Journal of Computing and Electronic Information Management

ISSN: 2413-1660

Analysis of influencing factors of forest ecological function index based on random forest

Jinbao Liu [1,2,3,4]

[1]Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China

[2] Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710075, China

[3] Shaanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China

[4] Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China

Corresponding author e-mail: liujinbao1212@163.com

Abstract: Chlorophyll is an important factor to measure the normal growth and development of plants. This study took the chlorophyll content of rape leaves as the research object, and analyzed the influence of spectral data method on feature extraction and chlorophyll content prediction model. Use machine learning to analyze the spectral reflectance curve characteristics of different chlorophyll content, and combine the first-order differential and second-order differential to establish the optimal chlorophyll spectrum estimation model. The results show that the combination of differential transformation helps to improve the correlation between leaf chlorophyll content and spectrum. (2) In the PCR model, the model established by the second-order differential of the reciprocal logarithm after NOR processing the reflectance spectrum has the best effect, Rc2=0.7957, RMSEC=10.5436, SEC=10.6070.

Keywords: Influencing Factors, Forest Ecological Function Index, Random Forest.

1. Introduction

As the largest ecosystem on land, the change of ecological function of forest directly affects the change of human living environment [1-2]. The study on the catastrophic flood disaster in 1998 shows that over cultivation of forest and deforestation are the main causes of the flood disaster. The study on the three North Shelterbelt shows that with the increase of the area of the three North Shelterbelt, the wind and sand fixation capacity of the ecosystem is significantly increased, and the amount of soil erosion is significantly reduced, which greatly improves the local ecological environment and living conditions. With the emergence of more and more research results related to forest ecological function, people gradually realize that the change of forest ecological function is closely related to human daily life, and the awareness of forest ecological protection is becoming stronger and stronger.

The German federal forest law distinguishes the economic uses of forests and their environmental effects, especially in the aspects of ecosystem sustainability, natural performance and climate [3-4]. This includes water balance, air purification, soil fertility, landscape, agriculture and infrastructure and their importance for human recreation. The United States, Japan and other developed forestry countries have long realized the importance of forest ecological function. They attach great importance to the forest ecological function, confirm the status of forest ecological products, and include them into the cost of forest management, so that the forest ecological function products have commodity attributes and realize the economic compensation of forestry. These countries have also developed empirical formulas or models for economic accounting of forest ecological function, which makes the evaluation results intuitive and clear, and provides a scientific basis for the rational allocation of forest resources and forestry development.

The continuous inventory data of forest resources has the advantages of high quality, providing long time series data and high survey accuracy [5-7]. In addition, more than 60 attributes of the sample plot contain many investigation factors directly related to forest ecological function. Therefore, the fixed sample data of continuous inventory of forest resources as the main data source can accurately describe the regional forest ecological function of Guangzhou from 1979 to 2012, and provide the basis for spatial-temporal dynamic analysis and driving factor analysis. The water conservation function of forest is mainly reflected in the increase of available water resources, the purification of water quality and the regulation of runoff; the water and soil conservation function is mainly reflected in reducing surface erosion, reducing the loss of soil fertility, improving soil structure, and reducing the surrounding sand hazards.

2. Materials and methods

2.1 Soil profile

The soil in the study area is acidic with pH between 6.7 and 4.5. According to the vertical zone of the land, the landforms can be divided into the following types: the mountains with an altitude of 400 m ~ 500 m are called middle and low mountains, with a slope of > 25 ° and mainly distributed in the northeast of Guangzhou City; the altitude of 400 m ~ 500 m is called the middle and low mountains In the vertical zone below m is called slope land, which is mainly distributed between mountains, basins, valleys and plains.

The parent material of soil formation is mainly composed of sand shale, granite and metamorphic rock; the altitude is 80 The gentle slope land or low flat slope land with slope less than 15 ° m is called terrace land. The parent material of soil formation is mainly accumulated red soil, red rock series and sand shale. The alluvial plain mainly exists in the Pearl River delta plain. The tidal flat is mainly distributed in the Nansha and Wanqingsha coastal areas.

2.2 Data standardization

Data are different in magnitude and dimension. In order to ensure the objectivity and scientificity of data, it is necessary to standardize the evaluation factors according to certain standards. Principal component analysis (PCA) is a multivariate statistical analysis method, which uses the idea of dimension reduction and transforms a group of linearly related variables into a group of linearly unrelated variables through orthogonal transformation, and explains the relationship between variables. Firstly, the original data are standardized, and the variance contribution rate and cumulative contribution rate of principal components are calculated to determine the number of principal components. Secondly, the eigenvector in the characteristic equation is calculated to construct the comprehensive evaluation function. By calculating the score and comprehensive score of each principal component, the comprehensive evaluation value of the principal component is finally obtained

2.3 Random forest regression

Random forests is a statistical method. When building the model, the program uses a bootstrap method to sample multiple samples from the modeling samples, and models the sampling. The values of multiple prediction models are combined to get the final prediction results. In this paper, the construction of random forest regression model is mainly realized by forest random forest model in R language software rat package. NTree is the number of regression trees, that is, the number of sampling times using bootstrap method. The selection of the number should not be too much, but also ensure that each data in the original training data set has a chance to be selected. Mtry is the number of random features, which is 1/3 of the number of independent variables selected when building the random forest regression model. When the number of independent variables is less than 3, the value of mtry is 1 by default.

3. Results and analysis

3.1 Spatial autocorrelation analysis

Spatial autocorrelation refers to the potential interdependence of some variables in the same distribution area [76]. The basic measure of spatial autocorrelation agglomeration is Moran's I value. When Moran's I > 0, it indicates that forest ecological function index is positively correlated in space, and the significant degree of spatial positive correlation increases with the increase of value; when Moran's I < 0, it indicates that there is spatial negative correlation in forest ecological function index, and the smaller the value, the greater the spatial difference. Because the existence of spatial autocorrelation is the premise of geographic weighted regression analysis, this paper uses spatial autocorrelation (Moran's I) function in ArcGIS toolbox to analyze the spatial autocorrelation of forest ecological function index in each period of the study area.

Table 1 Spatial autocorrelation of forest ecological function index

Years	Moran's I	Z scores
1979	0.173	1.053
1983	0.223	0.410
1988	0.265	4.280
1992	0.056	0.993
1997	0.148	2.455
2002	0.172	2.808
2007	0.162	2.678
2012	0.203	3.329

3.2 Random forest index analysis

Random forest regression (RF) algorithm is a popular ensemble learning method. In this experiment, the forest model in rat package of R language was used to simulate the random forest regression, and the nTree and mtry values were tested for many times by empirical trial method. When nTree = 500 and mtry = 2, the mean square residual error of the random forest model reaches the minimum, which can explain 52.29% of the variables.

Vegetation coverage is the most important driving factor for forest ecological function index, followed by slope and altitude. The greater the vegetation coverage, the greater the number of vegetation in the forest, the more obvious the improvement effect and impact on the stand and soil, and the more suitable for forest growth and ecological benefits. The area with higher altitude and slope is more difficult to cut, and the degree of forest resources can be cut down in this area is relatively light. Sufficient precipitation is a necessary condition for plant growth. To a certain extent, the increase of precipitation accelerates the growth and development of trees, and accelerates the formation of stable community structure of forest, so as to better play the ecological function and benefit of forest.

Table 2 Statistics on contribution rate of random forest regression factors

Index	%IncMSR	IncNodePurity
Recover	27.97	2.42
Altitude	17.01	2.55
Slope	12.87	2.06
Rain	11.17	1.28
Light	8.97	1.17
People	8.60	1.15
GDP	7.39	1.08
Temperatur	6.03	1.08

4. Conclusion

The similarity between GWR, Pearson and random forest regression results is that the three factors with the highest correlation or contribution rate are vegetation coverage, slope and altitude. That is to say, no matter what method is used to analyze the driving factors of forest ecological function index in Guangzhou in 2012, the three factors with the highest correlation or contribution rate are vegetation coverage, slope and altitude. The difference is that the results of GWR analysis and Pearson correlation analysis show that terrain factor is the most important driving factor, followed by vegetation coverage factor; while random forest regression results show that vegetation coverage is the most important driving factor, followed by terrain factor. The reason may be that R^2 of GWR is 0.55, which can only explain 55% variance, while random forest regression can explain 52.29% variance, both of which are relatively low, so there will be a slight difference in the results.

Geographic weighted regression analysis showed that the main driving factors were altitude, vegetation coverage, slope and GDP. Altitude, vegetation coverage and slope have a positive impact on forest ecological function index, while GDP has a negative impact on forest ecological function index.

References

[1] Snelder, T. , Lehmann, R. , Lamouroux, R. , Leathwick, R. , & Allenbach, R. . (2010). Effect of classification procedure on the performance of numerically defined ecological regions. Environmental Management, 45(5), 939-952.

[2] Dong, Hui, Ko, Shin, Taek, & Jeong, et al. (2017). Distribution and trend analysis of the significant wave heights using kma and ecmwf data sets in the coastal seas, korea. Journal of Korean Society of Coastal and Ocean Engineers, 29(3), 129-138.

[3] Anna Rita Germani. (2007). The environmental enforcement in the civil and the common law systems. a case on the economic effects of legal institutions. Quaderni Dsems, 157(1), 63.

[4] Twerefou, D. K. , Danso-Mensah, K. , & Bokpin, G. A. . (2017). The environmental effects of economic growth and globalization in sub-saharan africa: a panel general method of moments approach. Research in International Business and Finance, 42(DEC.), 939-949.
[5] Gardingen, P. R. V. , Mcleish, M. J. , Phillips, P. D. , Fadilah, D. , Tyrie, G. , & Yasman, I. . (2003). Financial and ecological analysis of management options for logged-over dipterocarp forests in indonesian borneo. Forest Ecology & Management, 183(1-3), 1-29.
[6] Karim, A. , Fazlina, Y. D. , Syakur, & Fathanah, N. . (2020). Evaluation of the critical level of mangrove ecosystems using spatial technology in east aceh coastal areas. IOP Conference Series: Earth and Environmental ence, 425(1), 012015 (7pp).
[7] Nunes, L. M. , Zhu, Y. G. , Stigter, T. Y. , Monteiro, J. P. , & Teixeira, M. R. . (2011). Environmental impacts on soil and groundwater at airports: origin, contaminants of concern and environmental risks. Journal of Environmental Monitoring, 13(11), 3026.