression coefficient is the spatial error model, followed by the OLS, and the spatial lag model. It is possibly because that some of the prediction results obtained by the spatial lag model are based on the prediction of spatial autoregression.

For the traditional models of typical linear regression,  $R^2$  can be used as an evaluation indicator of the goodness of fit. However, it may not be suitable for spatial autoregressive models as the value output from

Table 6. Comparison of Spatial Distribution Probability of Cultivated Land in Multi-scale GWR Model.

Probability interval	0.5 km×0.5 km Statistical probability	Simulated probability				
		1 km×1 km	5 km×5 km	10 km×10 km	20 km×20 km	30 km×30 km
0-15	43.79	42.66	41.29	37.63	39.51	41.30
16-30	17.65	18.52	20.16	27.03	28.32	29.71
31-45	14.20	15.39	16.70	18.53	20.63	21.74
46-60	12.68	12.32	11.43	11.27	9.79	6.52
61-100	11.68	11.11	10.43	5.54	1.75	0.72

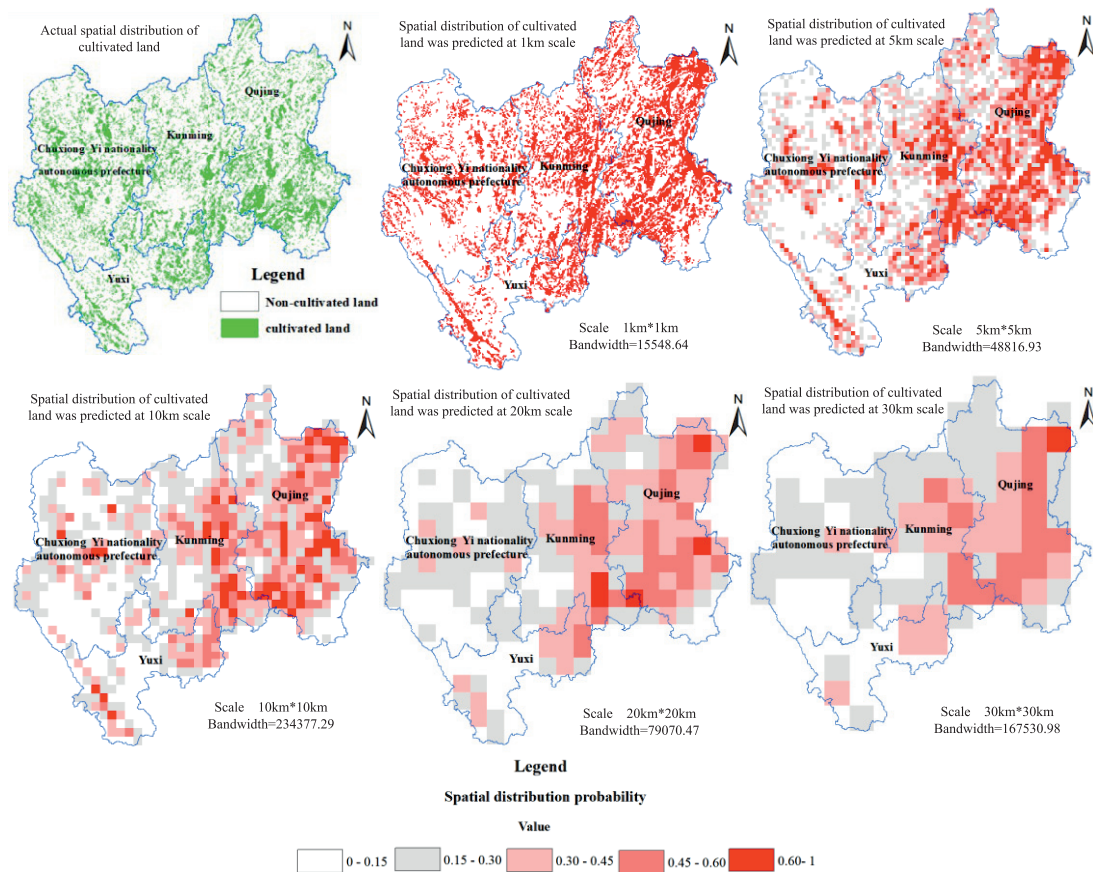


Fig. 4. Probability map of cultivated land spatial distribution of multi-scale GWR model.

spatial autoregressive models is a pseudo  $R^2$  value [31]. Currently, LIK, AIC, and SC are commonly used to evaluate spatial autoregressive models for the goodness of fit (Table 7).

As shown in Table 7, the LIK value of the spatial error model is the largest while the AIC and SC values are the smallest at the same scale. At the 1km scale, the LIK value is the largest while the AIC and SC values are the smallest, i.e., the LIK value decreases as the scale increases while the AIC and SC values rise as the scale decreases. That is to say, the spatial error model performs best in the goodness of fit. Overall, the spatial autoregressive based model outperforms the OLS model in terms of explanatory power.

Below is a comparison of the parameters simulated by the spatial multi-scale spatial autoregressive model and the GWR model to analyze the advantages and disadvantages of various models as well as the regularities of scaling (Table 8).

In Table 8, the spatial multi-scale model is comparatively analyzed using three modeling parameters:  $R^2$ , AIC and residual square, the results of which are as follows:

(1)  $R^2$  indicates the proportion of the variance of the dependent variable used in the regression model. Besides, the larger the value, the higher accuracy. According to Table 8, at the same scale, the models

are in the following order by magnitude: GWR model>spatial error model>spatial lag model>OLS regression model.

(2) The smaller the AIC value, the closer the model reflects the prediction is closer to the true value, while a larger value is less effective. As can be seen from Table 8, at the scales of 1 km, 5 km and 10 km, AIC is ranked as follows: OLS model>GWR model>spatial lag model>spatial error model; at the scales of 20 km and 30 km, AIC is ranked as follows: OLS model>spatial lag model>spatial error model>GWR model. That is to say, the spatial error model performs best in carrying out simulation at the scales of 1 km, 5 km and 10 km, while the GWR model performs best at the scales of 20 km and 30 km. It can also be found out that the AIC values of each model increase with the scale.

(3) The sum of squared residuals is referred to as the sum of squares of the errors between the actual and measurement values, indicating the accuracy of the simulation by the model. The smaller the value, the higher accuracy of the model, and vice versa. From Table 8, it can be found out that each model tends to have a reduction in the sum of squared residuals as the scale increases. At the scales of 10 km and 30 km, the order of the sum of squared residuals is as follows: OLS model<spatial lag model<spatial error model<GWR model. However, it is exactly the opposite at the scales of 1 km, 5 km and 20 km.

Table 7. Spatial Multi-scale Cultivated Land Goodness Index Comparison.

Indicators	Scale	OLS	SL	SE
LIK	1 km	922.78	1363.54	1473.35
	5 km	661.335	744.321	860.01
	10 km	510.39	604.62	627.59
	20 km	203.30	220.02	224.32
	30 km	94.28	104.56	100.76
AIC	1 km	-1817.56	-2701.08	-2924.70
	5 km	-1439.26	-1956.88	-2051.57
	10 km	-1000.78	-1189.23	-1237.18
	20 km	-392.60	-424.04	-436.64
	30 km	-178.55	-197.13	-195.52
SC	1 km	-1730.42	-2620.16	-2856.23
	5 km	-1254.86	-1844.76	-2011.28
	10 km	-952.39	-1140.84	-1193.63
	20 km	-368.42	-396.39	-415.91
	30 km	-165.48	-181.44	-187.68

Table 8. Spatial Auto-regressive Model and GWR Model Simulation Parameters Comparison.

Indicators	Scale	OLS	SL	SE	GWR
R <sup>2</sup>	1 km	0.40	0.57	0.61	0.65
	5 km	0.42	0.59	0.63	0.66
	10 km	0.48	0.62	0.65	0.66
	20 km	0.58	0.68	0.69	0.75
	30 km	0.55	0.67	0.65	0.68
AIC	1 km	-1817.56	-2701.08	-2924.70	-2564.36
	5 km	-1439.26	-1956.88	-2051.57	-1856.88
	10 km	-1000.78	-1189.23	-1237.18	-1012.62
	20 km	-392.60	-424.04	-436.64	-511.97
	30 km	-178.55	-197.13	-195.52	-244.35
Residual sum of squares	1 km	142.53	133.21	132.93	116.92
	5 km	56.83	47.22	32.54	27.54
	10 km	17.96	18.33	19.53	22.58
	20 km	2.41	2.23	2.01	1.99
	30 km	0.55	0.77	0.91	1.10

In general, with an increase in the scale, the values tend to be smooth and the goodness of fit of the simulation is enhanced. However, it also leads to a considerable difference between the actual location information and the data used in the constructed model, i.e., the residual sum of squares tends to increase.

#### *Residual Spatial Autocorrelation Comparison Analysis*

Allowing for the spatial autocorrelation between the residuals of each model, the residuals of the OLS model have a significant spatial autocorrelation.



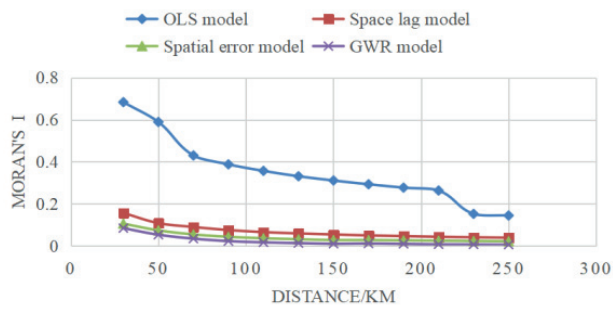


Fig. 5. Residuals of spatial auto-correlation models.

In comparison, the spatial autocorrelation between the residuals of the spatial lag model and spatial error model tends to diminish. In addition, the residuals of the GWR are relatively better than those of the spatial error model and the spatial lag model (Fig. 5). Therefore, the GWR model is the optimal.

#### *Comparison of the Number and Spatial Scale of Impact Factors*

Land use patterns are dependent on scale, the types and driving factors of which show spatial autocorrelation. By analyzing the four types of

cultivated land models in the study area through their influencing factors (Table 9) at multiple scales in space, the following characteristics were identified:

(1) At the same scale, the number of influencing factor variables in each cropland model is OLS, spatial lag model, spatial error model, and GWR in descending order, and the overall number of factors in the classical linear regression model is larger than that in the spatial regression model and geographically weighted regression model.

(2) At different scales, the number of influencing factor variables of the same cultivated land model tends to decrease with the increase of scales, and the coefficients of the spatial lag model and spatial error model also follow this pattern.

(3) Even for the same cultivated land model at different scales, or for different cultivated land models at the same scale, there are significant variations in their influencing factor variables. In addition, there are common influencing factors in different models of cultivated land at the same scale. In general, the number of common influencing variables decreases as the scale increases. For example, at the 1 km scale, the common driving factors are average elevation, average slope, average slope direction, the distance from road, the distance from railroad, the distance from rural

Table 9. Comparison of Impact Factors Number of Cultivated Land Models in Different Scale.

Model	Scale	Variable number	Coefficient	P values
OLS	1 km	13	-	-
	5 km	11	-	-
	10 km	9	-	-
	20 km	6	-	-
	30 km	4	-	-
SL	1 km	11	0.584616	0.00000
	5 km	8	0.572160	0.00000
	10 km	6	0.530487	0.00000
	20 km	4	0.469405	0.00000
	30 km	2	0.448854	0.00000
SE	1 km	10	0.674875	0.00000
	5 km	8	0.668807	0.00000
	10 km	5	0.659153	0.00000
	20 km	2	0.595943	0.00000
	30 km	1	0.553298	0.00000
GWR	1 km	9	-	-
	5 km	8	-	-
	10 km	6	-	-
	20 km	2	-	-
	30 km	1	-	-

road, the distance from ditch, the distance from water system, the distance from village and GDP per farmer. That is to say, topographic factors, traffic conditions and water conditions are the main influencing factors for cultivated land distribution at this scale. Differently, there are 6 common factors at the 5km scale, including average slope, the distance from road, the distance from railroad, the distance from ditch, the distance from village and GDP per farmer. At the 10km scale, there are three common factors: average slope, the distance from road and the distance from ditch. At the scales of 20km and 30km, each model has only one common driving variable, that is, average slope.

It is speculated that the above characteristics are attributable to the spatial pattern of cultivated land distribution and its scale size. The average slope is the only common variable at all scales, indicating that slope is the most significant influencing factor in the distribution pattern of cultivated land in the study area. It is also affected by the characteristics of the fragmented topography of the plateau.

For the selection of driving factors, the dependent variable and the independent variable data of the driving factor used in the OLS model, spatial lag model, and spatial error model are taken as the research object to focus on the global situation. The global model constructed in this study contains a larger amount of information, and the fit of the model produces a better effect on the whole. The GWR is localized, focusing on the local area, which may not lead to a high level of model fit. In spite of this, it performs better in taking into account the local characteristics of land use types and driving factors. Besides, it is advantageous in showing the spatial multi-scale characteristics, local non-smoothness and spatial pattern divergence of the spatial distribution of cultivated land. As revealed by the overall comparison of the advantages and disadvantages

of the models, the GWR model is superior to the spatial error model, the spatial lag model is the second best, and the OLS model is the worst.

### Comparison of Land Use Spatial Distribution Simulation

Based on two modeling methods including OLS Logistic and GWR Logistic, the simulation model was constructed through combination with the CLUE-S model. The spatial distribution patterns of land use in the study area in 2020 were simulated for comparison with the current situation of land use in 2020, respectively (Fig. 6).

According to an analysis of the above figure, the simulated spatial pattern of land use in the study area in 2020 based on a combination of OLS Logistic, GWR Logistic and CLUE-S models is highly consistent with the current land use map in 2020. The Kappa coefficients of OLS and GWR simulations are 0.886 and 0.921 respectively, indicating that the results of the two simulations are coherent with the current state and that the model is effective. Apart from that, both can be used to predict the spatial pattern of land use in the future. In terms of accuracy, the CLUE-S model based on the improved GWR Logistic slightly outperforms the OLS Logistic CLUE-S model. After a further comparative analysis of the probability spatial distribution maps of the two models (Fig. 7), it was discovered that the probability spatial pattern distribution map of land use adaptation as simulated by the GWR Logistic model is more comparable to the real land use pattern distribution than if the simulation was performed by the OLS Logistic regression model. Besides, it performs better in reflecting the characteristics of spatial variation in the extent to which each driver affects land use types (Fig. 8). Based on the analysis shown in Fig. 8,

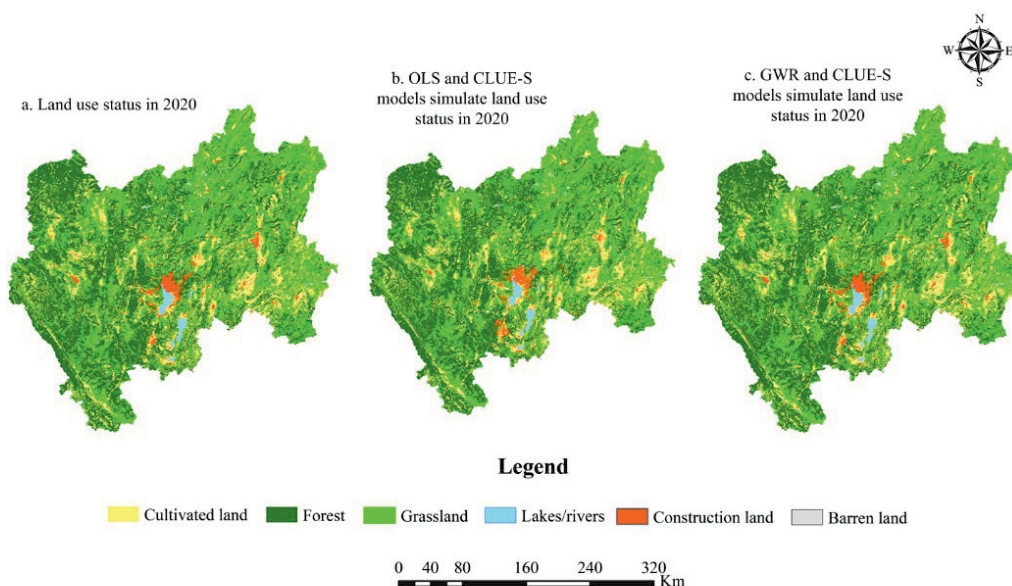


Fig. 6. Status of land use in 2020(a) and simulation (b, c).

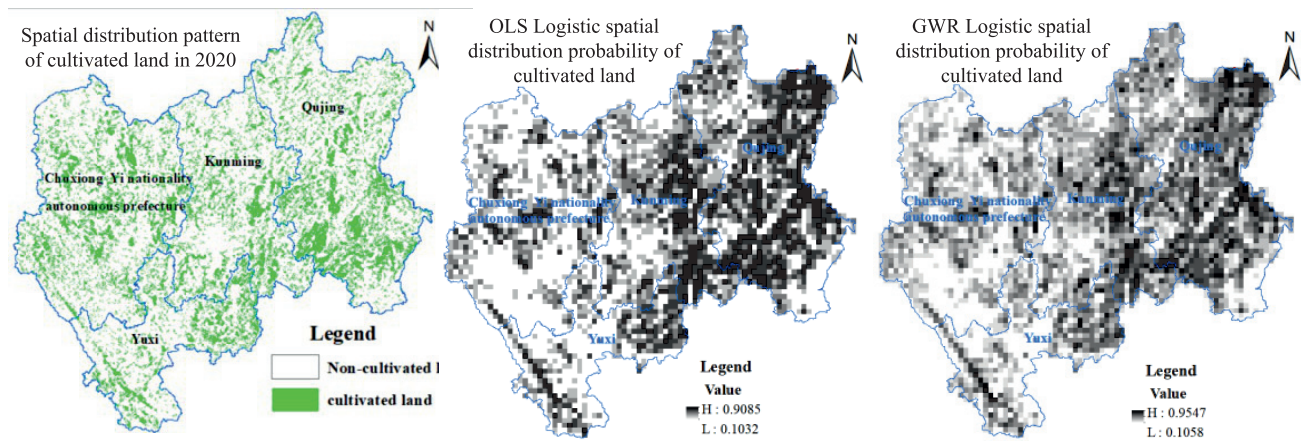


Fig. 7. Comparison results of OLS Logistic and GWR Logistic regression.

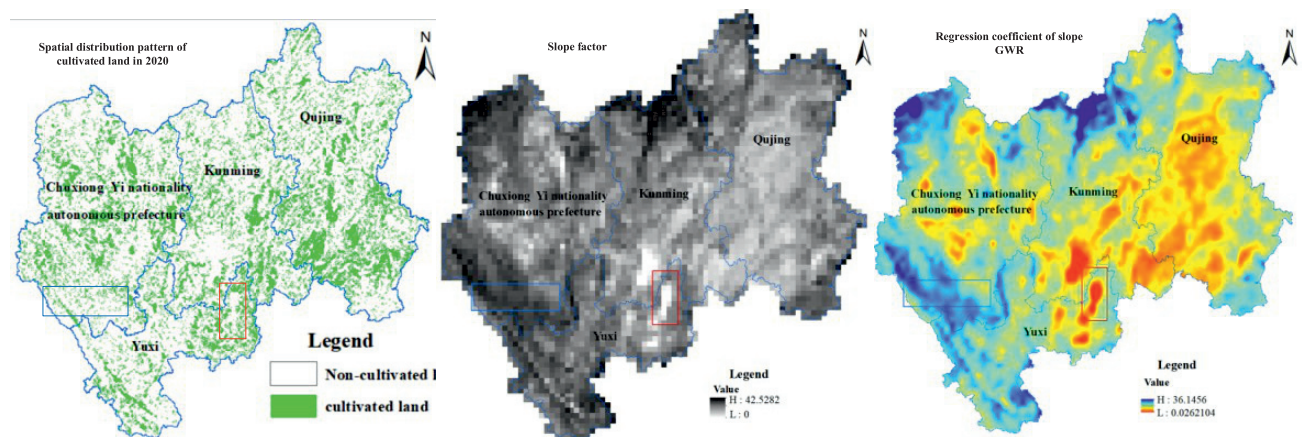


Fig. 8. Spatial differentiation diagram of the impact of slope factor on cultivated land.

the spatial instability of the driving factors was quantified through the regression coefficients of each raster cell. It was found out that the regression coefficient of the slope factor was higher in the blue area, the influence of which on the spatial distribution of cropland was clearly significant. It suggests a significant increase in the aggregation of cropland distribution in the area of gentle slope. Conversely, the regression coefficient of slope factor is low in the red area, and the steepness of slope makes little difference to the distribution probability of cultivated land and the extent to which aggregation diminishes. The spatial variation in the effects of other driving factors on land use types is similar, which will not be reiterated due to space limit.

## Conclusions

Land use is highly scale-dependent, scale-coupled and non-stationary in nature, showing spatial autocorrelation among various land use patterns at different spatial scales. Given different natural

environmental factors and socio-economic factors affecting the formation and evolution of the spatial pattern of cultivated land, it is essential to develop a suitable model for the simulation and prediction as to the spatial distribution characteristics of cultivated land. In this paper, the spatial distribution of cultivated land is simulated and thoroughly analyzed by using a multi-scale model. By taking the Yunnan central urban agglomeration area, which is a typical representative mountainous area with relatively drastic land evolution, as the research object, the main conclusions are drawn as follows:

(1) Spatial autocorrelation analysis is conducted with 16 driving factors on five scales: 1 km, 5 km, 10 km, 20 km and 30 km. It is found out that spatial autocorrelation exists for both multi-scale land use patterns and driving factors, and that the spatial pattern of cultivated land and its driving factors are changing spatially. They show a greater spatial dependence when the weight distance is small and a higher spatial heterogeneity as the weight distance increases. The overall characteristics of cultivated land agglomeration areas are highly consistent at spatial



scale, despite some differences in the specific spatial location and the degree of detail expression. The larger the scale, the more insignificant the agglomeration. Conversely, as the scale decreases, the more obvious the agglomeration and the more significant its local spatial variability.

(2) Comparative analysis of spatial multiscale models is performed using three model parameters:  $R^2$ , AIC and the residual sum of squares. According to the comparison of  $R^2$ , the GWR model achieves high accuracy and the OLS regression model performs worst in accuracy at the same scale. Besides, the overall accuracy is improved with the increase of scale. According to the comparison of AIC, the spatial error model produces the best simulation results at the scales of 1 km, 5 km and 10 km, while the GWR model performs best at the scales of 20 km and 30 km. Besides, the AIC value of each model rises with the increase of scale. As revealed by the comparison of the residual sum of squares parameters, as the scale increases, the accuracy of all models is decreasing. At the scales of 10km and 30km, the order of the residual sum of squares is as follows. The OLS model is the best, followed by the spatial lag model and the spatial error model, while the GWR model is the worst. At the scales of 1 km, 5 km and 20 km, it is exactly the opposite.

(3) Two models, namely, the spatial autoregressive model and the geographically weighted regression model, are adopted to perform simulation for predicting the spatial distribution pattern of cropland and the change of cropland. Since the local spatial object of cropland is taken as the study unit in the GWR model, the spatial local characteristics of cropland are more detailed to a certain extent. By taking into account the geospatial local characteristics and geospatial variability, the goodness of fit achieved by different spatial location models can be determined. Besides, it is possible to obtain the weights of each driver of land use with the change of spatial location. By comparing the fit and residual sum of squares of the spatial autoregressive model, it can be found out that the GWR model performs well in terms of local spatial characteristics, simulation accuracy and driving factor coefficients. By using this model, the spatial pattern of cropland in the study area in 2020 is simulated and compared with the actual distribution of cropland in 2020. The Kappa coefficient reaches 87.68%, which is 2.35 percent higher than that of the traditional OLS Logistic probability model. It evidences the applicability of the proposed model.

Since LUCC is a highly complex system, the study and identification of its model and driving factors have become a research hotspot. This paper is based on the selection of different models to explore the spatial distribution characteristics and factors in the evolution of cropland. Given many factors affecting the distribution and evolution of cropland in different time periods, the conclusions drawn in this paper are limited

to partially reflecting the characteristics of cropland. Besides, different research scales can lead to variations in the magnitude of impact on the construction of models as well as the simulation and prediction of cropland. In addition, the formation and evolution of cultivated land are affected by many socio-economic factors and national policies, which makes it difficult to reflect them quantitatively to a certain extent. Therefore, the reasonable selection of appropriate influencing factors is crucial to the construction of the model. Therefore, the selection of scale and the quantification of influencing factors must be considered for the construction of model, so as to provide an effective theoretical reference and technical support for the protection of cultivated land and the formulation of national policy.

## Acknowledgments

The authors gratefully acknowledge the financial support from Yunnan philosophy and social science planning project (PY202129).

## Conflict of Interest

The author declare no conflict of interest.

## References

1. SUN X.L., ZHAO R., HE W.R., ZHENG Y., YUAN L., GUO C.Y. Analysis on Grassland Resources Change and Driving Forces in Huize County Based on Logistic Regression Model. *Southwest Journal of Agriculture*, **6**, 1309, **2021**.
2. LUO G.P., ZHANG A.J., YIN C.Y. Progress and Prospects of Multi-scale Evaluation of Land Change. *Arid Zone Research*, **2**, 187, **2009**.
3. MA S.F., CAI Y.M., NIAN P.H. Models of Land Use Planning: A Review. *China Land Science*, **3**, 34, **2014**.
4. WEI X.J., JIANG P., MA G.P. Simulation on Spatial Land Use Patterns Based on Different Logistic Models: A comparative study of Xuan'en and Xianfeng Counties, Hubei Province. *Scientific and Technological Management of Land and Resources*, **1**, 93, **2016**.
5. WANG X.L., LIU Y., ZHANG Y., LIU J.P. Exploration and prediction of land use cover change in western Jilin province based on CA-markov model. *Science Technology and Engineering*, **21**(19), 7942, **2021**.
6. LI S.F., HONG Z.L., XUE X.P., ZHANG F.J., SHI W. Multi-Scenario Simulation of LUCC in Binzhou City Based on Logistic-CA-Markov Coupling Model. *Research of Soil and Water Conservation*, **29** (04), 292, **2022**.
7. ZHANG X., GU R.X. Spatio-temporal pattern and multi-scenario simulation of land use conflict: A case study of the Yangtze River Delta urban agglomeration. *GEOGRAPHICAL RESEARCH*, **41** (5), 1311, **2022**.
8. DENG Y.J., HOU M.Y., ZHANG X., JIA L., LI Y.Y., YAO S.B., GONG Z.W., LIU G.Q. Drivers of forestland change in the Qinba Mountain region of Shaanxi based on the

- Logistic regression model. *Journal of Nanjing Forestry University (Natural Sciences Edition)*, **46** (1), 106, **2022**.
9. WU Y.W., LIU G. Simulation and prediction of land-use change in Nanjing based on artificial bee colony cellular automata model. *Bulletin of Surveying and Mapping*, **02**, 95, **2022**.
  10. ZHAO M.S., XU S.J., DENG L., LIU B.Y., WANG S.H., WU Y.J. Simulation of Land Use Change in Typical Coal Mining City Based on CLUE S Model. *Transactions of the CSAE*, **53** (05), 158, **2022**.
  11. WANG X.R., PAN P.P., WANG X.X., WANG X.M. Simulation of landscape pattern for land use in Hebei province based on GeoSOS-FLUS model. *Jiangsu Journal of Agricultural Sciences*, **37** (03), 667, **2021**.
  12. HU Y.T., LI T.H. Forecasting Spatial Pattern of Land Use Change in Rapidly Urbanized Regions Based on SD-CA Model. *Acta Scientiarum Naturalium Universitatis Pekinensis*, **58** (02), 372, **2022**.
  13. ZHU Y., LIANG D.D., XU Y.T. Decision Process, Mechanism and Simulation of Farmers' Farmland Transfer Behavior Based on MAS Model: A Case Study of Anhui Province. *Areal Research and Development*, **39** (05), 126, **2020**.
  14. QUAN B. Introduction to Land Use and Land Cover Change Science. Beijing: China Environmental Science Press, **2010**.
  15. CHEN G.P. Multi-scale spatio-temporal pattern evolution and dynamic simulation of land use. Kunming University of Science and Technology, Kunming, China, **2018**.
  16. LI X., ZHOU J.C., JIN T.T., WANG J.L. LUCC spatial simulation of urban agglomeration in Central Yunnan. *Journal of Ecology and Rural Environment*, **38** (06), 1, **2022**.
  17. GOODCHILD M.F. Spatial Autocorrelation. Concepts and Techniques in Modern Geography. Norwich, UK: Geo Books, **1986**.
  18. CHEN G.P., ZHAO J.S., LI H.B., WU X.W. Analysis Method of Temporal and Spatial Changes of Cultivated Land in Yunnan Central Urban Agglomeration Area. *Journal of Kunming University of Science and Technology (Natural Science Edition)*, **41** (02), 24, **2016**.
  19. LIN L., FAN H., JIN Y. Multi-Scale and Multi-Model Simulation of Land Use /Land Cover Change in the Mountainous County: A Case Study of Mengla County in Yunnan Province, China. *MOUNTAIN RESEARCH*, **38** (04), 630, **2020**.
  20. ZHANG W.B., ZHANG Z.B., DONG J.H., ZHANG H.L., GAO F.W., GONG W.M. Transformation and driving forces of cultivated land utilization function from a multi-scale perspective in Gansu Province. *Scientia Geographica Sinica*, **41** (5), 900, **2021**.
  21. TURNER M.G. Spatial and temporal analysis of landscape patterns. *Landscape Ecology*, **4** (1), 21, **1990**.
  22. WANG Y.F., HE H.L. Spatial data analysis methods. Beijing: Science Press, **2007**.
  23. ZHANG H., ZHAO X. M., OU YANG Z.C., GUO X., LI W. F., KUANG L. H., YE Y. C., HUANG C., WANG X. Y. Multi-scale spatial autocorrelation analysis of cultivated land quality in China's southern hillside areas: A case study of Lichuan County, Jiangxi Province. *Chinese Journal of Eco-Agriculture*, **26** (2), 263, **2018**.
  24. YAN C., HUAN J.L. Multi-scale spatial autocorrelation analysis of cultivated land quality in Da'an city. *Rural economy and science & technology*, **31** (10), 8, **2020**.
  25. WANG J.F., LI L.F., GE Y. A theoretical framework for spatial analysis of geographic information. *Journal of Geography*, **1**, 92, **2000**.
  26. ZHANG S.L., ZHANG K.K., Zhang S. L. Comparison between General Moran's Index and Getis-Ord General G of Spatial Autocorrelation. *Journal of Zhongshan University (Natural Science Edition)*, **4**, 93, **2007**.
  27. DU W.T., LI S.J., CAO J.W., Qi L. Spatial autocorrelation analysis of cultivated land quality in Hunchun city under multi-scale. *Journal of Northeast Normal University (Natural Science Edition)*, **50** (04), 134, **2018**.
  28. DING Y.N., LIU H.L., WANG W.Q. Spatial-temporal evolution and spatial correlation analysis of ecosystem services based on grid in Changzhi. *Journal of Shanxi University of Technology (Natural Science Edition)*, **37** (04), 85, **2021**.
  29. FAN Q., GUO A.J. Research Progress on Geographical Weighted Regression Models: A Perspective of Literature Reviews. *Journal of Quantitative Economics*, **12** (02), 134, **2021**.
  30. YANG Y.Y., YAO Y., HAO S. Spatial-Temporal Variations and Influencing Factors of Intensified Cultivated Land Use Level in Guizhou Province Based on GWR Model. *Research of Soil and Water Conservation*, **29** (1), 326, **2022**.
  31. CHEN A.N. Introduction to spatial metrology and GeoDa software application. Hangzhou: Zhejiang University Press, **2014**.