ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202108.00074

Postprint: Vegetation Change Simulation Based on BP-SVM Model

Authors: Jia Songtao

Date: 2021-08-12T12:08:15+00:00

Abstract

Vegetation serves as a crucial link connecting the biosphere, atmosphere, and hydrosphere, exerting significant influences on the ecological environment and water-heat regime variations within watersheds. Current research predominantly focuses on correlation analyses between NDVI and climate factors, while few studies predicting NDVI have neglected to consider lag effects and the impact of incorporating additional predictor variables on improving model accuracy. Based on this, the present study compares Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Support Vector Machine (SVM) models to identify the model with superior accuracy. Beyond conventional factors (precipitation and temperature), soil moisture and sunshine factors influencing vegetation growth were incorporated, and lag effects between different factors and NDVI were considered. The research results demonstrate: (1) The Support Vector Machine (SVM) model exhibits the strongest fitting capability and highest NDVI prediction accuracy, with Root Mean Square Error (RMSE) reduced by over 1.8% in both the Jinghe and Beiluo River basins. (2) The inclusion of soil moisture, sunshine, and other factors improves model prediction accuracy in the Jinghe River basin, reducing RMSE by 8.8%. (3) After accounting for lag effects, RMSE in the Jinghe and Beiluo River basins is reduced by 15% and 11%, respectively, further enhancing NDVI prediction accuracy and increasing model reliability. The prediction results can serve as an important reference for formulating future ecological protection strategies and guiding ecological restoration efforts.

Full Text

Simulation of Vegetation Change Based on BP-SVM Model

JIA Songtao[1], HUANG Shengzhi[1], WANG Hao[2], LI Ziyan[1], HUANG Qiang[1], LIANG Hao[1]

[1]State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi' an University of Technology, Xi' an 710048, Shaanxi, China
[2]State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

Abstract

Vegetation serves as a critical link connecting the biosphere, atmosphere, and hydrosphere, exerting significant influence on watershed ecological environments and water-heat regime variations. Current research predominantly focuses on correlation analyses between the Normalized Difference Vegetation Index (NDVI) and climatic factors, while few predictive studies consider the effects of time-lag relationships or how incorporating additional predictive factors might enhance model accuracy. This study compares multiple linear regression (MLR), artificial neural network (ANN), and support vector machine (SVM) models to identify the most accurate approach. Building upon conventional factors (precipitation and temperature), we incorporate soil moisture and sunshine duration—factors that significantly affect vegetation growth—and examine the time-lag effects between different factors and NDVI. The results demonstrate: (1) The SVM model exhibits the strongest fitting capability and highest prediction accuracy, reducing root mean square error (RMSE) by over 1.8% for both the Jing River and Beiluo River basins. (2) Adding soil moisture and sunshine factors improves model prediction accuracy in the Jing River Basin, with RMSE decreasing by up to 8.8%. (3) When time-lag effects are considered, RMSE decreases by 15% and 11% for the Jing River and Beiluo River basins, respectively, further improving NDVI prediction accuracy and enhancing model reliability. These predictions provide important reference value for future ecological protection strategy formulation and ecological restoration guidance.

Keywords: support vector machine; prediction accuracy; prediction model; time-lag effect; vegetation coverage

Introduction

Vegetation functions as a natural link between the atmosphere, soil, and watershed, representing a vital component of terrestrial ecosystems and forming an integrated system where the climate, hydrosphere, lithosphere, biosphere, and human activities interact. As one of the most important indicators for measuring surface vegetation conditions, vegetation coverage plays crucial roles in plant transpiration, photosynthesis, and soil moisture evaporation. Land cover change, with vegetation change as its primary manifestation, indirectly reflects the degree of human impact on natural ecosystems—an influence that significantly affects water cycling processes. The Normalized Difference Vegetation Index (NDVI), calculated as the ratio of the difference to the sum of near-infrared and red band reflectance, serves as an effective parameter for characterizing surface vegetation coverage and plant growth trends, widely recognized for its sensitivity in measuring vegetation conditions.

Numerous scholars have investigated relationships between NDVI and climatic factors through extensive correlation analyses. Du et al. expanded short-term GIMMS NDVI series to long-term sequences using MODIS NDVI data, analyzing climate change impacts on vegetation and finding that temperature primarily promotes vegetation growth in spring and autumn, while summer precipitation represents the main factor regulating vegetation growth. Tao et al. employed trend and exponential analyses to examine temporal variations and future trends of vegetation coverage at annual and seasonal scales, revealing that vegetation growth is influenced by multiple factors including elevation, mean annual temperature, soil type, and population density. Li et al. analyzed relationships between vegetation changes and climatic factors (precipitation, temperature) in the wind-sand area along the Great Wall in northern Shaanxi, concluding that precipitation constitutes the primary factor causing NDVI interannual fluctuations, with non-climatic factors such as human activities also contributing to vegetation changes. For arid and semi-arid regions, vegetation growth additionally relates to soil characteristics and root systems, with these factors collectively determining plant growth status and coverage changes. Wang et al. calibrated productivity models for different vegetation types using carbon flux data, identifying total radiation and precipitation as key factors affecting seasonal vegetation growth in the study area.

Although these studies demonstrate that vegetation coverage results from multiple factors, most NDVI prediction research employs only precipitation and temperature as input variables. Zhang et al. constructed an NDVI prediction model using SVM optimized with three algorithms, achieving effective predictions. Cong et al. used precipitation and evapotranspiration as independent variables in a multiple regression model, effectively predicting seasonal vegetation dynamics. However, vegetation growth quality depends not on single climatic factors but on the combined effects of multiple climatic elements. Therefore, incorporating additional input factors closely related to NDVI is essential for improving prediction accuracy and model reliability.

Soil moisture, a critical input factor for terrestrial ecosystems, plays a decisive role in water and energy exchange and plant growth. Studies indicate that soil water content affects vegetation-available water storage, with temperature and sunshine showing stronger driving effects on vegetation growth than precipitation in high-latitude regions. Pei et al. quantitatively analyzed factors influencing maximum annual vegetation coverage, identifying precipitation, soil type, and land use as dominant factors. Liu et al. examined mechanisms affecting NDVI at annual and seasonal scales, finding that humidity and sunshine influence vegetation growth. Consequently, incorporating soil moisture and sunshine data as predictive factors, while considering the non-linear relationships and time-lag effects between vegetation growth and climatic elements, is crucial for constructing reliable multi-factor NDVI prediction models. Such models provide important references for future ecological protection strategies and local ecological restoration efforts.

This study compares MLR, ANN, and SVM models for simulating vegetation changes using precipitation and temperature data, then incorporates soil moisture and sunshine factors while considering time-lag effects between different factors and NDVI growth conditions, aiming to improve NDVI prediction accuracy and provide a basis for agricultural production, drought early warning, and ecological management in the study region.

1.1 Study Area Overview

The Jing River and Beiluo River, two major tributaries of the Wei River Basin (Fig. 1), are located in eastern Northwest China and experience a continental monsoon climate with concentrated summer precipitation and scarce winter rainfall. The region exhibits pronounced seasonal temperature variations. Loess constitutes the primary soil type with low water content. Sunshine duration is relatively short, averaging approximately 2,200 hours annually. Dominant vegetation includes larch, apricot, peach, locust trees, grasslands, and crops. The study focuses on these two tributaries: the Jing River Basin, where irrigated areas account for 0.38% of the basin area, and the Beiluo River Basin, with a larger irrigated area comprising 0.66% of its basin area.

1.2 Data Sources

Data include meteorological observations, soil moisture measurements, and NDVI data from the China Meteorological Data Service Center (http://data.cma.cn), the Global Land Data Assimilation System (GLDAS, http://disc.sci.gsfc.nasa.gov/hydrology/data_{holdings}), and the Geospatial Data Cloud (http://www.gscloud.cn). Monthly precipitation, temperature, and sunshine data from 2000–2015 were collected for ten meteorological stations (Huanxian, Guyuan, Pingliang, Xifeng, Changwu in the Jing River Basin; Wuqi, Yan' an, Luochuan, Tongchuan in the Beiluo River Basin) and converted to basin-scale data using Thiessen polygons. Station locations are shown in Fig. 1.

1.3 Methods

1.3.1 Multiple Linear Regression Multiple linear regression explains a dependent variable using two or more independent variables, assuming linear relationships and calculating regression coefficients via least squares method:

$$Y = \beta_0 + \beta_{1X}1 + \beta_{2X}2 + \cdots + \beta_{mX}m + \varepsilon$$

where Y is the dependent variable, X_n are independent variables ($n = 1, \ldots, m$), β_n are regression coefficients, and ε represents random error.

1.3.2 Support Vector Machine Support Vector Machine determines kernel functions to construct optimal hyperplanes for solving complex non-linear fitting problems. Through non-linear mapping $\phi(x)$, SVM transforms data into high-dimensional feature space to seek optimal regression functions:

$$f(x) = (w, \phi(x)) + b$$

where $\phi(x)$ is the non-linear mapping function, w represents weights, x is the input variable for sample i, and b is the threshold.

1.3.3 Artificial Neural Network Artificial Neural Network employs backpropagation to continuously update weights and thresholds, demonstrating strong non-linear mapping and self-learning capabilities:

$$X_n = \sum_{i=1}^{l} w_{in} \cdot f\left(\sum_{k=1}^{m} w_{ik} x'_k - \theta_i\right) - \theta_n$$

where X_n is the output layer, X_i represents hidden layer values, w_{ik} denotes weights from input neuron k to hidden neuron i, x'_k are input values, w_{in} represents weights from hidden neuron i to output neuron n, and θ_i, θ_n are thresholds.

The network comprises input variables, output variables, hidden layers, and activation functions. During training, samples are preprocessed to determine network structure, initialize weights and thresholds, and iterate until achieving preset accuracy.

1.3.4 Multi-factor Prediction Model Considering Time-lag Effects To enhance model reliability and NDVI prediction accuracy, this study considers factor lag effects and incorporates additional factors (soil moisture and sunshine) beyond precipitation and temperature.

First, monthly precipitation and temperature data serve as input variables for MLR, ANN, and SVM models to compare performance and identify the optimal model. Second, soil moisture and sunshine data are added as input factors to assess accuracy improvements. Finally, time-lag data for NDVI, precipitation, temperature, soil moisture, and sunshine are integrated into the SVM model.

Model performance is evaluated using Nash-Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE), and Mean Relative Error (MRE):

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - t_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i are observed values, t_i are predicted values, and $\bar{y}$ is the mean observed value. NSE approaching 1 indicates good model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - t_i)^2}$$

Smaller RMSE values indicate closer agreement between predicted and observed values.

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - t_i}{y_i}\right| \times 100\%$$

The modeling process is illustrated in Fig. 2.

2 Results

2.1 Optimal Prediction Model Selection

Using 2000–2015 monthly precipitation and temperature data from the Jing River Basin as input vectors, MLR, ANN, and SVM models were applied to predict NDVI (training period: 2000–2010; validation period: 2011–2015). Prediction accuracies varied significantly across models (Table 1).

In the Jing River Basin, MLR achieved a validation NSE of only 0.62 with RMSE of 0.089 and MRE of 8.91%. ANN improved validation NSE to 0.73 (RMSE: 0.074; MRE: 7.39%). SVM outperformed both, achieving validation NSE of 0.75—4.04% higher than ANN and 21.9% higher than MLR. Training period NSE for SVM was 14.64% higher than ANN and 2.53% higher than MLR. In the Beiluo River Basin, SVM similarly demonstrated superior performance, with validation NSE 8.52% higher than ANN and 1.47% higher than MLR, and RMSE reductions of 21.9% and 8.52% respectively.

These results confirm that while ANN effectively captures complex non-linear relationships and outperforms MLR, SVM' s advantage lies in transforming optimization problems into convex quadratic programming, thereby avoiding local optima issues that plague ANN while handling non-linear relationships more effectively than MLR. Consequently, SVM provides the highest prediction accuracy.

2.2 Input Data Selection for Prediction Models

Most studies consider only precipitation or temperature effects on NDVI, yet Fig. 3 reveals complex non-linear relationships between NDVI and various climatic factors plus soil moisture. To improve accuracy, additional relevant input factors are necessary. This study added soil moisture and sunshine to precipitation-temperature inputs in the SVM model (selected for its superior performance) and compared results.

In the Jing River Basin, the two-factor model (precipitation + temperature) achieved NSE of 0.75, while the multi-factor model (adding soil moisture and sunshine) improved NSE and reduced RMSE by 8.806% and MRE by 5.649% (Table 2). In the Beiluo River Basin, the multi-factor model increased NSE by 1.06% and reduced RMSE by 34.93% and MRE by 39.46%. The improvement was less pronounced in the Beiluo Basin, likely due to its larger irrigated area (0.66% vs. 0.38%) and greater human influence.

2.3 Time-lag Effects of Regional Vegetation Response to Climate Factors

Vegetation growth depends not only on current but also on preceding climate conditions. Correlation coefficients between NDVI and each factor were calculated to determine optimal lag times. Factors with correlation coefficients exceeding 0.3 (significant at 95% confidence level) were selected. In the Jing River Basin, NDVI showed 1-month lag for precipitation and temperature, 2-month lag for soil moisture, and 1-month lag for sunshine. In the Beiluo River Basin, optimal lags were similarly identified.

Using these lagged factors as SVM inputs significantly improved predictions (Table 4). In the Jing River Basin, the lag-aware model reduced RMSE by 24.39% and MRE by 23.50% compared to the non-lagged model. In the Beiluo River Basin, NSE improved substantially with RMSE decreasing by 51.3%. Fig. 4 illustrates the enhanced simulation results for both basins.

3 Discussion

The superior performance of lag-aware models likely stems from plant physiological resilience and water storage capacities, enabling delayed responses to climate changes. This study’ s incorporation of soil moisture and sunshine as sensitive factors improved Jing River Basin predictions, suggesting vegetation growth responds to multiple factors rather than single climatic elements. The minimal improvement in the Beiluo River Basin may reflect its extensive irrigation (0.66% of basin area) and stronger human intervention compared to the Jing River Basin (0.38% irrigation). While region-specific sensitive factors require local optimization, the proposed modeling framework is transferable. Future research should explore NDVI influence mechanisms, integrate teleconnection factors, and develop hybrid models using decomposition techniques to further improve accuracy and quantitatively separate climate change from human activity impacts.

4 Conclusions

This study compared MLR, ANN, and SVM models for predicting NDVI in the Jing and Beiluo River basins using precipitation and temperature data, then incorporated soil moisture and sunshine factors while analyzing input factor time-lag effects. Key conclusions include:

1) Model prediction accuracy ranks as: SVM > ANN > MLR, confirming complex non-linear relationships between meteorological factors and vegetation coverage.

2) Adding soil moisture and sunshine data to SVM inputs improved accuracy and reliability over precipitation-temperature-only models, demonstrating vegetation growth responds to multiple factors. Beiluo Basin improvements were less significant due to greater human influence.

3) Lag-aware models substantially outperformed non-lagged models in both basins, as vegetation responses to different climate factors exhibit varying time lags. Considering these lag effects is crucial for revealing vegetation-climate interactions.

The prediction framework provides theoretical support for formulating ecological protection strategies, guiding restoration efforts, and informing agricultural production and drought early warning systems under changing environmental conditions.

References

[1] Liu Jiafu, Ma Shuai, Li Shuai, et al. Changes in vegetation NDVI from 1982 to 2016 and its responses to climate change in the black soil area of Northeast China[J]. Acta Ecologica Sinica, 2018, 38(21): 7647-7657.

[2] Tian Tian, Li Shaocai, Chen Min, et al. Analysis on vegetation index' s long time series dynamics of Yalong River basin[J]. Journal of Hydroelectric Engineering, 2012, 31(2): 159-164.

[3] Yang Feng, Li Jianlong, Qian Yurong, et al. Estimating vegetation coverage of typical degraded grassland in the Northern Tianshan Mountains[J]. Journal of Natural Resources, 2012, 27(8): 1340-1348.

[4] He Hui, Yusufujiang Rusuli. Analysis of the relative role of vegetation cover changes and its influencing factors in Yili area from 2001 to 2015[J]. Journal of Central South University of Forestry & Technology, 2019, 39(10): 76-87.

[5] Guo Ni, Guan Xiaodan. An improvement of the vegetation condition index with applications to the drought monitoring in Northwest China[J]. Advances in Earth Science, 2007, 22(11): 1160-1168.

[6] Pei Zhilin, Yang Qinke, Wang Chunmei, et al. Spatial distribution of vegetation coverage and its affecting factors in the upper reaches of the Yellow River[J]. Arid Zone Research, 2019, 36(3): 546-555.

[7] Wu D H, Zhao X, Huang K C, et al. Time lag effects of global vegetation responses to climate change[J]. Global Change Biology, 2015, 21(9): 3520-3531.

[8] Du Jiaqiang, Gao Yun, Jiaerheng Ahati, et al. Spatio-temporal patterns and driving factors of vegetation growth anomalies in Xinjiang over the last three decades[J]. Acta Ecologica Sinica, 2016, 36(7): 1915-1927.

[9] Tao Shuai, Kuang Tingting, Peng Wenfu, et al. Analyzing the spatiotemporal variation and drivers of NDVI in upper reaches of the Yangtze River from 2000 to 2015: A case study of Yibin City[J]. Acta Ecologica Sinica, 2020, 40(14): 5029-5043.

[10] Li Dengke, Guo Ni, He Huijuan. Vegetation change and its relationship with climate in the region along the Great Wall in northern Shaanxi[J]. Acta Ecologica Sinica, 2007, 27(11): 4620-4629.

[11] Sun Rui, Chen Shaohui, Su Hongbo. Spatiotemporal variation of NDVI in different ecotypes on the Loess Plateau and its response to climate change[J]. Geographical Research, 2020, 39(5): 1200-1214.

[12] Wang Lunche. Regional Variations of Atmosphere Radiation and its Quantitative Effects on the Terrestrial Ecosystem Productivity[D]. Wuhan: Wuhan University, 2015.

[13] Zhang Mantun, Huang Chunmeng, Mi Na, et al. Combination forecast model of NDVI based on support vector machine regression[J]. Journal of Hebei University of Technology, 2017, 46(4): 39-45.

[14] Cong Xiaohong, Shi Bing, Yu Xida, et al. Prediction of sediment discharge at Lijin station of the Yellow River based on PSO-BP neural network[J]. Yellow River, 2020, 42(1): 1-8.

[15] Tian Siyuan, Van Dijk A I J M, Paul Tregoning, et al. Forecasting dryland vegetation condition months in advance through satellite data assimilation[J]. Nature Communications, 2019, 10: 469.

[16] Yang Xi, Wu Jianjun, Yan Feng, et al. Assessment of regional soil moisture status based on characteristics of surface temperature vegetation index space[J]. Acta Ecologica Sinica, 2009, 29(3): 1205-1216.

[17] Liu Jing. Vegetation Cover Change Process and Future Distribution Prediction on the Loess Plateau after Grain to Green[D]. Beijing: University of Chinese Academy of Science, 2019.

[18] Chen Mo, Lu Wenxi, Hou Zeyu, et al. The assessment of groundwater quality based on support vector machine in Western Jilin[J]. Water Saving Irrigation, 2013, 38(5): 29-33.

[19] Mao Huihui, Yan Yaoxing, Zhang Jie. The present situation and prospect of the hydrologic forecasting methods[J]. Journal of Library and Information Science, 2005, 15(19): 172-173.

[20] Liang Hao, Huang Shengzhi, Meng Erhao, et al. Runoff prediction based on multiple hybrid models[J]. Journal of Hydraulic Engineering, 2020, 51(1): 112-125.

[21] Dai Meng, Huang Shengzhi, Huang Qiang, et al. Dynamic assessments of drought multi-attribute risks and analysis of its driving force[J]. Journal of Hydroelectric Engineering, 2019, 38(8): 15-26.

[22] Wang Xiuying, Liao Liufeng, Wang Junjie. A forecast model for flash heavy rainfall in southwestern Yunnan province based on a multiple linear regression method[J]. Journal of Meteorology and Environment, 2019, 35(2): 15-22.

[23] Meng Erhao, Huang Shengzhi, Huang Qiang, et al. Runoff prediction incorporating anomalous atmospheric circulation factors[J]. Journal of Hydroelectric Engineering, 2017, 36(8): 34-42.

[24] Qian Jianping. A Study on Runoff Prediction in Qira River Basin Based on Artificial Neural Network[D]. Urumqi: Xinjiang University, 2018.

[25] Zhang Jun. Mid-Long Term Hydrological Forecasting and Operation Techniques Research and Application[D]. Dalian: Dalian University of Technology, 2009.

[26] Yang Hanbo, Lyu Huafang, Hu Qingfang, et al. Comparison of parameterization methods for calculating the downward long wave radiation over the North China Plain[J]. Journal of Tsinghua University (Science and Technology), 2014, 54(5): 590-595.

[27] Zhang Xia, Li Zhanbin, Zhang Zhenwen. Application and comparison of two prediction models for groundwater dynamics[J]. Acta Ecologica Sinica, 2012, 32(21): 6788-6794.

[28] Xu Dongmei, Zhao Xiaoshen. Review on study of mid and long term hydrological forecasting technique[J]. Water Conservancy Science and Technology and Economy, 2010, 16(1): 1-7.

[29] Anderson L O, Malhi Y, Aragão Luiz E O C, et al. Remote sensing detection of droughts in Amazonian forest canopies[J]. New Phytologist, 2010, 187(3): 733-750.

[30] Zhou Zhaoqiang, Ding Yibo, Shi Haiyun, et al. Analysis and prediction of vegetation dynamic changes in China: Past, present and future[J]. Ecological Indicators, 2020, 117: 106642.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.