J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 9137--9156

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Quantitative analysis of the impact of urban sprawl on land resources: algorithm-based spatio-temporal modeling

Xiaoxiao Wang[1,✉], Zhenchen Lin[1], Xuanyi Li[1], Benzheng Zhang[1]

[1] *College of Architecture and Civil Engineering, Chongqing Metropolitan College of Science and Technology, Chongqing, 402167, China*

ABSTRACT
The global urbanization process is rapidly increasing, and a reasonable and scientific analysis of the relationship between urban land expansion and land resources plays an important role in the rational allocation and coordinated development of land resources. This paper constructs a spatio-temporal geographic weighted regression model coupled with geospatial and temporal coordinates, and incorporates temporal and spatial non-stationarity into the model. Then, using the method of hypothesis testing, the temporal non-stationarity and spatial non-stationarity of the spatio-temporal geographic weighted regression model are examined, and at the same time, the multiple covariance test and the variance expansion factor method are proposed to carry out further statistical inference of the model. As the degree of urban sprawl increases, the land resources weaken year by year from the center to the surrounding area.The global Moran's I for the three periods from 2003 to 2023 are 0.6289, 0.7159, and 0.7368, respectively, which show a trend of increasing year by year. It shows that land resources are strongly influenced by urban expansion, and the spatial distribution of land resources shows spatial aggregation. Several variables, such as building volume rate, population size, regional economic development, regional cultural level, infrastructure construction and urban fallow area, have significant effects on the spatial differentiation of land resources. The above differentiation characteristics provide insights into the rationalization of urban expansion and the scientific allocation of land resources.

Keywords: land resource allocation, spatio-temporal geographically weighted regression, variance inflation factor method, urban sprawl

1. Introduction

With the continuous growth of population, urban expansion has become a real problem in today's society. Urban expansion has a far-reaching impact on land resource utilization, involving land

✉ Corresponding author.
E-mail address: wangxiaoxiao_1216@126.com (X. Wang).
Received 01 February 2024; Revised 20 April 2024; Accepted 10 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-501

occupation, environmental quality, and sustainable development [1-3].

First of all, urban expansion leads to a large amount of land occupation. With the expansion of cities, land that was originally used for farmland, forests and other natural ecosystems has been forced to be used for the construction of residences, commercial districts and industrial land, etc. This occupation not only directly affects the stability of the ecological environment, but also leads to the waste of land resources [4-7]. Especially, some big cities have to develop the neighboring farmland or mountainous areas in order to meet the demand of population growth, which leads to the reduction of farmland and ecological damage [8-9]. Secondly, urban expansion has brought about unbalanced utilization of land resources. In the process of urban expansion, some rich land resources may be prioritized, while areas that cannot be developed are neglected. Such unbalanced utilization not only causes waste of land resources, but also leads to unbalanced urban development [10-13]. In some cities, resources are over-utilized in the central areas while the surrounding areas are relatively poor, thus exacerbating the gap between cities. In addition, urban expansion has a negative impact on the efficiency of land use. In the process of urban expansion, the utilization efficiency of land is often low due to asymmetric information and imperfect policies [14-15]. Some land has been used for inefficient construction projects, and there is also the phenomenon of large areas of vacant land. This not only wastes valuable land resources, but also brings challenges to the sustainable development of cities [16-17].

vanVliet, J. emphasizes that forest loss due to displacement of cropland exceeds that caused by urban expansion. This was verified with global data. It is pointed out that guiding urban development towards a sustainable trajectory is conducive to the protection of forests and other natural areas [18]. Wu,Y. et al. used Landsat images to assess urban expansion and LULC changes from 1979 to 2013. It was pointed out that urban growth had a serious impact on land cover. And taking Guangzhou as an example, they emphasized that the obvious changes in LULC and urban sprawl were highly correlated with indicators such as economic development and population growth [19]. Colsaet, A. summarizes the relationship between land occupation and different explanatory factors based on a literature review. It was pointed out that demographic, income, political and institutional factors were mentioned, suggesting that urban sprawl is influenced by public policy in addition to being the result of “market forces” [20]. Wei,Y.D. analyzed the relationship between urban sprawl and equitable and sustainable development. It also explored the pattern, structure and dynamics of urban sprawl and its impact on spatial inequality from various economic and social dimensions [21]. Brend'Amour, C. combines spatially explicit projections of urban expansion with datasets of global farmland and crop yields. The results of the study illustrate that urban sprawl will cause a global loss of cropland of 1.8-2.4% by 2030, with significant regional variations. And the loss of arable land will be accompanied by other risks that pose a threat to livelihoods [22]. Şen, G. aims to determine the change in land cover in a city from 1999-2014. This change was assessed based on relevant data from the last 15 years and using remote sensing methods. The results showed that agricultural and forested areas became artificial land surface during many years [23]. Song,X. et al. explored the quantitative and qualitative correlations between ecosystem service functions and urban expansion indicators in Wuhan using spatial autocorrelation analysis and linear regression. It was pointed out that urban expansion had the greatest impact on food production, and the expansion area was an important factor contributing to the decline of ecosystem service function [24]. Huang,Q. et al. analyzed the global urban expansion and its impact on cropland NPP from 1992-2016 based on urban expansion data. The results showed that urban expansion caused a sharp decline in the area of arable land and led to food security risks in developing countries [25].

This paper mainly applies the comprehensive evaluation method to measure the effective degree of land resource allocation, and adopts the index of effective degree of land resource allocation to quantify the level of land resource allocation. On this basis, for the traditional spatio-temporal model that does not take into account the spatial and temporal changes of the data, a spatio-temporal geographically weighted regression model that integrates temporal and spatial features is proposed, which further grasps the spatio-temporal non-stationarity of the land resource allocation data. Chinese cities were selected as examples for empirical analysis, and the law of urban expansion on the spatio-temporal changes of land resources was explored. Eight variables are selected as the factors influencing the spatial differentiation of land resource allocation for research and analysis.

2. Methodology for quantifying land resources

2.1. Comprehensive evaluation method

This part mainly applies the comprehensive evaluation method to measure the effective degree of land resource allocation [26], and the formula of the comprehensive evaluation method is as follows:

$$S_i = \sum_{j=1}^{n} W_j z_{tij} \tag{1}$$

where S_i is the comprehensive evaluation index, which represents the degree of effectiveness of land resource allocation in the evaluation of land resource allocation, respectively. W_j is the weight of each index; z_{tij} is the standardized value of each city in each index.

On the determination of indicator weights, in the evaluation of land resource allocation effectiveness, since there are only two indicators in the evaluation of land resource allocation effectiveness, and this paper believes that these two indicators are equally important to regional land resource allocation, and there is no difference between the advantages and disadvantages; therefore, the coefficients to be determined are determined to be equal weights.

Entropy value method is an objective empowerment method, which determines the weights based on the degree of dispersion of the indicators. If the greater the degree of data dispersion, the smaller the information entropy, the greater the amount of information provided by the indicator, the greater the impact on the comprehensive evaluation results, so its weight in the comprehensive evaluation should be greater. Most of the existing researches use the traditional entropy method, i.e., they can only calculate the two-dimensional table, which has certain limitations. On this basis, the global entropy value method introduces the time sequence in the general cross-section data, and makes the data comparable among years by establishing the three-dimensional data of "region, time and indicator". Therefore, this paper introduces the global entropy method to determine the weights of each indicator, which helps to dynamically analyze the new urbanization level, the effective degree of land resource allocation and the level of regional high-quality development of Chinese cities and municipalities. The specific steps for determining the weights of each index using the global entropy method are as follows:

1) Establishment of three-dimensional time series data table. There are m municipality, n evaluation indicators, and T years, which together form the initial global evaluation matrix:

$$X = \left\{ x_{nj} \right\}_{mT \times n} \tag{2}$$

where x_{tij} represents the j th indicator for the i rd municipality in year t.

2) Data standardization. The polar value method was used to standardize the indicators to avoid the interference of the inter-indicator scale and positive and negative orientation. At the same time, in order to eliminate the influence of zero on the subsequent model, the standardized data were distributed between 0.1 and 0.9. The specific standardization formula is shown below:

Positive indicators:

$$z_{iij} = 0.1 + 0.8\left(\frac{x_{iij} - Min(x_j)}{Max(x_j) - Min(x_j)}\right) \tag{3}$$

Negative indicators:

$$z_{tij} = 0.1 + 0.8(1 - \frac{x_{tij} - Min(x_j)}{Max(x_j) - Min(x_j)}) \tag{4}$$

where x_{tij} is the j th indicator for the i rd municipality in year t, z_{tij} is the normalized value of x_{tij}, and $Max(x_j)$ and $Min(x_j)$ are the maximum and minimum values of the j th indicator for all municipalities.

3) Indicator normalization.

$$P_{tij} = \frac{z_{tij}}{\sum_{t=1}^{T}\sum_{i=1}^{m} z_{tij}} \tag{5}$$

4) Calculate the entropy value.

$$E_j = -k\sum_{t=1}^{T}\sum_{i=1}^{m} P_{tij} \ln P_{tij}, k = \frac{1}{\ln(T \times m)} \tag{6}$$

5) Calculate the redundancy of each indicator.

$$D_j = 1 - E_j \tag{7}$$

6) Determination of weights.

$$W_j = \frac{D_j}{\sum_{j=1}^{n} D_j} \tag{8}$$

2.2. Land Resource Allocation Effectiveness Index (LREI)

The primary issue in quantitative research on land resource allocation is to rationally quantify the level of land resource allocation. Existing studies usually equate agreement transfer with industrial land, and measure land resource allocation by the ratio of the area of land transferred by agreement to the total area of land transferred. Land resource mismatch is the distorted behavior of urban industrial land and commercial land supply. Combined with existing studies, in order to reduce the problem of bias in research results caused by a single variable, this paper decomposes the degree of land resource allocation effectiveness into the degree of quantitative allocation effectiveness and price allocation effectiveness, and portrays the degree of land resource quantitative allocation effectiveness by the ratio of the area of land granted by way of bidding and auctioning to the total area of land granted. The ratio of the unit price of the land granted by agreement to the unit price of the land granted by auction is used to measure the effective degree of price allocation of land resources. The specific formula for calculating the effective allocation of land resources is as follows:

$$La = a \cdot La - q + b \cdot La - p \tag{9}$$

where La is the effectiveness of land resource allocation; $La - q$. Allocate the effectiveness of the quantity of land resources; $La - q$. Allocate the effectiveness of land resource prices; a, b are the coefficients to be determined, considering that both the effectiveness of quantity allocation and the effectiveness of price allocation have a positive impact on the effectiveness of regional land resource allocation, and both are equally important, and there is no difference between advantages and disadvantages. Therefore, the pending coefficients for quantity allocation validity and price allocation validity are determined to be equally weighted, i.e., $a = b = 1/2$.

3. Spatio-temporal geographically weighted regression models

3.1. *Definition of spatio-temporal geographically weighted regression models*

The spatio-temporal geographically weighted regression model [27] combines the original geospatial coordinates with temporal coordinates to form the spatio-temporal three-dimensional coordinates (u_i, v_i, t_i). The mathematical expression of the model is as follows:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i, t_i) x_{ik} + \varepsilon_i \quad i = 1, 2, \cdots, n \tag{10}$$

where (u_i, v_i, t_i) is the spatio-temporal 3D coordinates of the i nd sample point, n is the number of sample points, $\beta_k(u_i, v_i, t_i)$ is the regression coefficient of the k th independent variable for the i th sample point. ε_i is the random error of the i th sample point, $\varepsilon_i \sim N(0, \sigma^2)$ and the random errors of the different sample points i and j are independent of each other with a covariance of zero.

3.2. *Estimation of spatio-temporal geographically weighted regression models*

The estimation of the spatio-temporal geographically weighted regression model is similar to that of the geographically weighted regression model, and the regression coefficients can be estimated by the following equation, based on the weighted least squares method:

$$\min \sum_{j=1}^{n} w_{ij} (y_j - \beta_{i0} - \sum_{k=1}^{p} \beta_{ik} x_{ik})^2 \tag{11}$$

Here w_{ij} is the spatio-temporal weight matrix whose value decreases as the distance between the sample point and the data point increases.

The regression coefficient valuation $\hat{\beta}_i$ for the i nd sample point is:

$$\hat{\beta}_i = (\hat{\beta}_0 \hat{\beta}_1 \ldots \hat{\beta}_d)^T = (X W_i X)^{-1} X W_i y \tag{12}$$

The fitted value of dependent variable Y at point i is:

$$\hat{y}_i = X_i \hat{\beta}_i = X_i (X W_i X)^{-1} X W_i y \tag{13}$$

Let $\hat{Y} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n)^T$ be the vector consisting of the fitted values of $y_1, y_2, \ldots, y_n$ and $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n)^T$ be the corresponding residual vector, then:

$$\begin{cases} \hat{Y} = SY \\ \hat{\varepsilon} = Y - \hat{Y} = (I - S)Y \end{cases} \tag{14}$$

Among them:

$$S = \begin{bmatrix} X_1 (X^T W_1 X)^{-1} X^T W_1 \\ X_2 (X^T W_2 X)^{-1} X^T W_2 \\ \cdots \\ X_n (X^T W_n X)^{-1} X^T W_n \end{bmatrix} \tag{15}$$

The residual vector e is the difference between the observed and fitted values:

$$e=\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} - \begin{bmatrix} S_1 \\ S_2 \\ \cdots \\ S_n \end{bmatrix} y = (I-S)\,y \tag{16}$$

The residual sum of squares RSS is:

$$RSS = e^{\mathrm{T}} e = \left[(I-S)y\right]^{\mathrm{T}}\left[(I-S)y\right] = y^{\mathrm{T}}(I-S)^{\mathrm{T}}(I-S)y \tag{17}$$

In general the lower $tr(S)$ is so close to $tr(S^T S)$ that the effective degrees of freedom can be approximated as $n-tr(S)$, thus leading to a simplified formula:

$$\sigma^2 = \frac{RSS}{n-tr(S)} \tag{18}$$

3.3. Spatio-temporal weight function and bandwidth

The core of the spatio-temporal geographic weighted regression model is the spatio-temporal weight matrix, and different spatio-temporal weight functions are selected to describe the spatio-temporal relationship between data. The mutual combination of temporal and spatial information can be realized by the Gaussian function method to establish the spatio-temporal weight function of the spatio-temporal geographic weighted regression model.

The spatio-temporal distance can be expressed as a linear combination of temporal distance d_{ij}^f and spatial distance d_{ij}^S :

$$d_{ij}^{ST} = \sqrt{\lambda (d_{ij}^S)^2 + \mu (d_{ij}^T)^2} \tag{19}$$

where for λ and μ are scaling factors indicating the balance between spatial and temporal distances. d_{ij}^f is the temporal distance between sample points i and j, Eq:

$$(d_{ij}^T)^2 = (t_i - t_j)^2 \tag{20}$$

d_{ij}^S is the spatial distance, Eq:

$$(d_{ij}^S)^2 = (u_i - u_j)^2 + (v_i - v_j)^2 \tag{21}$$

Thus the spatio-temporal distance can be expressed as:

$$d_{ij}^{ST} = \sqrt{\lambda\left[(u_i - u_j)^2 + (v_i - v_j)^2\right] + \mu (t_i - t_j)^2} \tag{22}$$

The weights at row i and column j are calculated by Gaussian function:

$$
\begin{aligned}
w_{ij}^{ST} &= \exp\left[-\left(\frac{d_{ij}^{ST}}{h^T}\right)^2\right] = \exp\left\{-\frac{\lambda\left[\left(u_i-u_j\right)^2+\left(v_i-v_j\right)^2\right]+\mu(t_i-t_j)^2}{\left(h^{ST}\right)^2}\right\} \\
&= \exp\left\{-\left[\frac{\left(u_i-u_j\right)^2+\left(v_i-v_j\right)^2}{\left(h^S\right)^2}+\frac{\left(t_i-t_j\right)^2}{\left(h^T\right)^2}\right]\right\} \\
&= \exp\left\{-\left[\frac{\left(d_{ij}^S\right)^2}{\left(h^S\right)^2}+\frac{\left(d_{ij}^T\right)^2}{\left(h^T\right)^2}\right]\right\} \\
&= \exp\left\{-\frac{\left(d_{ij}^S\right)^2}{\left(h^S\right)^2}\right\}\exp\left\{-\frac{\left(d_{ij}^T\right)^2}{\left(h^T\right)^2}\right\} \\
&= w_{ij}^S \times w_{ij}^T
\end{aligned}
\tag{23}
$$

where h^S is the spatial bandwidth, h^T is the temporal bandwidth, and h^{ST} is the spatio-temporal bandwidth, which is linearly related to the two, respectively:

$$(h^S T)^2 = \lambda(h^S)^2, (h^{ST})^2 = \mu(h^T)^2 \tag{24}$$

4. Statistical inference of spatio-temporal geographically weighted regression models

Statistical inference is the estimation and making of inferences about one or some characteristics of the population by analyzing data, models and the reasonableness of the inference method, including parameter estimation and hypothesis testing.

4.1. *Tests for spatio-temporal non-stationarity*

The goodness-of-fit can be used as an evaluation metric to verify whether the geographically weighted regression model is better than the ordinary linear regression model.

Based on the above analysis, the original hypothesis H_0 and alternative hypothesis H_1 are constructed as follows:

$$H_0 : y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \; i = 1,2,\cdots,n \tag{25}$$

$$H_1 : y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad i = 1,2,\cdots,n \tag{26}$$

If the original hypothesis is true, it proves that the model is not spatially non-stationary. If the alternative hypothesis is true, the original hypothesis is rejected, indicating that the model is spatially smooth, i.e., the regression relationship varies with geographic location.

From the previous section, the residual sum of squares RSS_H of the general linear regression model is calculated as:

$$RSS_H = \sum_{i=1}^{n} e_i^2 = e^T e = y^T (I-H)^T (I-H) y \tag{27}$$

The residual sum of squares and RSSs for the geographically weighted regression model with embedded spatial location are calculated as:

$$RSSs = \sum_{i=1}^{n} e_i^2 = e^T e = y^T (I-S)^T (I-S) y \tag{28}$$

Construct the test statistic $F_i = \dfrac{RSS_H - RSS_S}{RSS_c} = \dfrac{y^{'}(R_H - Rs)y}{v^{'}Rsv}$, where R_H and R_s are both semipositive definite real symmetric matrices.

$$R_H = (I - S_H)^T (I - S_H) \tag{29}$$

$$R_S = (I - S_S)^T (I - S_S) \tag{30}$$

Under the condition of $E(\hat{y}_s) = E(y), E(\hat{y}_H) = E(y)$

$$F_1 = \frac{\varepsilon'(R_H - R_s)\varepsilon}{\varepsilon' R_s \varepsilon} \tag{31}$$

If the original hypothesis H_0 is not true, there is a tendency for F_i to be biased toward the larger, H_0 test p value for alternative hypothesis H_i:

$$p_1 = P_{H_0}\left(F_1 > f_1\right) = P_{H_0}\left(\varepsilon'(R_H - (f_1+1)R_S)\varepsilon > 0\right) \tag{32}$$

where f_1 is the observed value of the test statistic, given a significance level of α. If p_1 is greater than α, the original hypothesis is accepted, i.e., the geographically weighted regression relationship is smooth with respect to changes in spatial location. Otherwise, if p_1 is greater than α, spatial non-stationarity is considered to exist.

Whether the spatio-temporal geographically weighted regression model is superior to the geographically weighted regression model is also verified here using goodness-of-fit.

Based on the above analysis, the original hypothesis H_0 and alternative hypothesis H_1 are constructed as follows:

$$H_0 : y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad i = 1, 2, \cdots, n \tag{33}$$

$$H_1 : y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad i = 1, 2, \cdots, n \tag{34}$$

If the original hypothesis is true, the model is shown not to be time non-stationary. If the alternative hypothesis is true, the model has temporal smoothness, i.e., the regression relationship changes over time.

Under the condition that the original hypothesis is true, the residual sum of squares RSS_s of the geographically weighted regression model is obtained and calculated as:

$$RSSS_1 = \sum_{i=1}^{n} e_i^2 = e^{\mathrm{T}} e = y^{\mathrm{T}} (I - S_1)^{\mathrm{T}} (I - S_1) y \tag{35}$$

The residual sum of squares RSSs of the spatio-temporal geographically weighted regression model are obtained and calculated as under the condition that the alternative hypothesis holds:

$$RSSS_2 = \sum_{i=1}^{n} e_i^2 = e^{\mathrm{T}} e = y^{\mathrm{T}} (I - S_2)^{\mathrm{T}} (I - S_2) y \tag{36}$$

Constructing test statistics:

$$F_2 = \frac{RSS_{S_1} - RSS_{S_2}}{RSS_{S_1}} = \frac{y'(R_{S_1} - R_{S_2})y}{y'R_{S_1}y} \tag{37}$$

Under the condition of $E(\hat{y}_{S_1}) = E(y), E(\hat{y}_{S_2}) = E(y)$:

$$F_2 = \frac{\varepsilon'(R_{S_1} - R_{S_2})\varepsilon}{\varepsilon' R_{S_1}\varepsilon} \tag{38}$$

If the original hypothesis H_0 is not true, then F_2 there is a tendency for it to be large, otherwise F_2 there is a tendency for it to be small, and H_0 the value of test p for alternative hypothesis H_1 is:

$$p_2 = P_{H_0}(F_2 > f_2) = P_{H_0}(\varepsilon'(R_H - (f_2+1)R_{S_1})\varepsilon > 0) \tag{39}$$

The F-distribution approximation can be used to reduce the difficulty of the operation. The F distribution approximation method means that the distribution of the value of the statistic is approximated by the F distribution. The distribution of the quadratic form of the normal variable in the numerator and denominator of the test statistic F is approximated by taking the appropriate multiplicity and degree of freedom of the χ^2 distribution, and the distribution of the test statistic F is approximated by using the F distribution with the corresponding degree of freedom.

Let $Q = \xi^T \mathrm{A}\,\xi, \xi \sim N(0,1)$, A denote semipositive definite real symmetric matrices of order n. If A is an idempotent matrix, then Q obeys a χ^2 distribution with $tr(A)$ degrees of freedom. A common treatment is to approximate Q by the distribution of $b\chi^2$. The constants b and r can be computed so that the variances of matrices A and $b\chi_r^2$ are the same as the expectation.

Since $E(b\chi_r^2) = br, Var(b\chi_r^2) = 2b^2 r$, and $E(Q) = tr(A), Var(Q) = 2tr(A^2)$, it follows that $br = tr(A), 2b^2 r = tr(A^2)$ yields $r = \frac{[tr(A)]^2}{tr(A^2)}, b = \frac{tr(A^2)}{tr(A)}$.

For the ratio of quadratic forms of normal variables $F = \frac{1}{\xi}\xi\xi^T A_2\xi, \xi \sim N(0,1)$. in the case of A_1 and A_2 idempotents and $A_1 A_2 = 0$, F obeys a F distribution with degrees of freedom $tr(A_1)$ and $tr(A_2)$, and in the usual case the distribution of $b_i\chi_{r_i}^2$ can be de-approximated to the distribution of $\xi A \xi$ first, $i = 1,2$. And then let:

$$\tilde{F} = \frac{\left(\frac{1}{b_1}\xi^T A_1 \xi\right) \Big/ \sqrt{r_1}}{\left(\frac{1}{b_2}\xi^T A_2 \xi\right) \Big/ \sqrt{r_2}} = \frac{tr(A_2)}{tr(A_1)} F \tag{40}$$

is the $\tilde{F}$-distribution approximated by the F-distribution with degrees of freedom r_1 and r_2.

4.2. *Multicollinearity test*

Multicollinearity refers to the highly correlated relationship between two or more explanatory variables in a linear regression model. When multicollinearity occurs among the variables in the model, the estimation of the model will be distorted or difficult to estimate accurately.

1) The economic meaning of the parameter estimates is unreasonable. If the two explanatory variables in the model appear to be correlated, $x_2 = \lambda x_1$, the parameter β_1, β_2 at this point does not respond to its relationship with the explanatory variable $x_1 x_2$, but to their joint influence with the explanatory variable $x_1 x_2$.

2) The significance test of the variables loses its significance. When there are multiple covariances in the model, it will lead to the variance and standard deviation of the estimated values of the model parameters to become larger, and therefore the explanatory variables that originally existed will be excluded.

3) Failure of the predictive function of the model. Multicollinearity causes the variance of the model parameters to become larger, which leads to the failure of prediction. Therefore, when analyzing with the model, it is necessary to perform a multicollinearity test first, and the most commonly used method is the variance inflation factor (VIF) [28].

The inverse matrix M of the correlation matrix of the variance inflation factor method is:

$$M = (X'X)^{-1} \tag{41}$$

The variance inflation factor v_i for variable x_i is:

$$v_i = m_{ii} = \frac{1}{1-(R_i)^2} \tag{42}$$

The VIF method assumes that if the value of the variance inflation factor is larger, the stronger the covariance is and the regression analysis can be performed. Conversely, the smaller the variance inflation factor, the weaker the multicollinearity, indicating that it is not suitable for analysis.

5. Analysis of the efficiency of land resource allocation

Stata software was used to measure the value of land resource allocation efficiency. Since agricultural land is mostly defined as carbon sink land, which has less correlation with the apparent CO2 selected in this paper, only the factor efficiency of urban villages and industrial, mining and transportation land is analyzed here. The time-series changes of the mean values are shown in Figure 1. From the overall time series change, the mean value of factor efficiency of land resource allocation in China's provinces under the carbon emission constraint is higher than the total efficiency mean value during the study period, and all of them show a two-stage change, the first half of which is a wave-type dynamic change, and the second half of which shows a downward trend, but with different demarcation points for the downward turn.

Specifically:

1) The average value of total efficiency decreases from 0.725 in 2003 to 0.471 in 2020, which is contrary to the conclusion that the green land use efficiency of urban land in China shows an upward trend from 2003 to 2017, which may be attributed to, firstly, the difference in non-desired outputs, and the fact that the efficiency measurements of the existing studies do not include the constraints of carbon emissions, which implies that disregarding the outputs of carbon emissions will overestimate land utilization efficiency, and the second is the difference in the choice of input land factors.

2) From the perspective of factor efficiency, the average value of carbon emission efficiency of urban villages and industrial and mining land has been significantly higher than the average value of carbon emission efficiency of total efficiency and transportation land since 2010, which is partly due to the fact that the construction land can play a higher role in economic efficiency while bringing about high carbon emission, and partly due to the fact that the rigid constraints of the construction land index have reduced the supply of urban villages and industrial and mining land and its inputs, which is

equivalent to shrinking the supply of land and inputs of urban villages and industrial and mining land, under the premise of lack of constraints of the input of other factors. On the other hand, the rigid constraint of the construction land target reduces the supply and input of urban villages and industrial and mining land, which is equivalent to narrowing the gap between the real input of urban villages and industrial and mining land and the target input in the absence of other factor input.

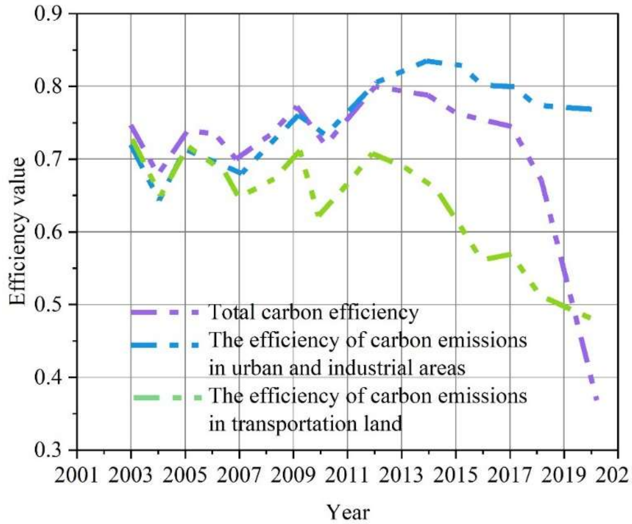

Fig. 1. Mean time sequence change

Gaussian kernel density estimation method was used to obtain the kernel density distribution of total and factor efficiency of land resource allocation in Chinese provinces, and the kernel density distribution of total efficiency, the kernel density distribution of carbon emission efficiency of urban villages and industrial and mining land use, and the kernel density distribution of carbon emission efficiency of transportation land use are shown in Fig. 2, Fig. 3 and Fig. 4, respectively. The kernel density distributions of total efficiency and factor efficiency differ greatly.

1) From the viewpoint of total efficiency, the kernel density curve shifted to the left overall from 2003 to 2020, and shifted to the left even more obviously from 2010 to 2020, and the peak of the curve continued to rise, indicating that the total efficiency of land resource allocation firstly showed a downward trend, and secondly, the gap between domains narrowed, and shifted from multi-polarization to "agglomeration". "This corresponds to the fact that during the sample period, the allocation of various types of land resources was not balanced enough, and "high factor inputs and low economic outputs" combined with increasing carbon emissions and undesired outputs led to inefficiency in the allocation of land resources and a narrowing of the inter-regional gap. This is the result of a common decline in the efficiency of most provinces and a decreasing number of model efficiency highlands.

2) In terms of the carbon emission efficiency of urban villages and industrial and mining land use, the kernel density curve from 2003 to 2020 showed a dynamic upward trend in the y-axis direction, and then declined in 2020, with a higher peak and a smaller change interval, representing a narrowing of the gap between regions, and the curve shifted to the right with a left trailing tail, which implies that the threshold for low efficiency has been increasing, and that the gap between the low and the high values is narrowing, and the above indicates a narrowing of the gap between regions. All of the above indicate that the carbon emission efficiency gap between urban villages and industrial and mining land use between provinces and regions is narrowing.

3) In terms of the carbon emission efficiency of transportation land, the peak of the curve from 2003 to 2016 is a single wave, and the peak is constantly shifted upward, indicating that the gap in the carbon emission efficiency of transportation land between provinces is decreasing, but it is still in the

state of polarization, which corresponds to the rapid development of China's transportation industry, and the allocation of transportation land is more balanced between provinces, but the first-mover advantage of some provinces still exists. The difference in the statistical caliber of the transportation land use classification in the "three surveys" data in 2020 has caused the kernel density curve of the carbon emission efficiency of transportation land use to change in the opposite direction, which is shown by a leftward shift of the curve, with the wave peak located in the low-efficiency range, which is related to the fact that transportation land use carries a large amount of carbon emissions, and under the constraint of carbon emissions, the difference between provinces and regions is still polarized, and the difference between provinces and regions in carbon emission efficiency has continued to narrow. emission constraints, the differences between provinces and regions are narrowed but the overall efficiency is lower.

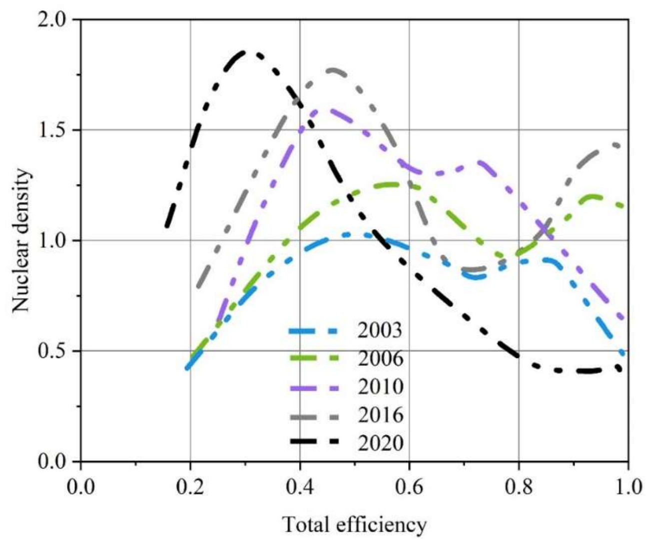

Fig. 2. Total efficiency of nuclear density distribution

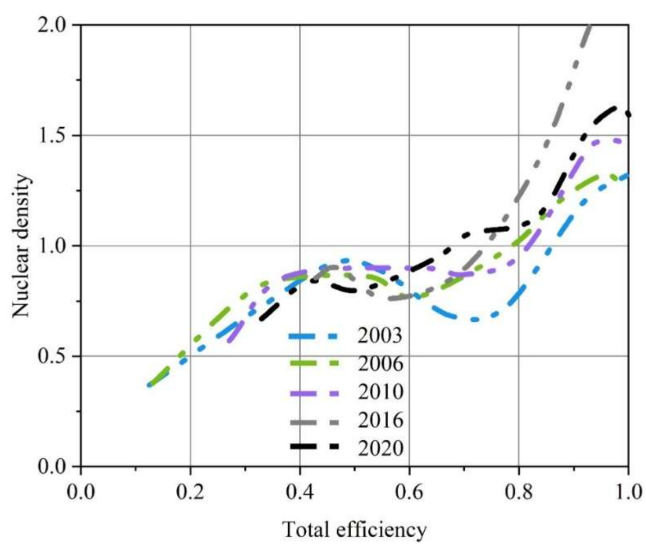

Fig. 3. Distribution of carbon emission efficiency in towns and sites

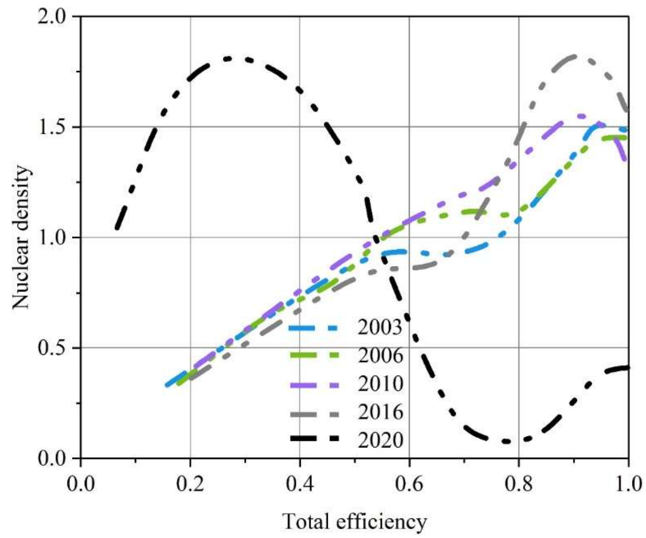

Fig. 4. Distribution of nuclear density of traffic land

6. Analysis of spatial and temporal changes in land resources due to urban expansion

6.1. Characterization of spatial differentiation

The three periods of data of the relevant parameters fitted by the semi-variance function of China's land resource values from 2003 to 2023 were calculated by applying the trend analysis function of the Geostatistical Analyst tool in the extension module of Arc GIS 10.2 software, and the spatial trend distributions of China's land resource indices for 2003-2023 were obtained, and the spatial trend distributions of the land resource indices in 2003, 2013 and 2023 are shown in Figures 5-7, respectively. The X and Y axes represent the spatial locations of the data points, and the Z axis represents the values of the data points. The lines on the X and Z sides are the east-west trend lines, and the lines on the Y and Z sides are the north-south trend lines. It is possible to analyze the spatial layout trend of China's LRI based on these two trend lines.

China's LRIs from 2003 to 2023 all show a spatial trend direction of high in all directions and low in the middle. In terms of the east-west trend line, the curves of the three periods are bending more and more, indicating that the degree of spatial variation of China's LRI in the east-west direction gradually strengthens and the value of LRI in the middle gradually decreases. From the trend of the north-south trend line, the trend line in 2003 is more gentle, and the spatial variability of the LRI in that year is not particularly obvious on the surface. However, the trend lines in the following two periods are getting lower and lower from the perimeter to the center, and the spatial variability of the LRI is getting more and more obvious. Overall, China's land resource index shows a certain degree of decline in both the east-west and north-south directions, and the value in the central region declines most significantly. This indicates that during the period of 2003-2023, due to the expansion of cities, China's land resources are weakening year by year from the central region to the surrounding area, and the overall land resources are strongly influenced by human activities.

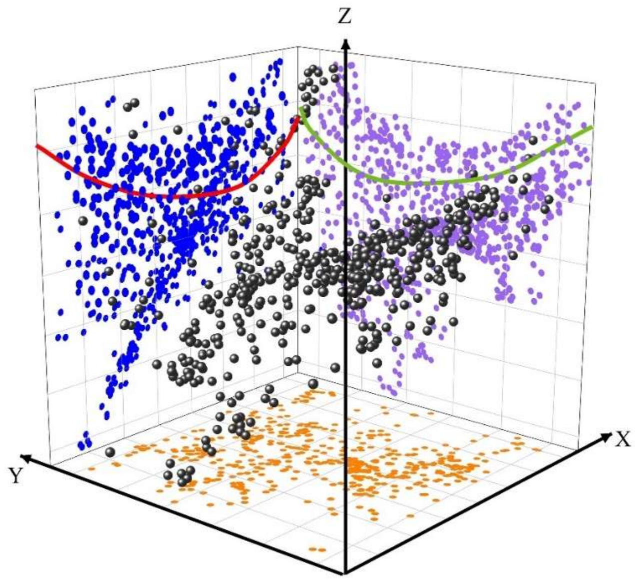

Fig. 5. Distribution of land resource space in 2003

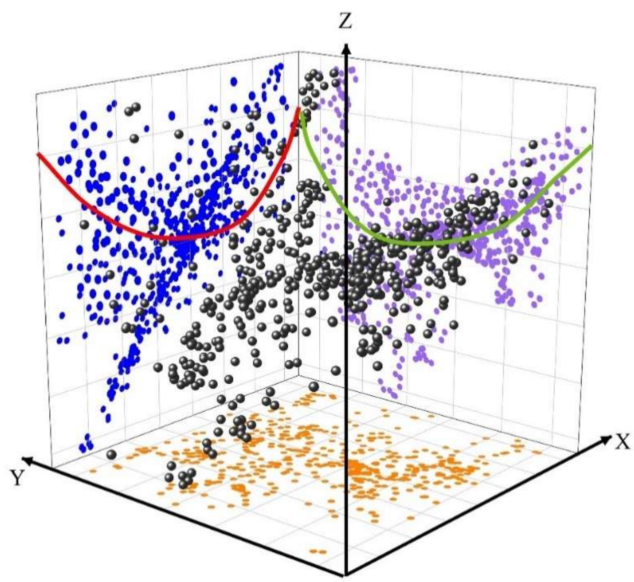

Fig. 6. Distribution of land resource space in 2013

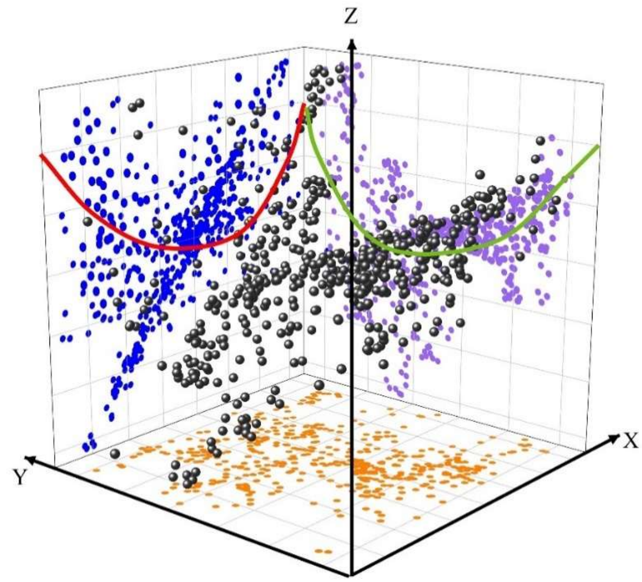

Fig. 7. Distribution of land resource space in 2023

6.2. *Global spatial autocorrelation*

With the help of GeoDa1.18 software to spatially analyze the impact of distribution characteristics of land resources in China under urban expansion in three phases, the global spatial autocorrelation results and Moran scatter plots obtained are shown in Table 1 and Figures 8-figure supplement 10, respectively.The global Moran's I for the years 2003-2023 are 0.6289, 0.7159, and 0.7368, respectively, which show an overall increasing trend. The lowest Z value is 71.25 and the highest is 87.59, and the P values are all <0.01, which tends to be close to 0, indicating that the spatial distribution of land resources in China from 2003 to 2023 all show a positive correlation, and the correlation is high, and the spatial distribution is more aggregated. In addition, the values of Moran's I index and Z-value from 2003 to 2023 show an upward trend, indicating that the spatial distribution of land resources in China has become more and more obvious with the socio-economic development and urbanization process in the past 20 years.

Table 1. Moran's I index and test value

Year	Moran's I	P	Z
2003	0.6289	0.000	71.25
2013	0.7159	0.000	83.56
2023	0.7368	0.000	87.59

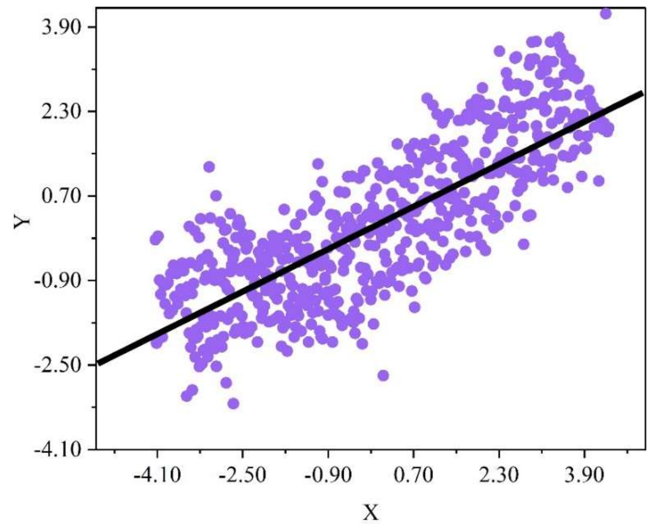

Fig. 8. 2003 Moran's I map

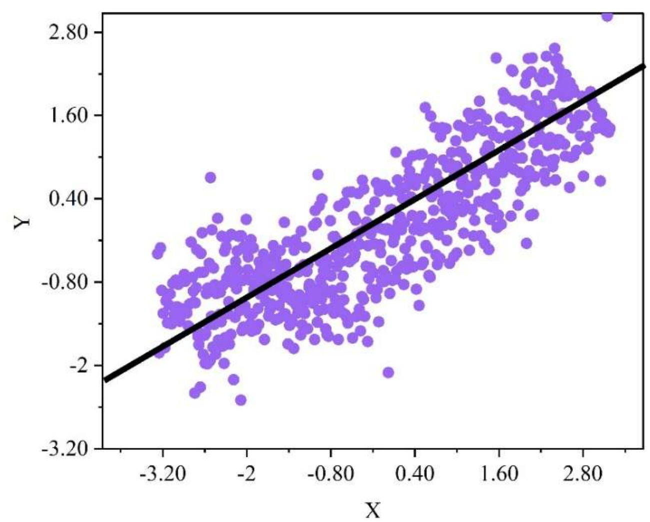

Fig. 9. 2013 Moran's I map

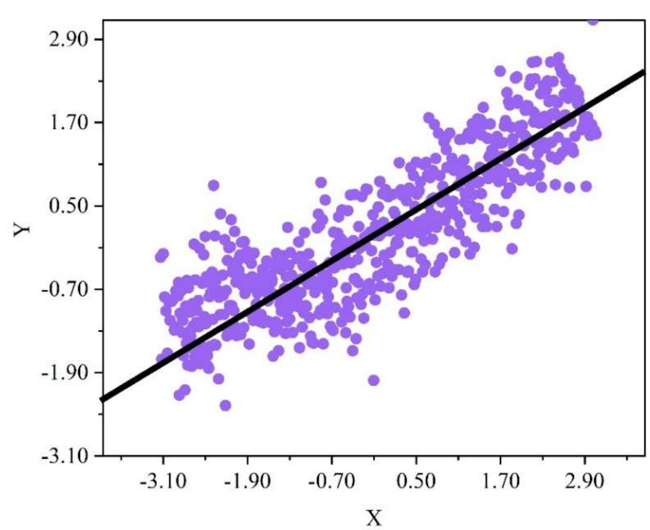

Fig. 10. 2023 Moran's I map

7. Analysis of factors influencing the spatial differentiation of land resources

In this study, the ordinary least squares (OLS) method in ArcGIS was first utilized to calculate the significance (p-value) and global impact of the explanatory variables. Building volume rate, population size, regional economic development, regional cultural level, infrastructure construction, recreational facilities construction, medical facilities construction, and urban fallow forest area were selected as explanatory variables.The results of the OLS regression are shown in Table 2.The building volume rate, population size, regional economic development, regional cultural level, infrastructure construction, and urban fallow forest area have the most significant impacts on the spatial differentiation of land resources at the global level. The influence is most significant. Meanwhile, according to the coefficient analysis, the building volume rate and the area of returning farmland to forest are inversely proportional to the spatial differentiation of land resources, and the lower their values are, the higher the spatial differentiation of land resources is.

In order to ensure that there is no strong statistical correlation or linear relationship between the explanatory variables, the variance inflation factor (VIF) was calculated in this study. When the VIF value of an explanatory variable is >5, it is considered that there is multicollinearity and should be removed.The variable of urban fallow area was chosen to be removed in the result of the VIF value. Spatio-temporal geographically weighted regression analysis was performed on the indicators that passed the test. In the process of constructing the spatio-temporal geographic weighted regression model, referring to the research of previous scholars, this study chooses the adaptive Gaussian function and optimizes it with the AICc criterion. In order to verify the superiority of the spatio-temporal geographic weighted regression model, this study compares the running results of the OLS model, the GWR model and the spatio-temporal geographic weighted regression model, and the specific results are shown in Table 3. The comparison results show that the spatio-temporal geographically weighted regression model has a higher fitting accuracy compared with the other two models, while the AICc value is lower, 64958.This indicates that the spatio-temporal geographically weighted regression model can more accurately capture the spatial heterogeneity of the spatial differentiation of the land resources, and shows a higher superiority.

Table 2. Common least squares results are summarized

Variable	Coefficient	Standard deviation	Probability	VIF
Construction volume	-278.495	105.69	0.012*	1.021
Population scale	451.46	1542.05	0.000*	0.108
Regional economic development	758.91	456.78	0.000*	2.616
Regional cultural level	496.12	145.69	0.037*	1.511
Infrastructure construction	125.69	577.81	0.000*	1.375
Recreational facilities construction	226.14	458.69	0.415	1.315
Medical facilities construction	598.78	1254.26	0.121	1.516
Urban reforestation area	-789.58	1112.26	0.000*	5.205

Table 3. The comparison of the three models

	OLS	GWR	This model
R^2	0.403	0.495	0.678
Adjusted R^2	0.395	0.474	0.583
AICc	67893	65895	64958

8. Conclusion

This paper constructs a spatio-temporal geographically weighted regression model to explore the impact of urban expansion on land resources based on the collation of quantitative methods for land resources. The main conclusions include the following:

1) From the point of view of the temporal change of land resource allocation efficiency, in the past 20 years, China's land resource allocation efficiency has shown a wave-like change in the first half of the period and a downward trend in the second half. The total efficiency trend of land resource allocation efficiency evolution is declining. The gap between different values of carbon emission efficiency of urban villages and industrial and mining land narrows, and the gap between provinces narrows.

2) From the point of view of spatial differentiation characteristics, China's land resource index shows a trend of high surrounding and low center. The trend bending amplitude of the three periods on the east-west trend gradually increases. The trends of the north-south trend line and the east-west trend are consistent. It shows that China's urban expansion gradually increased during the 20-year period, and the land resources weakened year by year from the center to the periphery, and the overall land resources were strongly influenced by urban expansion. From the results of global spatial autocorrelation analysis, China's three-period global Moran's I are 0.6289, 0.7159, and 0.7368, respectively, increasing year by year. It shows that the spatial distribution of land resources in China presents a strong spatial aggregation feature, and the feature becomes more and more significant with the increase of urban expansion.

3) Urban expansion features such as building volume rate, population size, regional economic development, regional cultural level, infrastructure construction and urban area of returning farmland to forests have the most significant influence on the spatial differentiation of land resources in the whole world. At the same time, the lower the values of building volume rate and urban forest area, the higher the spatial differentiation of land resources.

Funding

This work was supported by the 2024 Scientific and Technological Research Projects of Chongqing Municipal Education Commission, "Research on the Landscape Adaptability of Mountainous Industrial Heritage" (KJQN202402504).

References

[1] Zhao, S., Zhou, D., Zhu, C., Sun, Y., Wu, W., & Liu, S. (2015). Spatial and temporal dimensions of urban expansion in China. *Environmental science & technology*, 49(16), 9600-9609.

[2] Zhang, Z., Liu, F., Zhao, X., Wang, X., Shi, L., Xu, J., ... & Liu, B. (2018). Urban expansion in China based on remote sensing technology: a review. *Chinese Geographical Science*, 28, 727-743.

[3] Radwan, T. M., Blackburn, G. A., Whyatt, J. D., & Atkinson, P. M. (2019). Dramatic loss of agricultural land due to urban expansion threatens food security in the Nile Delta, Egypt. *Remote Sensing*, 11(3), 332.

[4] Fenta, A. A., Yasuda, H., Haregeweyn, N., Belay, A. S., Hadush, Z., Gebremedhin, M. A., & Mekonnen, G. (2017). The dynamics of urban expansion and land use/land cover changes using remote sensing and spatial metrics: the case of Mekelle City of northern Ethiopia. *International journal of remote sensing*, 38(14), 4107-4129.

[5] Dadi, D., Azadi, H., Senbeta, F., Abebe, K., Taheri, F., & Stellmacher, T. (2016). Urban sprawl and its impacts on land use change in Central Ethiopia. *Urban Forestry & Urban Greening*, 16, 132-141.

[6] Cao, H., Liu, J., Fu, C., Zhang, W., Wang, G., Yang, G., & Luo, L. (2017). Urban expansion and its impact on the land use pattern in Xishuangbanna since the reform and opening up of China. *Remote Sensing*, 9(2), 137.

[7] Mahtta, R., Fragkias, M., Güneralp, B., Mahendra, A., Reba, M., Wentz, E. A., & Seto, K. C. (2022). Urban land expansion: the role of population and economic growth for 300+ cities. *Npj Urban Sustainability*, 2(1), 5.

[8] Wang, K., Li, Z., Zhang, J., Wu, X., Jia, M., & Wu, L. (2020). Built-up land expansion and its impacts on optimizing green infrastructure networks in a resource-dependent city. *Sustainable Cities and Society*, 55, 102026.

[9] Song, S., Liu, Z., He, C., & Lu, W. (2020). Evaluating the effects of urban expansion on natural habitat quality by coupling localized shared socioeconomic pathways and the land use scenario dynamics-urban model. *Ecological Indicators*, 112, 106071.

[10] Ren, Q., He, C., Huang, Q., Shi, P., Zhang, D., & Güneralp, B. (2022). Impacts of urban expansion on natural habitats in global drylands. *Nature Sustainability*, 5(10), 869-878.

[11] Huang, C. W., McDonald, R. I., & Seto, K. C. (2018). The importance of land governance for biodiversity conservation in an era of global urban expansion. *Landscape and Urban Planning*, 173, 44-50.

[12] Wang, C., Myint, S. W., Fan, P., Stuhlmacher, M., & Yang, J. (2018). The impact of urban expansion on the regional environment in Myanmar: A case study of two capital cities. *Landscape ecology*, 33, 765-782.

[13] Nor, A. N. M., Corstanje, R., Harris, J. A., & Brewer, T. (2017). Impact of rapid urban expansion on green space structure. *Ecological Indicators*, 81, 274-284.

[14] Ebrahimi, A., & Kiani Salmi, E. (2019). Assessing the impact of urban expansion and land cover changes on land surface temperature in Shahrekord city. *Journal of RS and GIS for Natural Resources*, 9(4), 102-118.

[15] Qu, S., Hu, S., Li, W., Wang, H., Zhang, C., & Li, Q. (2020). Interaction between urban land expansion and land use policy: An analysis using the DPSIR framework. *Land Use Policy*, 99, 104856.

[16] Puplampu, D. A., & Boafo, Y. A. (2021). Exploring the impacts of urban expansion on green spaces availability and delivery of ecosystem services in the Accra metropolis. *Environmental Challenges*, 5, 100283.

[17] Cecchini, M., Zambon, I., Pontrandolfi, A., Turco, R., Colantoni, A., Mavrakis, A., & Salvati, L. (2019). Urban sprawl and the 'olive'landscape: Sustainable land management for 'crisis' cities. *GeoJournal*, 84, 237-255.

[18] van Vliet, J. (2019). Direct and indirect loss of natural area from urban expansion. *Nature Sustainability*, 2(8), 755-763.

[19] Wu, Y., Li, S., & Yu, S. (2016). Monitoring urban expansion and its effects on land use and land cover changes in Guangzhou city, China. *Environmental monitoring and assessment*, 188, 1-15.

[20] Colsaet, A., Laurans, Y., & Levrel, H. (2018). What drives land take and urban land expansion? A systematic review. *Land Use Policy*, 79, 339-349.

[21] Wei, Y. D., & Ewing, R. (2018). Urban expansion, sprawl and inequality. *Landscape and urban planning*, 177, 259-265.

[22] Bren d'Amour, C., Reitsma, F., Baiocchi, G., Barthel, S., Güneralp, B., Erb, K. H., ... & Seto, K. C. (2017). Future urban land expansion and implications for global croplands. *Proceedings of the National Academy of Sciences*, 114(34), 8939-8944.

[23] Şen, G., Güngör, E., & Şevik, H. (2018). Defining the effects of urban expansion on land use/cover change: a case study in Kastamonu, Turkey. *Environmental monitoring and assessment*, 190, 1-13.

[24] Song, X., Liu, Y., Zhu, X., Cao, G., Chen, Y., Zhang, Z., & Wu, D. (2022). The impacts of urban land expansion on ecosystem services in Wuhan, China. *Environmental Science and Pollution Research*, 1-14.

[25] Huang, Q., Liu, Z., He, C., Gou, S., Bai, Y., Wang, Y., & Shen, M. (2020). The occupation of cropland by global urban expansion from 1992 to 2016 and its implications. *Environmental Research Letters*, 15(8), 084037.

[26] Wangwei Chu,Yingguang Li,Changqing Liu,Wenping Mou & Limin Tang. (2014). A manufacturing resource allocation method with knowledge-based fuzzy comprehensive evaluation for aircraft structural parts. International Journal of Production Research(11),3239-3258.

[27] Haoyan Duan,Qian Cao,Lunche Wang,Xihui Gu & Khosro Ashrafi. (2024). Exploring the relationships between 3D urban landscape patterns and PM2.5 pollution using the multiscale geographic weighted regression model. Progress in Physical Geography: Earth and Environment(3),368-388.

[28] Kermarrec Gaël,Lösler Michael,Guerrier Stéphane & Schön Steffen. (2022). The variance inflation factor to account for correlations in likelihood ratio tests: deformation analysis with terrestrial laser scanners. Journal of Geodesy(11).