ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

Spatial and Temporal Evolution of Land Use Change and Precipitation Based on SARIMA and Random Forest Models

Xintong Dong[1,*], Wanxin Liu[2], Yu Chen[3]

[1]*School of International Economics and Politics, Liaoning University, Shenyang, China*
[2]*School of Mathematics and Statistics, Liaoning University, Shenyang, China*
[3]*Faculty of Information, Liaoning University, Shenyang, China*
[*]*Corresponding author: 18741982001@163.com*

Abstract: *In the context of rapid urbanisation, land use change has become a key factor affecting ecological environment and socio-economic development [1]. In this paper, the spatial and temporal evolution characteristics of precipitation and land use/land cover types and their driving factors are analysed by constructing various prediction models. For the spatio-temporal changes of precipitation, this paper adopts statistical methods to calculate the mean and standard deviation, and predicts the future monthly precipitation by the SARIMA model, which is combined with the least-squares method and one-way linear regression analyses to reveal its inter-annual change trends [2-4]. Meanwhile, the transfer probability of different land types was analysed by Markov chain model, and the dynamic characteristics of land use change were demonstrated by detecting the mutation points through CPT method [5]. In exploring the influence of topographic and climatic factors on extreme weather, this paper constructed an orographic precipitation model (OPM) and a multiple regression model [6], and used the random forest algorithm to capture the complex nonlinear relationships, and finally verified the validity of the model through assessment methods such as mean square error and spatial autocorrelation [7]. The results of the study provide an important reference for understanding land use change and its driving mechanism, and provide a scientific basis for related policy formulation.*

Keywords: *Topographic precipitation modelling (OPM), random forest regression, Markov chain model, SARIMA model*

1. Introduction

With the acceleration of urbanisation globally, land-use changes have had far-reaching impacts on ecological and socio-economic systems. The transformation of land-use patterns not only changes the distribution and use of natural resources, but also affects regional climate, ecological balance and socio-economic structure. Therefore, an in-depth analysis of the spatial and temporal characteristics of land use change and its driving factors is of great significance for achieving sustainable development and rational planning of land resources. Meanwhile, precipitation, as an important variable in the climate system, has a direct impact on land use change, especially in the context of intensifying climate change and the frequent occurrence of extreme weather events, which poses a great challenge to land use and socioeconomics. This paper systematically analyses the characteristics of land use change, the spatial and temporal evolutionary trend of precipitation and its interaction mechanism based on multiple prediction models, combining climate and topographic factors, with the aim of providing a scientific basis for policy makers and theoretical support for future land resource management and ecological protection.

2. Characterisation of the spatial and temporal evolution of precipitation and land use/cover type

2.1 Data preprocessing

Precipitation and land use/land cover type are two variables with different natures, so different processing methods are needed; precipitation is a continuously changing variable suitable for the use of time-series analysis methods, while land use/land cover type is a discrete variable suitable for categorical and mutation analyses.

ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

For the dataset collected in this paper, precipitation data at latitude and longitude outside China are set to -99.9 in the dataset, and in the subsequent data processing, statistical indicators and statistical charts can be established, and -99.9 can be set to 0, to prevent these data from influencing the precipitation indicators in some border areas.

2.2 Constructing descriptive statistical methods

2.2.1 Continuous variable description of spatio-temporal evolution

For the spatial and temporal variations of precipitation, the characteristics can be summarised by temporal statistics and spatial distribution analysis. Through the day-by-day precipitation data, the annual precipitation data are counted, using the annual average precipitation and its standard deviation as the first statistical index. The calculation formula is:

$$\bar{P} = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{1}$$

where P_i denotes the amount of precipitation in the ith region and N denotes the number of all regions in China.

2.2.2 SARIMA model - time series analysis

From a practical point of view, the monthly precipitation data from 1990 to 2020 are inevitably cyclical, so they are fitted with seasonal autoregressive integral moving average model (SARIMA).

Seasonal autoregressive integral moving average model is an effective method to deal with time series data with seasonal patterns, usually denoted as $\mathrm{SARIMA}(\mathrm{p,d,q})(\mathrm{P,D,Q})_\mathrm{S}$. By combining the autoregressive, differential, moving average and seasonal components, the SARIMA model can capture the complex dynamic characteristics of the time series, and its mathematical expression is:

$$\phi_p(B)\Phi_P(B_s)(1-B_s)^d(1-B_s)^D x_t = \theta_q(B)\Theta_Q(B_s)\varepsilon_t \tag{2}$$

where B is a delay operator, B_s is a seasonal delay operator, $\phi(B)$) and $\theta(B)$ are delay polynomial operators, $\Phi(B_s)$ and $\Theta(B_s)$ are seasonal delay polynomial operators, and there:

$$\begin{aligned}
&\phi(B) = 1-\phi_1 B - \ldots - \phi_p B^p \\
&\theta(B) = 1-\theta_1 B - \ldots - \theta_q B^q \\
&\Phi\left(B_s\right) = 1-\Phi_1 B_s - \ldots - \Phi_p {B_s}^P \\
&\Theta\left(B_s\right) = 1-\Theta_1 B_s - \ldots - \Theta_q {B_s}^Q
\end{aligned} \tag{3}$$

The SARIMA model is suitable for time series data with seasonal patterns and performs well in short- and medium-term forecasting. By choosing the appropriate model order and seasonal period, the forecasting accuracy can be significantly improved.

2.2.3 Spatial distribution characteristics

Spatially averaged precipitation: The spatial average of precipitation for each region of the country is calculated annually, which can be done using the geographically weighted average (GWR) technique. in order to map the spatial distribution of precipitation, showing the changes in the distribution of precipitation in different regions of China.

Spatial heterogeneity of precipitation: it is possible to calculate precipitation extremes (maxima and minima) for each region and to analyse the variation of precipitation with spatial distance using geostatistical functions of variation.

Spatial heterogeneity equation for precipitation:

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[P(x_i) - P(x_i+h)]^2 \tag{4}$$

where h is the spatial distance, N(h) is the number of point pairs at distance h, and $P\left(x_i\right)$ is the precipitation at the ith point.

ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

2.2.4 Analysis of the area of land use by type

Land use/land cover is discrete data, and the spatial and temporal distribution characteristics of each type of land use are analysed and summarised by visualising the annual average area of each type of land use.

2.2.5 Mutational transfer analysis

Transfer analysis can be used to analyse the transfer probability of different land use types through Markov chains. This method is suitable for analysing the conversion process of land types from one type to another. Mutation analysis uses the CPT (Change Point Test) method to detect points of mutation in land cover and identify years or areas of sudden change.

There are many real-world phenomena in which a system's future moments are only relevant to the present and not to its past history, given that the present conditions of the system are known. For example, when studying the cumulative sales of a shop, if the cumulative sales at the present moment are known, the cumulative sales at a future moment are not related to the cumulative sales at any moment prior to the present moment. Construction of the transfer probability matrix:

$$P_{ij} = \frac{A_{ij}}{A_i} \tag{5}$$

where P_{ij} is the probability of transformation from type i to type j, A_{ij} is the area of transformation from type i to type j, and A_i is the total area of type i.

2.2.6 Model solving

Firstly, the precipitation data were extracted and descriptive statistics of the data were carried out to obtain a description of the precipitation related data as shown in Table 1:

Table 1: Descriptive statistics of annual precipitation

Average value	Median	Standard deviation
260.0023005222334	259.4772677951389	16.139796014943133

Then SARIMA prediction was performed on the month-by-month precipitation data, and the month-by-month precipitation time series is shown in Figure 1:

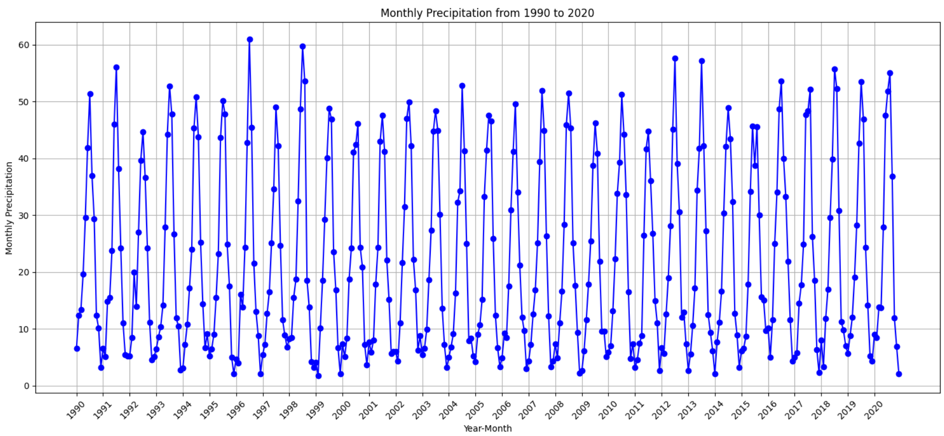


Figure 1: Monthly precipitation time series, 1990-2020

The time series plot shows that there is a great difference in monthly precipitation in the same year, and it can be identified as having a seasonal effect with a period of 12. Therefore, the first-order difference and the first-order seasonal difference are performed to eliminate the instability and periodicity as shown in Figure 2 and Figure 3.

ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

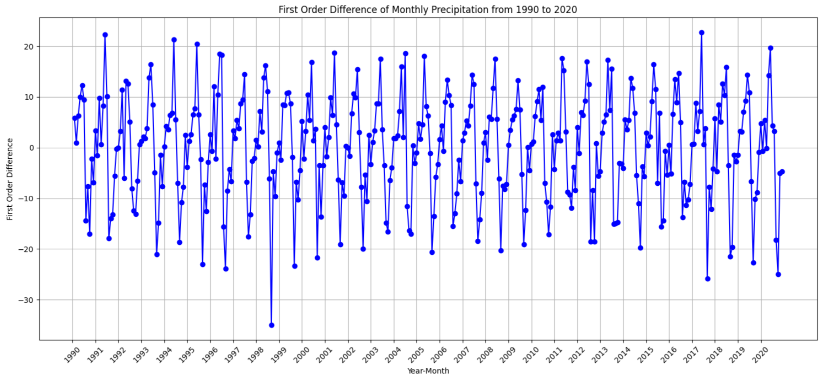


Figure 2: First order differential timing diagram

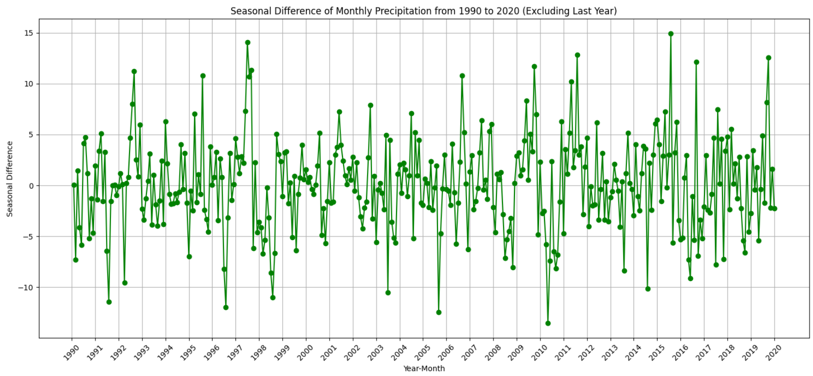


Figure 3: First-order seasonal difference timing diagram

The model parameters are obtained by building a SARIMA model for the differenced time series, which is well tested and has a good model fit.

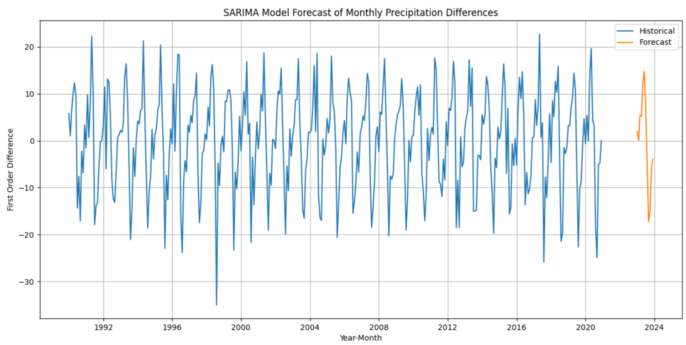


Figure 4: Forecast map of future precipitation trends

Figure 4 shows that precipitation will maintain a similar historical fluctuating trend for some time to come.

3. Study on the mechanism of influence of topographic and climatic factors on weather extremes

3.1 Topographic and climatic interactions

When moist air encounters a mountain range, it is forced to rise, temperatures drop, and water vapour condenses to form precipitation, especially orographic precipitation (Orographic Precipitation), which forms on windward slopes. Basin topography tends to accumulate cold air, creating a low-temperature inversion layer that increases humidity and further affects the precipitation process. Slope gradient affects the speed of airflow, and slope orientation determines the amount of solar radiation received, thus affecting the local distribution of air temperature and water vapour.

ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

3.2 Impact of climatic factors on extreme weather

Temperature determines the water content of the air; the higher the temperature, the more water vapour the air can carry. High humidity means that there is enough water vapour in the air to easily form precipitation. Strong winds accelerate air currents, leading to water vapour transport and speeding up the condensation process.

3.3 Modelling the effect of topography on precipitation

Here the use of Orographic Precipitation Model (OPM) can be used to explain the mechanism by which topography affects precipitation. The common formula is:

$$P = P_0 + kH \tag{6}$$

where: P denotes total precipitation, P_0 denotes precipitation when there is no topographic obstruction, k is a topographic factor indicating the incremental effect of topography on precipitation, and H is the height (elevation) of the terrain.

We can further refine the model by introducing slope gradient and slope direction to describe the effect of different terrain on rainstorms:

$$P = P_0 + k_1 H + k_2 \mathrm{sinh}_1 + k_3 \mathrm{sinh}_2 \tag{7}$$

This allows for more accurate modelling of the distribution of heavy rainfall over mountainous terrain.

3.4 Climate modelling

The relationship between climate variables and precipitation can be described by multivariate regression or more sophisticated machine learning models. For example, we can use Random Forest Regression or Support Vector Regression to establish a non-linear relationship between climate variables (temperature, humidity, wind speed) and extreme precipitation.

The formula for the random forest regression model is expressed as:

$$P = f(T, H_u, W) + \varepsilon \tag{8}$$

Random forest regression is suitable for dealing with complex climate datasets by constructing multiple decision trees to fit the input variables nonlinearly.

3.5 Topography-climate interaction term

To capture the effect of terrain-climate interactions on weather extremes, we can introduce interaction terms.

$$P = \alpha_1 H + \alpha_2 HT + \alpha_3 HH_u + \alpha_4 SW \tag{9}$$

The formula describes the interaction between terrain elevation and climatic variables, in particular how temperature, humidity, and wind speed can amplify or inhibit the formation of heavy rainfall in different terrains.

3.6 Modelling Algorithms

3.6.1 Data fitting and regression analysis

To solve for the parameters in this model, we can use the following methods:

Least Squares: for linear models, least squares can be used to solve for the coefficients.

Machine learning algorithms: for non-linear relationships, algorithms such as Random Forest Regression and Support Vector Machines can be used.

Specific steps:

Prepare the training dataset to correlate the terrain and climate variables with the actual precipitation.

Select appropriate algorithms for model training and validation.

Predict future extreme precipitation events and test the accuracy of the model.

ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

3.6.2 Spatial analysis and geographically weighted regression

Geographically weighted regression (GWR) is used here to describe the spatially heterogeneous effects of topography on climate, and GWR can take into account local variations in topographic and climatic variables, capturing differences in heavy rainfall across regions.

The formula for GWR is:

$$P(x_i, y_i) = \beta_0(x_i, y_i) + \beta_1(x_i, y_i)H + \beta_2(x_i, y_i)T + \beta_3(x_i, y_i)H_u + \varepsilon \tag{10}$$

where (x_i, y_i) is the geographic location coordinates and $\beta_i(x_i, y_i)$ is the regression coefficient with geographic location.

The temperature data were analysed and processed to obtain images for selected years as shown in Figure 5 below:

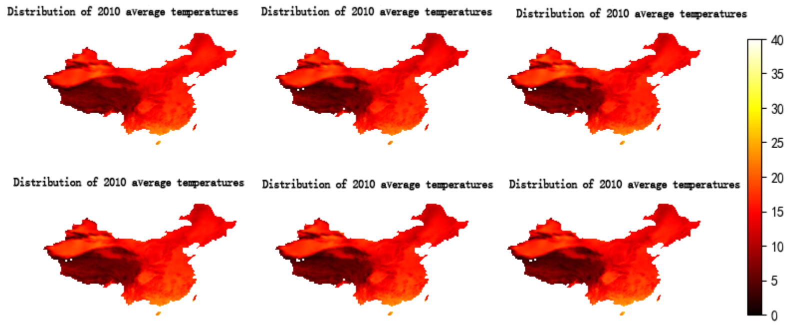
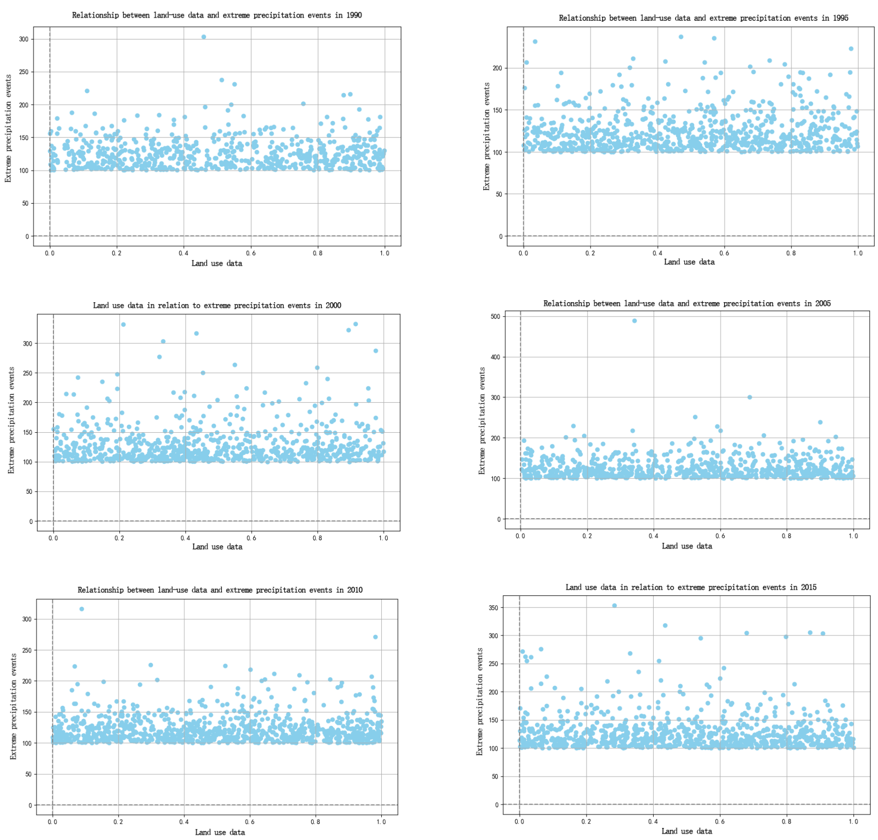


Figure 5: Average temperature map

Calculations were made based on the model developed above, and Figure 6 shows the correlation between land use data and extreme precipitation events for only some of the years.



ISSN 2616-5872 Vol.6, Issue 6: 26-32, DOI: 10.25236/AJEE.2024.060604

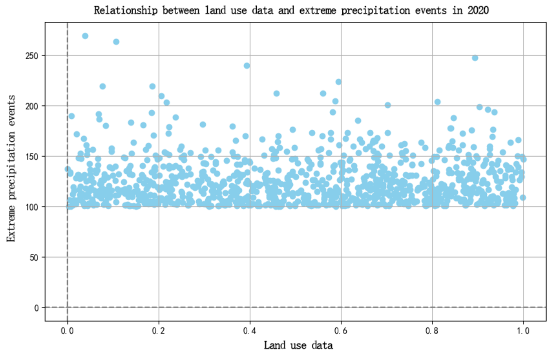


Figure 6: Correlation map between land use data and extreme precipitation events

In this study, we aim to explore the relationship between precipitation and land use types, particularly the impact of extreme precipitation events (precipitation exceeding 100 mm). By analysing the precipitation data from 1961 to 2022 and combining it with the changes in land use types, some important conclusions and observations were made.

4. Conclusions

In this paper, the spatial and temporal evolution characteristics of precipitation and land use/cover types and their interrelationships are deeply explored by establishing multiple prediction models. The results show that the interannual changes of precipitation show obvious fluctuations and trends, and the combination of adaptive stepwise sampling and SARIMA model can better predict the future monthly precipitation. The transfer probability analysis of land use types shows that the changes between different land types have significant dynamic characteristics, and the Markov chain model effectively reveals this dynamic process. Meanwhile, the construction of terrain precipitation model and multiple regression model further clarifies the complex mechanism of the role of terrain and climate factors in the formation of extreme weather. The study not only provides data support and theoretical basis for future land use management and ecological protection, but also provides scientific decision-making reference for coping with extreme weather risks in the context of climate change.

References

[1] JI Xiongjuan, CHENG Wenshi. Simulation prediction of land use change based on CLUE-S and Markov model[J]. Land and Natural Resources Research, 2024, (06):17-21. DOI: 10. 16202/ j. cnki. tnrs.2024.06.004.

[2] J.T. Huang, Z.X. Wang, Y.Q. Teng, et al. Prediction of unit output volume based on DE-SARIMA method [J]. Modern Information Technology, 2024, 8(12):116-120.DOI:10.19850/j.cnki.2096-4706. 2024. 12. 025.

[3] Bras A L , Gomes D , Filipe P A ,et al.Trends, seasonality and forecasts of pulmonary tuberculosis in Portugal[J].The International Journal of Tuberculosis and Lung Disease, 2014, 18(10):1202-1210.DOI:10.5588/ijtld.14.0158.

[4] LI Zhiqi, JI Yingjun, CHEN Xiangmin, et al. Research on ultra-wideband roadheader positioning algorithm based on constrained overall least squares[J]. Electronic Design Engineering, 2024, 32(20):144-148.DOI:10.14022/j.issn1674-6236.2024.20.030.

[5] M. Zhang, H.L. Liu, N.Z. Du, et al. Forecasting merchandise sales based on GM(1, 1)-Markov model[J]. Inner Mongolia Science and Economy, 2024, (05):77-82.

[6] Zhang Shengtang, Kang Shaozhong. Study on distributed precipitation model based on Lagrangian interpolation method to correct the influence of terrain[J]. Hydrology, 2004, (06):6-9.

[7] Jia Haoxin, Zou Menghua, Tang Xin, et al. Evaluation of construction suitability of geothermal power station based on random forest method[J/OL]. China Geology, 1-28[2024-10-02].http://kns.cnki. net/ kcms/ detail/11.1167.P.20240924.1714.003.html.