

# *Prediction and assessment model of the impact of different grazing strategies on Xilingole grassland*

**Bo Li, Yixin Cui**

*School of Geophysics and Information Technology, China University of Geosciences (Beijing),  
Beijing, China*

**Keywords:** Grassland grazing strategy, Mechanism analysis, Random forest classification algorithm, Model

**Abstract:** Grassland grazing strategies are crucial to the balance of grassland ecosystems and are closely related to the livelihoods of herders. The purpose of this paper is to study the effects, future trends of different grazing strategies on soil physics properties and vegetation biomass in Xilingole grassland. Based on the 19th graduate student mathematical modeling dataset, the raw data were preprocessed using Python, including missing value interpolation, normality test, Laeta criterion (also known as  $3\sigma$  criterion) to eliminate outliers, and then combined with random forest classification algorithm and logistic regression algorithm to establish a mathematical model of the effects of different grazing strategies on soil physical properties and vegetation biomass, and analyzed in detail by their relationships. The results showed that (1) the grazing strategy and the changes of soil and vegetation in the area were not significant. The results show that (1) the grazing strategy contributes more to vegetation biomass than soil physical properties, and the accuracy of the training set for random forest classification is 0.98, and the accuracy of the test set is 0.74. (2) The linear function of grazing intensity on soil physical properties and vegetation biomass is established by combining with logistic regression algorithm. For different grazing methods, the final grazing strategy results can be obtained by multiplying the grazing intensity by a weighting factor, and subject to the influence of multiple factors, the soil physical properties and vegetation biomass do not always maintain a positive or negative correlation with the change of grazing strategy, which is consistent with the actual grazing strategy influence.

## **1. Introduction**

Grassland grazing strategies have always been a hot topic of research in grassland ecosystems. China, as the second largest grassland country in terms of area, has actively adopted policies such as "returning pasture to grass" to protect and improve the grassland ecosystem. Among them, Xilingole grassland in Inner Mongolia is one of the four major grasslands in China and an important base for livestock industry, as well as an important ecological barrier that has a strong inhibitory effect on the occurrence of dust storms and severe weather. Therefore, a reasonable grassland grazing strategy can effectively protect the grassland ecosystem, promote regional economy, prevent grassland desertification and protect people's livelihoods, and provide a scientific basis for national and governmental grazing policies and grassland management decisions[1].

Therefore, researchers use emerging technological tools and algorithms to study grassland ecosystems to give scientific predictions and evaluations for sustainable resource development and ecological balance. To study the effect of wind and dust on soil desertification, Wu, Jing et al. used multiple regression algorithms to analyze a time series dataset from 2001-2018 obtained from the Xilinguole grassland, China, and concluded that an increase in livestock load enhances the effect of wind and dust [2]. For vegetation and precipitation, Chi, Dengkai et al. combined vegetation net primary production (NPP) and precipitation variability analysis using residual trend method and fixed-size moving window method to conclude that grazing and precipitation are the main drivers of grassland vegetation change [3]. In addition, land degradation is one of the key concerns of grassland grazing strategies [4], Du, Shuai et al. analyzed the effect of soil resistance at different levels of grazing intensity using principal component analysis and concluded that grazing intensity had no significant effect on soil desertification and resistance [5], while Sun, Bin et al. used remote sensing to create a time-series of 2001-2012 Sun, Bin et al. used remote sensing to create time-series net initial yield data and determined the area of grassland degradation using the Sen+Mann-method and analyzed the 12-year grassland drivers using multiple and partial regression [6]. The influence of grassland biomass is also important for grassland development. Wang, Hao et al. analyzed the spatial variation, evolutionary characteristics and future trends of grassland biomass in Xilingole based on the Globeland30 dataset provided by China National Geospatial Information Center for 2000, 2010 and 2020 combined with logistic regression algorithm and concluded that overall as forest and grassland cover rises, the biospatial volume will also continue to grow [7]. The application of remote sensing technology also plays an active role in giving the grazing strategy of grassland ecosystems. Ye, Junzhi et al. applied random forest classification algorithm to create a land use and cover dataset of Xilinguole grassland for the past 20 years (2000-2020) using Google Map engine and Landsat satellite remote sensing image technology tools, and used principal component analysis and multiple linear stepwise regression for its spatial and temporal dynamics analysis and multiple linear stepwise regression to further analyze the driving mechanisms of grasslands with high progress.

In previous studies, the developmental impact on grassland ecology has been studied in terms of several drivers, however, the analysis of the impact of grazing strategy on plant growth and soil physical properties is basically absent. Among them, grazing strategies are mainly divided into grazing methods and grazing intensity. Grazing methods can be divided into five types: continuous grazing throughout the year, grazing ban, selective zoned rotational grazing, light grazing, and growing season rest, and grazing intensity can be divided into four types: control, light grazing intensity, moderate grazing intensity, and heavy grazing intensity. A reasonable grazing strategy can promote the cycle of grassland ecosystems and contribute to the global goal of sustainable development. Therefore, this paper distinguishes the importance of data sets according to a random forest classification algorithm and uses a logistic regression model to establish the effects of different grazing practices on soil physical properties and vegetation biomass. The study aims to achieve the following three objectives.

(1) To construct a set of prediction and assessment models applicable to the grazing strategies on soil physical properties and vegetation biomass in Xilingole grassland.

(2) To analyze the spatial and temporal evolution characteristics and future development trends of different grazing strategies.

(3) To provide recommendations for the government to address the potential impacts of grazing strategies.

## 2. Problem Analysis and Model Construction

### 2.1 Analysis of Question

According to the proposed problem, the effects of different grazing strategies, i.e. grazing methods and grazing intensity, on the physical properties of grassland soil (mainly soil moisture) and vegetation biomass were modeled for the Xilingole grassland. Considering that the data provided in the dataset have information under different grazing intensities, but their characteristics on soil physical properties and vegetation biomass are scattered, firstly, the systematic integration of all data is important for mining effective information. Secondly, the data provided inevitably have certain problems due to historical reasons, for example, the history is too long paper or computer data is not complete. Therefore, scientific and effective pre-processing of the data is decisive for the construction of models to determine the effects of different grazing strategies on the physical properties of grassland soils and vegetation biomass.

Data processing was performed to extract as many features as possible to ensure the accuracy of the modeling approach, and data integration was performed on the dataset using Pycharm software. After consolidation, 171564 samples with 48 features were obtained, which contained variables such as grazing intensity, time, latitude and longitude, soil moisture and runoff. The data processing showed that it was not difficult to find that there were missing data for nutrient seedlings, plant number and fresh weight, with missing proportions of 0.17 %, 2.50 % and 4.93%, respectively. Therefore, since the missing proportion was very small, the mean interpolation was selected to process the missing values for these three characteristic samples. Secondly, to reduce the effect of random errors coming from the outliers, it is also necessary to remove the outliers, and SPSS software can be used to test the normality of the data after processing the missing values to determine whether the outliers in the 171564 samples can be removed using the Lajda criterion. After the missing values and outliers were processed, a mathematical model was established using random classification and logistic regression methods, and finally the features were classified by random forest quadratic classification, and then the logistic regression coefficients were used to judge the model of the effect of their grazing intensity on the physical properties of grassland soils and the amount of vegetation growth. Since there is a strong correlation between grazing method and grazing intensity, i.e., it can be expressed in the form of the product of weight coefficient and grazing intensity. The detailed thought process is shown in Figure 1.

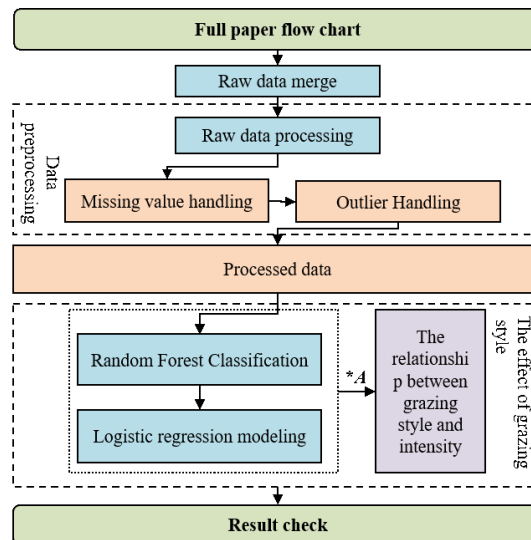


Figure 1: Flowchart of ideas

## 2.2 Study area

Xilinguole Meng is located in the central part of Inner Mongolia Autonomous Region, China, with its latitude and longitude range from to and to. In terms of administrative divisions, Xilinguole has 12 county-level administrative regions with a total area of more than 200,000 square kilometers, as shown in Figure 2.

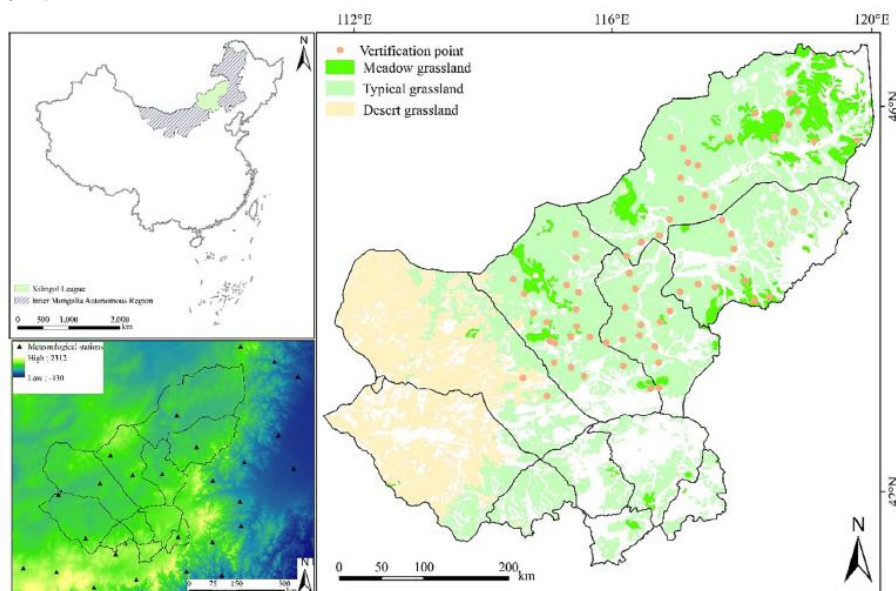


Figure 2: Location of the study area in China and distribution of grassland types

The total area of Xilinguole grassland is 192,000 km<sup>2</sup>, and the area of grassland available for grazing is 176,000 km<sup>2</sup>, accounting for 92% of the total grassland area. Grasslands are divided into three types from east to west: meadow grassland, typical grassland and desert grassland [9]. The representative plants of meadow grasslands are *Scutellaria baicalensis*, *Sheepgrass* and *Chrysanthemum*; the typical grasslands are *Big Needlegrass*, *Kessler's Needlegrass* and *Square-Scale Cryptomeria*; the desert grasslands are *Needlegrass spp.* and *Flat-Spike Cryptomeria spp.* Plants in Xilinguole grasslands usually turn green in early April and stop growing in September [10]. Herders start cutting grass from mid-August to early September so that they can store fodder for their livestock for winter.

## 2.3 Data processing and methods

This problem uses Python program to integrate data mining and pre-processing of the attachment data. First, the attachment data are extracted and merged to build a two-dimensional table data form after the original data through data processing shows statistics found that there are missing data for nutrient seedlings, plant number and fresh weight, the missing percentage is 0.17 %, 2.50 % and 4.93% respectively, but because the missing percentage is very small, this problem selects the basic mean interpolation for these three characteristics samples to handle the missing values. Mean interpolation is the interpolation of data using the mean or plurality of the sample data as its alternative value. For outlier detection, since the normal distribution has a very typical graphical feature of high in the middle and low on both sides, if the sample data does not obey the normal distribution, we can quickly distinguish it by histogram, so for the study of normality of data distribution, our preferred method is graphical observation. Using histograms and P-P plots for observation, assuming that the data obey normal distribution, as shown in Figure 3, we randomly selected four groups of data for testing, which can be seen from the graph that the normality is not very good, that is, there are outliers. Therefore,

the outliers need to be processed for their characteristic samples.

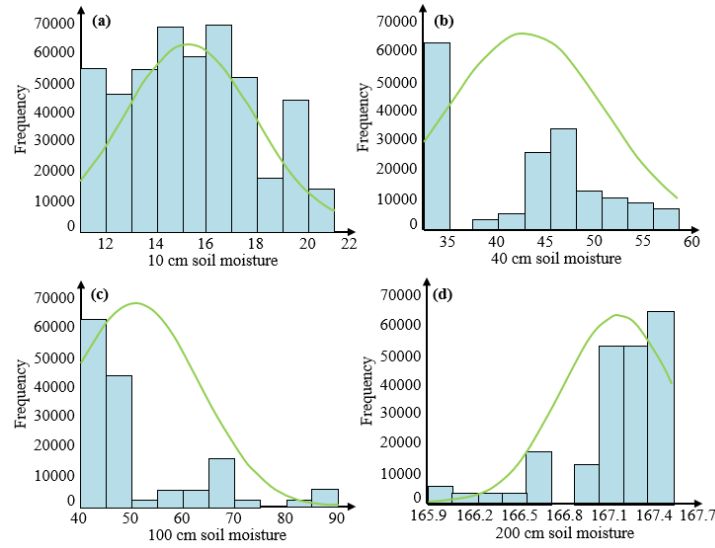


Figure 3: Histogram of normality test

## 2.4 Construction of predictive models

Random forest (RF) classification an integrated classifier that uses a set of decision trees to make predictions and applies a voting mechanism to the results. Specifically, each decision tree is judged independently and each node is segmented using a user-defined number of features. The final classification decision is made by averaging the class assignment probabilities computed from all spanning trees, and the class with the most votes is the final class selected. Despite the high dimensionality of the sample features, RF allows the training and classification processes to be highly parallelized and run efficiently, thereby improving the overall predictive performance of the model and reducing overfitting. The random forest classification code was written using Python language and the selected training set was 70% and the test set was 30% for random forest training. To facilitate the classification process for mapping grazing intensity to definite classes, the intensity relationships given in the title question, i.e., control (NG, 0 sheep/day/ha), light grazing intensity (LGI, 2 sheep/day/ha), moderate grazing intensity (MGI, 4 sheep/day/ha), and heavy grazing intensity (HGI, 8 sheep/day/ha), were divided into tables as follows (Table 1).

Table 1: Mapping relationship of grazing intensity

Grazing intensity (sheep/day/ha)	Mapping class variables
NG	0
LGI	1
MGI	2
HGI	3

To further analyze the importance of random forest classification for each indicator feature on the dependent variable, 18 features that can represent the physical properties of soil and vegetation biomass, respectively, were selected as shown in Figure 4, and now the importance of each feature can be obtained after normalizing the sample variables for random forest classification, and the importance indicates the degree of influence of the matter on the predicted results and grazing intensity, and the smaller the importance, the smaller the influence. It is easy to see that the physical properties are significantly less important than the chemical properties of the soil, which in turn makes the degree of plant influence greater, indicating that plant biomass is more influenced by grazing

intensity and less by the physical properties of the soil.

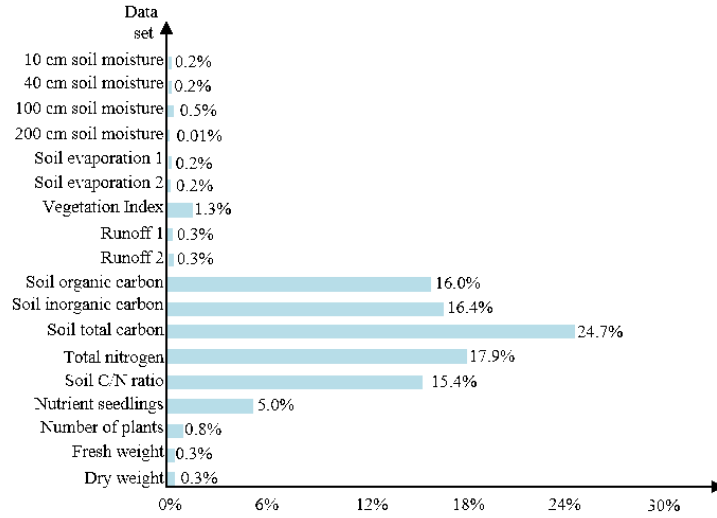


Figure 4: Importance of random forest classification features

Based on the random forest classification, in order to further build the prediction and evaluation model, this paper adopts the algorithm of logistic regression, with grazing intensity as the dependent variable and feature sample data set as the independent variable for model solving.

Logistic regression, also known as logistic regression analysis, is a generalized linear regression analysis model, which is a kind of supervised learning in machine learning. The derivation process and computation are similar to that of regression, but it is actually mainly used to solve classification problems. The model is trained by a given n sets of data (training set), and a given set or sets of data (test set) are classified at the end of the training. Each of these sets of data is composed of p indicators.

The general equation of logistic regression is

$$\ln \frac{y}{1-y} = w^T x + b \quad (1)$$

Among them, y is the independent variable, x is the dependent variable,  $w^T$  is the regression coefficient, b is the intercept. Generally, the probability method is used to solve the equations, such as the maximum likelihood estimation method. In this paper, the maximum likelihood estimation is used to solve the equation. Python is used to write a program to divide the pre-processed data into 70 % of the test set and 30 % of the data set. After that, it is substituted into the logical regression equation and solved. The corresponding regression coefficients and intercept results are obtained respectively. For convenient analysis, the regression coefficients are visualized by histogram, as shown in Figure 5. For the effect of soil physical properties, take soil moisture as an example, when grazing intensity is NG and LGI, soil moisture and grazing intensity show a negative correlation, that is, the lighter the grazing degree, the greater the soil moisture, corresponding to basic common sense. When the grazing intensity is MGI and HGI, the grazing intensity is positively correlated with the moisture content, i.e., the greater the grazing intensity, the greater the soil moisture content, mainly because a certain degree of grazing promotes the growth of vegetation and thus increases the soil moisture content, but once a certain limit is exceeded, the balance will be broken, i.e., overgrazing will cause a serious reduction in soil moisture content. Therefore, the hypothesis of this question is a model in which the number of livestock has not yet reached a threshold value, and it belongs to the ideal state of the effect of grazing intensity on soil physical properties.



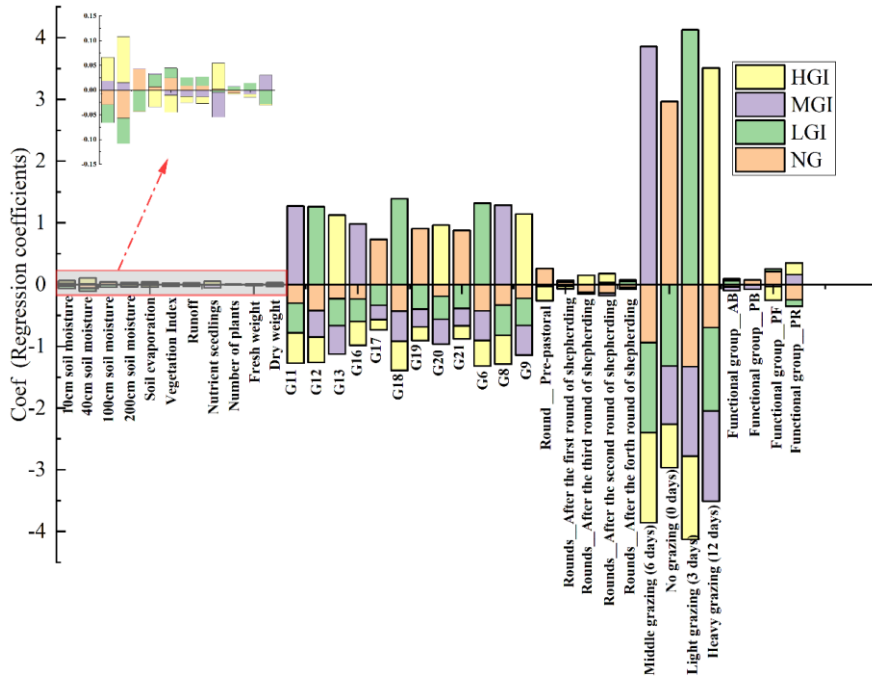


Figure 5: Histogram of regression coefficients of grazing intensity and characteristic samples

For the effect of plant biomass is expressed by vegetation index, and in this paper, to improve the final accuracy of logistic regression, the index is expanded to optimize the model with certain constraints. Similarly, it can be seen from the regression coefficients that the vegetation index has a greater influence for the model in this paper. When the grazing intensity belongs to NG or LGI, the vegetation index shows a positive correlation with the grazing intensity, that is, a certain amount of grazing can promote the ecological balance of grassland and strengthen the growth of grassland vegetation, especially the livestock manure indirectly promotes the improvement of the vegetation index of grassland. With the strengthening of grazing intensity, grazing intensity and vegetation index show an obvious negative correlation, and the regression coefficient is not obvious here because of the existence of self-regulation of the ecosystem, but when the grazing intensity gradually increases, it will gradually destroy the ecological balance and lead to the decrease of vegetation index, resulting in land desertification, which is the main reason why more features are needed to make the logistic regression more accurate.

Based on the equations of the logistic regression, it is not difficult to write the final mathematical function model.

$$GI = w^T x + b \quad (2)$$

where GI denotes grazing intensity, and thus for grazing mode, since it only adds a constraint in time, it is actually equivalent to multiplying a weighting factor A for different grazing intensities, so the following mathematical relationship can be established for grazing mode GI and grazing intensity WI, i.e.

$$WI = A \times GI \quad (3)$$

## 2.5 Testing of the model

The use of random forest classification method needs to check the accuracy and accuracy slightly of the corresponding training set and test set, as shown in Table 2, it is easy to see that the training set and test set accuracy of random forest classification are high, more ideal to meet the theoretical

analysis.

Table 2: Model evaluation results

	Accuracy	Recall	Precision
Training set	0.989	0.989	0.989
Test set	0.701	0.701	0.737

The accuracy and precision of the logistic regression method corresponding to four different intensities in both the training and test sets are close to 100%, which is sufficient to show that when the constraint features are increased, the results are closer to the actual grazing strategy on the physical properties of the soil and the amount of plant growth, so the regression model developed in this paper is an excellent inversion of the state of the grassland ecosystem in the region.

### 3. Conclusion

This paper constructs a set of prediction and assessment models based on random forest classification and logistic regression algorithms applicable to Xilinguole grassland grazing strategies on soil physical properties and vegetation biomass. The model fully illustrates the effects of different grazing strategies on soil physical properties and vegetation biomass in Xilinguole grassland in the recent decade of 2010-2022, and the grazing intensity is the main influencing factor, which is very likely to bring a greater impact on vegetation biomass. Under a certain grazing intensity, it can effectively improve the stability of grassland ecosystem and achieve the goal of sustainable development, but overgrazing or no grazing, etc. are not conducive to Therefore, under comprehensive consideration, the relevant local government can improve the quality of life of herding households through incentive funds, so that they can control the livestock load within a certain range and graze regularly and with appropriate intensity in order to effectively organize the problems of land degradation as well as land salinization and slabbing.

### References

- [1] Zhen, L., et al. "Herders' willingness to accept versus the public sector's willingness to pay for grassland restoration in the Xilingol League of Inner Mongolia, China." *Environmental Research Letters* 9.4 (2014): 045003.
- [2] Wu, Jing, Yasunori Kurosaki, and Chunling Du. "Evaluation of climatic and anthropogenic impacts on dust erodibility: A case study in Xilingol Grassland, China." *Sustainability* 12.2 (2020): 629.
- [3] Chi, Dengkai, et al. "Assessing the effects of grazing on variations of vegetation NPP in the Xilingol Grassland, China, using a grazing pressure index." *Ecological Indicators* 88 (2018): 372-383.
- [4] Wieland, Ralf, et al. "Identifying drivers of land degradation in Xilingol, China, between 1975 and 2015." *Land Use Policy* 83 (2019): 543-559.
- [5] Du Shuai, et al. "Grazing does not increase soil antibiotic resistome in two types of grasslands in Inner Mongolia, China." *Applied Soil Ecology* 155 (2020): 103644.
- [6] Sun Bin, et al. "Grassland degradation and restoration monitoring and driving forces analysis based on long time-series remote sensing data in Xilin Gol League." *Acta Ecologica Sinica* 37.4 (2017): 219-228.
- [7] Wang Hao, Yunfeng Hu, and Yuting Liang. "Simulation and spatiotemporal evolution analysis of biocapacity in Xilingol based on CA-Markov land simulation." *Environmental and Sustainability Indicators* 11 (2021): 100136.
- [8] Ye Junzhi, et al. "Analysis on Land-Use Change and its driving mechanism in Xilingol, China, during 2000–2020 using the google earth engine." *Remote Sensing* 13.24 (2021): 5134.
- [9] Tong Chuan, et al. "A landscape-scale assessment of steppe degradation in the Xilin River Basin, Inner Mongolia, China." *Journal of Arid Environments* 59.1 (2004): 133-149.
- [10] Chi, D., et al. "The variability of growing season of different vegetation types in Xilingol League." *Pratacultural Sci* 33 (2016): 1825-1834.