

Study on driving forces of wetland change in the Western Liaohe River basin based on random forest model

Menghong Wu^{1,*}, Changbao Yang¹, Yanhong Zhang¹ and Nan Lin²

¹School of College of Geo-exploratin Science and Technology, Jilin University, Changchun, China

²School of College of Surveying and Prospecting Engineering, Jilin Jianzhu University, Changchun, China

*Corresponding author e-mail: wxn920911@126.com

Abstract. Based on the platform of RS and GIS, random forest progression model is used for study driving force of wetland change in western Liaohe river basin, five influencing factors which include elevation, slope, temperature, precipitation and population density are chosen to establish random forest progression model about the wetland change and the driving factors. Using the the mean value of the prediction accuracy outside the bag calculated by the model to evaluate the importance of the variables. The result indicates that the coefficient of partial correlation between precipitation and wetland density is the largest among the five influencing factors, followed by temperature, population density, elevation and slope is smallest. The influence of natural factors on the change of wetland density is mainly reflected in precipitation and temperature factors, and the precipitation is obviously higher than that of temperature, under the influence of human factors, the influence of population density factor on wetland density is higher than that of elevation and slope factor. The result shows that in the past 40 years, the human activities in the study area have increased the density of wetland to some extent, but it is not the main factor.

1. Introduction

The study area is in the Western Liaohe River basin, which is located in agriculture and animal husbandry in Western Northeast Plain staggered along the eastern edge, the climate belongs to the transition band from warm temperate and semi humid to moderate temperate and semi-arid climate. The area is one of the largest wetland area in Northeast China, is also an important commodity grain producing areas in our country^[1]. Therefore, the analysis of the wetland ecological condition health has a great significance in explore main influence factors of wetland change. The natural factors and human factors on the study area is chosen to establish random forest regression model about the wetland change and the driving factors, the natural factors contains topography and meteorology, the human factors contains density of population. Mean value of out of pocket prediction accuracy is used to measure the degree of influence of each driving factor on wetland change, the grid overlay method is used to integrate the wetland data into a specific spatial unit to calculate, stack and evaluate the values, the quantitative analysis of driving forces of wetland change can provide reference basis for protect wetland of western Liaohe river basin effectively [2].



2. Study Area

The Western Liaohe River is seated in the southern part of China's northeast, flowing through 22 county-level administrative units in Inner Mongolia, Liaoning, Jilin and Hebei. The study area is located at 40°8'18.21"-45°1'20.36"N, 115°9'11.79"-123°00'13.93"E, the total area is $16.73 * 10^4 \text{km}^2$. The administration includes Tuquan, Horqin Right Wing Middle Banner, Jarud Banner, Horqin Left Middle Banner, kailu, tongliao, Horqin Left Back Banner, Ar Horqin Banner, Baarin Left Banner, Baarin Right Banner, Naiman Banner, Hure Banner, linxi, Hexigten Banner, Ongniud Banner, Aohan Banner, chifeng, jianping and Harqin Banner. The transportation in this area is convenient, railway and highway net is spread by radial pattern, Tongliao and Siping is the center of net.

3. The extraction of wetland change information and influencing factors

(1)Wetland change information

Choose MSS remote sensing image data in 1975, ETM remote sensing image data in 2000 and OLI remote sensing image data in 2015 as the remote sensing data source, artificial visual interpretation is used to obtain the distribution information of wetland in the past 40 years based on GIS software platform, using superposition analysis and transfer matrix to analysis spatiotemporal dynamic change rule of wetland change transformation in past 40 years in Western Liaohe River basin from temporal dynamic change, spatial dynamic change and wetland type transformation respectively^[3]. The dynamic change of wetland can be divided into three types: the increase of wetland, the stability of wetland and the decrease of wetland, the spatial distribution of wetland spatial dynamic change is as shown in Figure 1:



Fig.1 The wetland dynamic change figure in western Liaohe river basin

(2)Topographic factor

Topographic information is mainly extracted by DEM data, extracting terrain factors by DEM in an automatic way to spatial analysis has characteristics like fast, scientific, objective, which means that can replace the complex field observation in traditional operation. Elevation and slope factors which influence wetland distribution larger are selected in this study, as shown in figure 2, 3:

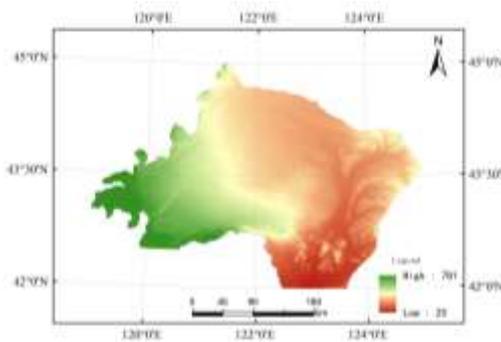


Fig.2 Elevation interpolation figure
In western Liaohe river basin

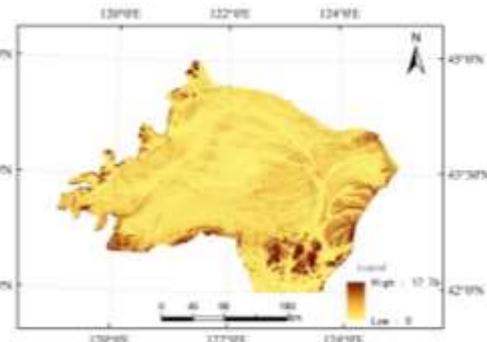


Fig.3 Slope interpolation figure
in western Liaohe river basin

(3) Meteorological factor

Temperature and precipitation which have a strong relationship with wetland growth are selected as study object, using kriging interpolation method to process interpolation of temperature and precipitation in study area^[4-6], the temperature interpolation figure and precipitation interpolation figure are shown in figure 4,5:

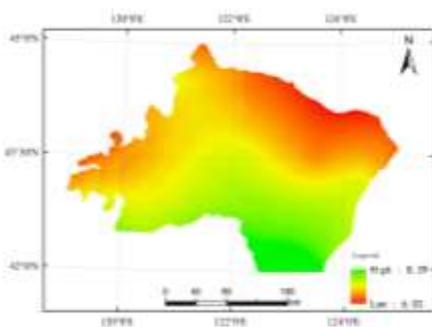


Fig.4 Temperature interpolation figure
In western Liaohe river basin

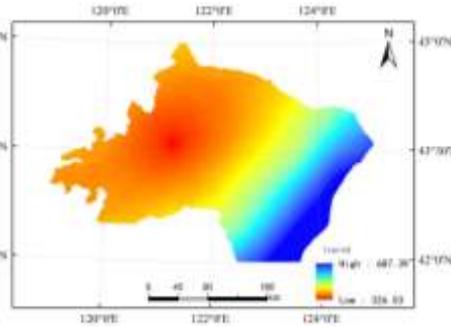


Fig.5 Precipitation interpolation figure
in western Liaohe river basin

(4) Population density factor

In a short term, human activity is the main driving factor. According to the actual situation of the study area, the paper selects a representative population density factor to explore the impact of human factors on wetland change. As shown in figure 6:

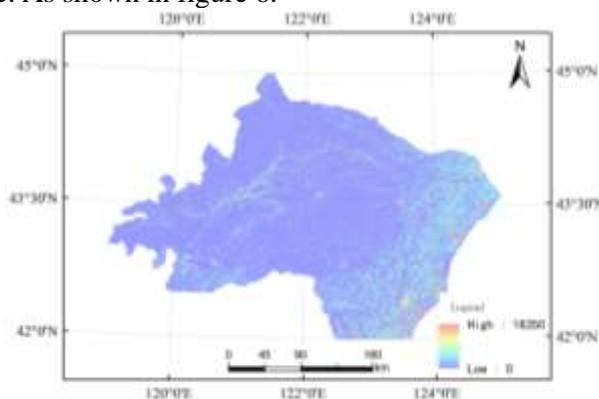


Fig.6 Population density interpolation figure in western Liaohe river basin

4. Random forest progression model

Random forest is ensemble classifier contains multiple non pruning classification and regression tree (CART), Breiman proposes the concept in 2001, Breiman combines the classification tree into a random forest, and each decision tree has two random processes, the generation process is as follows: assuming that the dependent variable Y has N samples, there are P independent variables associated with the dependent variable Y, In the process of constructing the regression tree, random forest will use Bootstrap sampling method for extracting k sample training set $(T_1, T_2 \dots T_k)$ randomly in the dependent variable Y, the size of each training set is about 2/3 of the original data set, the data which is not extracted is used for test sample set as out-bag data; Establishing classification and regression tree to generate K th trees $\{T(x_n, \theta_1), (x_n, \theta_2) \dots (x_n, \theta_k)\}$ for each data set, since each regression tree is likely to be random in nature, random forests are typically able to generate hundreds or even thousands of classification trees randomly [7,8]. The second stochastic process is to extract m features from the P features as the feature set of the current node, each classification regression tree grow recursively from top to bottom, setting the NTree value as the termination condition of the growth of the regression tree, all decision trees are combined into a random forest.

The importance evaluation of variables is an outstanding feature of random forest algorithm, among them, the evaluation of the importance of variables based on the prediction accuracy of out-bag data is widely used, the idea is to evaluate the importance of the variables by changing the value of the independent variables outside the bag randomly and to calculate the change of the accuracy of the model, the mean of the accuracy is proportional to the importance of the variable, and the larger the error caused by the change of the characteristic variable, the more important the variable is. The specific process is as follows:

(1)The prediction accuracy MSE_{OOB} of the corresponding bag data (OOB) is calculated for each training sample set, assuming that the number of samples outside the bag is q, the formula is:

$$MSE = \frac{1}{q} \sum_{i=1}^q (y_i - \hat{y}_i)^2 \quad (1)$$

In the formula, y_i represents the true value of the dependent variable in the outside bag data, \hat{y}_i represents the predict value of the regression model;

(2)The value of the j characteristic variable X_j is changed randomly, and the accuracy of the new outside bag error MSE_j is recalculated;

(3)The calculate formula of the importance of variable X_j is:

$$V(X^j) = \frac{1}{q} \sum_{j=1}^q (MSE - MSE_j) \quad (2)$$

5. Results

In order to quantitatively analyze the driving forces of wetland change, the grid superposition method is used to integrate the wetland data into a specific spatial unit to calculate, stack and evaluate. The size of the mesh affects the boundary range and the accuracy of the spatial unit directly. The study area is divided by orthogonal kilometer network, according to the actual situation of the study area; the grid size is 10km * 10km. After the grid is divided, the dynamic change trend of the wetland can be refined to each mesh, which is represented by discrete points. The following formula is to calculate the wetland change intensity:

$$P = \frac{[\sum U_b - \sum U_a]}{M_0} \quad (3)$$

Using the TIN model in DEM analysis in MAPGIS for contouring tracing analysis, P in the formula is the intensity of wetland change per unit area during the study period. $\sum U_a$ Represents the total area of wetland in each grid at the beginning of the study, $\sum U_b$ represents the final phase of the study, M_0 represents unit area. 1000 samples are selected randomly in the study area, While wetland change density is the dependent variable, elevation, slope, temperature, precipitation, population density are independent variables, the random forest progression model of wetland dynamic change is established to calculate the mean value of prediction accuracy out of pocket, the results are shown in Figure 7:

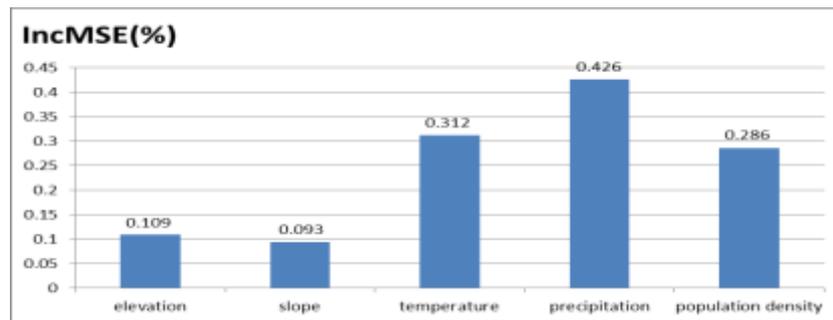


Fig.7 Analysis figure of driving factors

6. Conclusions

(1)According to the statistical results of the partial correlation coefficient of each factor and the wetland density, it can be seen that the temperature and precipitation factor under the meteorological factors have the greatest impact on the wetland change. Followed by the population density factor, the smallest impacts are the terrain factor elevation and slope factor. The high correlation coefficient of temperature and precipitation shows that the meteorological factors play an important role in controlling wetland change. Because the area is far away from the sea, the climate of the study area is more complex, which belongs to the transition zone of warm temperate semi humid climate to moderate temperate semi-arid climate. The continental climate is significant, which makes the temperature and precipitation play an important role in the development and utilization of wetland, and thus becomes the limiting condition of wetland change.

(2)Population density, elevation and slope also have a certain impact on wetland change, the impact of population density on vegetation is mainly reflected in the impact of human activities on Wetlands. The increase of population, the development of regional economy and the intensification of urbanization are the concrete manifestation of the influence of human activities on Wetlands. However, the partial correlation coefficient between elevation, slope and the density of wetland is low, because the terrain factors (elevation, slope) will not make a big change in a short period of time. Therefore, the change of wetland in the study area is relatively stable for a long time. It also shows that the elevation and slope of the whole study area are appropriate, and the adjustment ability of elevation and slope is strong.

Acknowledgements

This work was financially supported by Graduate Innovation Fund of Jilin University (No.2016199).

References

- [1] LI Yanqing. Spatial-temporal change of vegetation in the western Liaohe River Basin, Northeast China: master's degree thesis[D]. Changchun: Northeast Normal University, 2008.
- [2] Detenbeck N E, Taylor D L, Lima A, et al. Temporal and spatial variability in water quality of wetlands in the Minneapolis/ST. Paul, MN metropolitan area: implications for monitoring

- strategies and designs [J]. *Environmental Monitoring and Assessment*, 1996, 40 (1): 11-40.
- [3] Moorhead K K, Brinson M M. Response of wetlands to rising sea level in the lower coastal plain of North Carolina [J]. *Ecological Applications*, 1995, 5 (1): 261-271.
- [4] Acreman M, Blake J, Booker D, et al. A simple framework for evaluating regional wetland ecohydrological response to climate change with case studies from Great Britain [J]. *Ecohydrology*, 2009, 2(1): 1-17.
- [5] Burkett V, Kusler J. Climate Change: Potential impacts and interactions in wetlands of the United States, Virginia [J]. *Journal of the American Water Resources Association*, 2000, 36(2): 313-320.
- [6] LIN Nan. Study on the change and driving force of vegetation cover in eastern Jilin Province based on RS and GIS: master's degree thesis [D]. Changchun: Jilin University, 2010.
- [7] Wang Dong, Yue Cairong, Tian Chuanzhao, et al. Classification of TM remote sensing image based on random forests of Dayao country[J]. *Forest Inventory and Planning*, 2014,39(2):1-5.
- [8] Lei Zhen. Random forest and its application in remote sensing. [D]. Shanghai: Shanghai Jiao Tong University, 2012.