# Postprint: Simulation of Spatiotemporal Patterns and Future Evolution Trends of Grassland Net Primary Productivity in the Loess Plateau Based on Random Forest Models

**Authors:** Liu Huanhuan, Chen Yin, Liu Yue, Gang Chengcheng, Gang Chengcheng

**Date:** 2023-03-13T00:00:00+00:00

## Abstract

Accurate estimation of grassland Net Primary Productivity (NPP) is crucial for understanding carbon cycling processes in grassland ecosystems and scientifically assessing the response and adaptation mechanisms of grasslands to climate change. Taking the grassland ecosystem of the Loess Plateau as the study object, and based on 1788 grassland biomass data points and 19 environmental factor datasets (including climate, vegetation, soil, and topographic factors), this study simulated the spatiotemporal dynamics of grassland NPP on the Loess Plateau from 2002 to 2020 using a random forest model, and estimated the future evolution trend of grassland NPP on the Loess Plateau under four future climate scenarios of Shared Socioeconomic Pathways (SSPs). The results indicate: (1) The random forest model demonstrates good simulation accuracy and can be used for estimating grassland NPP on the Loess Plateau; (2) Grassland NPP on the Loess Plateau exhibits an overall spatial distribution pattern of "high in the southeast and low in the northwest," with an annual mean value of 276.55 g C·m-2, with the highest grassland NPP occurring in the Guanzhong region of Shaanxi Province; (3) From 2002 to 2020, grassland NPP on the Loess Plateau showed an overall increasing trend, with 55.01% of the region experiencing increased grassland NPP, mainly concentrated in the Guanzhong region of Shaanxi Province, western Gansu Province, and northern Shanxi Province; (4) Against the background of climate warming and humidification, grassland NPP on the Loess Plateau will show an increasing trend by the end of this century, with the greatest increase under the SSP585 scenario and the smallest increase under the SSP126 scenario. The use of the random forest model can effectively simulate the spatiotemporal patterns and future evolution trends of grassland NPP on the Loess Plateau, providing data support for the conservation and

sustainable development of grassland ecosystems on the Loess Plateau.

## Full Text

## Simulation of Spatial Pattern and Future Trends of Grassland Net Primary Productivity in the Loess Plateau Based on Random Forest Model

**LIU Huanhuan[1], CHEN Yin[1], LIU Yue[1], GANG Chengcheng[2,3]**
[1]College of Grassland Agriculture, Northwest A&F University, Yangling 712100, Shaanxi, China [2]Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, Shaanxi, China [3]Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, Shaanxi, China

### Abstract

Accurate estimation of grassland net primary productivity (NPP) is crucial for understanding carbon cycling processes in grassland ecosystems and scientifically assessing their response and adaptation mechanisms to climate change. Focusing on the grassland ecosystem of the Loess Plateau, this study simulated the spatiotemporal dynamics of grassland NPP from 2002 to 2020 using a random forest model based on 1,788 grassland biomass data points and 19 environmental factors (including climate, vegetation, soil, and topographic factors). Future trends of grassland NPP were estimated under four shared socioeconomic pathway (SSP) climate scenarios for the years 2030s, 2050s, 2070s, and 2090s. The results demonstrated that: (1) The random forest model exhibited good simulation accuracy and can be used to estimate grassland NPP in the Loess Plateau; (2) Grassland NPP showed a spatial pattern of "high in the southeast and low in the northwest," with a mean annual value of 276.55 g $C \cdot m^{-2}$. The highest NPP values were observed in the Guanzhong region of Shaanxi Province; (3) During 2002–2020, grassland NPP in the Loess Plateau showed an overall increasing trend, with 55.01% of the region exhibiting significant increases, concentrated mainly in Guanzhong region, western Gansu, and northern Shanxi; (4) Under a warming and wetting climate, grassland NPP will continue to increase by the end of this century across all scenarios, with the greatest increase under SSP585 and the smallest under SSP126. The random forest model can effectively simulate the spatiotemporal patterns and future trends of grassland NPP in the Loess Plateau, providing data support for grassland ecosystem conservation and sustainable development.

**Keywords:** grassland ecosystems; net primary productivity (NPP); random forest; Shared Socioeconomic Pathway (SSPs); Loess Plateau

## Introduction

Grassland ecosystems constitute one of the primary components of terrestrial ecosystems, covering approximately 40% of the global land area. They play vital roles in maintaining ecosystem functions and supporting socioeconomic services, including biodiversity conservation, soil and water retention, and greenhouse gas mitigation. Net primary productivity (NPP) is a key indicator for measuring ecosystem carbon sequestration capacity and directly reflects grassland productivity while serving as a critical metric for evaluating carbon budgets and regulating ecological processes in grassland ecosystems. Scientific assessment of grassland NPP and its spatiotemporal dynamics is therefore essential for grassland ecosystem management and resource conservation.

Traditional field sampling methods, while accurate, require substantial human, material, and financial resources, limiting their application for large-scale NPP estimation. The development of remote sensing technology and ecological models provides important alternative approaches. Remote sensing techniques estimate grassland NPP by leveraging relationships between remote sensing indices and grassland biomass. Model simulation represents another common method for large-scale NPP estimation, including process-based models such as BIOME-BGC and empirical models like Miami. In recent years, machine learning methods have been widely applied in vegetation productivity estimation. For instance, studies have used model tree ensemble algorithms to simulate global gross primary productivity and high-resolution total primary productivity. Random Forest (RF), an ensemble machine learning algorithm based on decision trees, has been successfully applied to estimate NPP across various grassland types globally, in western China, and on the Tibetan Plateau.

The Loess Plateau represents the world' s largest loess-covered region, characterized by arid climate, severe soil erosion, and extremely fragile ecological conditions. With the implementation of ecological restoration programs such as the Grain-for-Green Project, vegetation changes and their environmental effects have attracted considerable attention. Previous studies have investigated NPP spatiotemporal dynamics and driving factors in the Loess Plateau. However, due to complex terrain, significant discrepancies exist among different research results, and future NPP trends under global warming remain unclear. This study addresses this gap by simulating spatiotemporal NPP dynamics from 2002 to 2020 using RF modeling based on climate, vegetation, soil, and topographic factors, and projecting future trends under four SSP scenarios to provide scientific support for grassland ecosystem management and sustainable development.

## 1 Materials and Methods

### 1.1 Study Area

The Loess Plateau is located in the middle reaches of the Yellow River basin in China, spanning 100°54 -114°33 E and 33°43 -41°16 N, with a total area of 63$\times$10^{4}$ km$^2$. It extends across seven provinces including Shaanxi, Gansu, Ningxia, Shanxi, Qinghai, Henan, and Inner Mongolia, representing one of China' s four major plateaus and the world' s largest loess-covered region. The plateau encompasses arid, semi-arid, and semi-humid climate zones, with annual precipitation ranging from 150–800 mm and elevation varying from 200–3000 m, creating distinct regional differences. The dominant vegetation types include forest, grassland, and shrubland, with grassland covering approximately 40% of the total area.

### 1.2 Data Acquisition and Processing

Grassland aboveground biomass data were obtained from literature, the National Earth System Science Data Center, and the ORNL DAAC (https://daac.ornl.gov/), comprising 1,788 samples. Belowground biomass was estimated using root-to-shoot ratios, and grassland NPP was calculated by multiplying biomass by conversion coefficients (g $C \cdot m^{-2}$) to serve as training data. Nineteen feature parameters were selected across four categories: soil physicochemical factors (S), topographic factors (T), meteorological factors (A), and vegetation factors (B). Meteorological data (temperature and precipitation) were sourced from the Loess Plateau Sub-center of the National Earth System Science Data Center (http://loess.geodata.cn). Annual mean temperature, growing season mean temperature (TEM4-10), annual minimum temperature, annual precipitation (PRE), and growing season precipitation (PRE4-10) were synthesized from monthly data. Topographic data including elevation and slope were derived from the SRTM digital elevation model (http://srtm.csi.cgiar.org/) and calculated using ArcGIS. Soil physicochemical properties were obtained from the World Soil Database (https://data.isric.org/). Vegetation data included solar-induced chlorophyll fluorescence (SIF) from GOSIFv2 (https://globalecology.unh.edu), fraction of absorbed photosynthetically active radiation (FAPAR), normalized difference vegetation index (NDVI), and evapotranspiration (ET) from MODIS products (https://modis.gsfc.nasa.gov). Future precipitation and temperature data were obtained from WorldClim (https://www.worldclim.org/) under four SSP scenarios (SSP126, SSP245, SSP370, SSP585) from five climate system models (ACCESS-ESM1-5, CanESM5, CNRM-CM6-1, etc.) for the 2030s, 2050s, 2070s, and 2090s. The Loess Plateau boundary and grassland distribution data were sourced from the Chinese Academy of Sciences Resource and Environment Data Center (https://www.resdc.cn). All environmental factor data were resampled to 250–1000 m resolution and projected to D_{WGS}_{1984} using ArcGIS. Other environmental factors remained unchanged for future predictions.

summarizes the environmental factors used in the random forest model.

### 1.3 Methods

**1.3.1 Modeling Approach** This study employed the Random Forest (RF) algorithm, an ensemble method developed from Classification and Regression Trees (CART). RF combines multiple decision trees and averages their results to reduce generalization error and improve prediction accuracy. The algorithm is widely applicable for prediction problems, requires minimal parameter tuning, and can handle high-dimensional feature spaces effectively. Model performance was evaluated using the coefficient of determination ($R^2$) and root mean square error (RMSE). All modeling, parameter optimization, predictions, and spatial visualizations were conducted in R 4.0.5.

Theil-Sen median trend analysis, a robust non-parametric method, was used to calculate temporal trends. The slope (Senslope) is calculated as the median of all pairwise slopes between data points. A positive Senslope indicates an increasing trend, while a negative value indicates a decreasing trend. The Mann-Kendall test, another non-parametric statistical test, was used to assess trend significance. At a confidence level of $\alpha = 0.05$, a p-value less than 0.05 indicates a significant trend.

## 2 Results

### 2.1 Correlation Analysis Between NPP and Environmental Factors

Among vegetation factors, SIF showed the strongest correlation with NPP ($r = 0.91$), followed by FAPAR ($r = 0.89$) and NDVI ($r = 0.83$). In meteorological factors, growing season precipitation (PRE4-10) exhibited the highest correlation ($r = 0.71$), while growing season temperature (TEM4-10) also showed strong correlation ($r = 0.68$). Soil factors showed weak overall correlations, with sand content in the 0–30 cm layer (SAND_S) showing the strongest correlation ($r = -0.27$). In the random forest model factor importance analysis [Figure 2: see original paper], vegetation factors (SIF, FAPAR, NDVI) and meteorological factors (PRE4-10, TEM4-10) were the most influential for NPP estimation.

### 2.2 Model Accuracy Assessment

The RF model demonstrated strong performance with training set $R^2$ of 0.97 and RMSE of 27.34 g $\mathrm{C \cdot m^{-2}}$. After 10-fold cross-validation, the test set $R^2$ was 0.84 with RMSE of 59.73 g $\mathrm{C \cdot m^{-2}}$ [Figure 3: see original paper], indicating high prediction accuracy suitable for simulating grassland NPP in the Loess Plateau.

### 2.3 Spatial Dynamics of Grassland NPP

Grassland NPP exhibited strong spatial heterogeneity, decreasing gradually from south to north across the Loess Plateau [Figure 4: see original paper]. The mean annual NPP was 276.55 g $\mathrm{C \cdot m^{-2}}$, with the lowest values (76.52 g $\mathrm{C \cdot m^{-2}}$) in northern Shaanxi, Ningxia, and areas near Lanzhou, Gansu. The highest values (634.86 g $\mathrm{C \cdot m^{-2}}$) occurred in southern regions, particularly the Guanzhong Plain of Shaanxi. During 2002–2020, grassland NPP showed an overall increasing trend [Figure 5: see original paper], with 55.01% of the region experiencing increases. Specifically, 16.59% showed significant increases, 38.52% slight increases, 1.89% slight decreases, and 7.43% significant decreases, primarily in northwestern areas.

### 2.4 Future Grassland NPP Under Different Climate Scenarios

By the end of this century, grassland NPP is projected to increase under all four SSP scenarios [Figure 6: see original paper]. The smallest increase occurs under SSP126 (1.20 g $C \cdot m^{-2}$), while the largest occurs under SSP585 (8.10 g $C \cdot m^{-2}$). The increasing trends are 3.90, 4.00, and 4.00 g $C \cdot m^{-2}$ for SSP245, SSP370, and SSP585 scenarios, respectively. Increased NPP is concentrated in southern Loess Plateau, particularly under SSP585, while decreasing trends are limited to western regions, with the largest declines under SSP126.

## 3 Discussion

### 3.1 Applicability of Random Forest for Estimating Grassland NPP

Compared with traditional NPP estimation models, RF offers superior capability in capturing non-linear complex relationships between NPP and environmental factors while maintaining robust resistance to overfitting. The simulated mean NPP of 276.55 g $C \cdot m^{-2}$ in this study is higher than some previous estimates (e.g., 202.93 g $C \cdot m^{-2}$ from MOD17A3 data), likely due to differences in grassland distribution datasets and the spatial distribution of training samples, which were concentrated in western Loess Plateau. Nevertheless, the results align closely with high-value regions from other studies, demonstrating RF' s effectiveness for regional NPP simulation.

### 3.2 Dynamic Changes of Grassland NPP in the Loess Plateau

The increasing NPP trend during 2002-2020 corroborates previous findings. Precipitation is the primary limiting factor for vegetation growth in this arid region, and the observed increase in rainfall has created favorable conditions for NPP enhancement. The Grain-for-Green Program, implemented since 1999 through measures such as enclosure and grazing prohibition, has significantly improved grassland ecosystems, contributing to NPP increases. Projected future increases in precipitation and minimum temperature under all SSP scenarios will sustain NPP growth, particularly under the wettest SSP585 scenario.

### 3.3 Uncertainty Analysis of RF Estimation Results

Despite high model accuracy, several uncertainties remain. First, RF is a data-driven approach sensitive to training data quantity and quality. The uneven spatial distribution of field measurements, with more samples in northwestern and southwestern regions, may affect simulation precision elsewhere. Second, as a black-box model, RF lacks explicit control over internal mechanisms, requiring extensive parameter tuning. Future work should incorporate more comprehensive field data and integrate anthropogenic activity indicators to improve estimation accuracy.

## 4 Conclusions

This study simulated spatiotemporal dynamics of grassland NPP in the Loess Plateau using random forest modeling based on vegetation, climate, soil, and topographic factors, yielding four key conclusions:

1) Grassland NPP averaged 276.55 g $C \cdot m^{-2}$ during 2002–2020, exhibiting a "high in southeast, low in northwest" spatial pattern with an overall increasing trend, driven by increased precipitation and ecological restoration programs.

2) By the end of this century, grassland NPP will continue increasing under warming and wetting conditions across all SSP scenarios, with the greatest increase under SSP585.

3) The RF model effectively simulated grassland NPP spatiotemporal patterns, though uncertainties remain due to data limitations and model characteristics.

4) Future models should incorporate anthropogenic activity factors alongside environmental variables to further improve estimation accuracy for grassland vegetation productivity.

## References

[7] Guo D, Song X, Hu R, et al. Grassland type dependent spatiotemporal characteristics of productivity in Inner Mongolia and its response to climate factors[J]. Science of the Total Environment, 2021, 775: 145617.

[10] Jung M, Reichstein M, Margolis H A, et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations[J]. Journal of Geophysical Research: Biogeosciences, 2011, 116(G3): 1566-1582.

[11] Yao Y, Wang X, Li Y, et al. Spatiotemporal pattern of gross primary productivity and its covariation with climate in China over the last thirty years[J]. Global Change Biology, 2018, 24(1): 184-196.

[12] Sun Y, Feng Y, Wang Y, et al. Field-based estimation of net primary productivity and its above- and belowground partitioning in global grassland[J]. Journal of Geophysical Research: Biogeosciences, 2021, 126(11): 6472-6496.

[13] Li C, Dou T, Wang Y, et al. A method for quantifying the impacts of human activities on net primary production of grasslands in Northwest China[J]. Remote Sensing, 2021, 13(13): 2479-2497.

[15] Fu B. Soil erosion and its control in the loess plateau of China[J]. Soil Use and Management, 1989, 5(2): 76-82.

[16] Kang L, Han X, Zhang Z, et al. Grassland ecosystems in China: Review of current knowledge and research advancement[J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2007, 362(1482): 997-1008.

[17] Liu F, Yan H, Gu F, et al. Net primary productivity increased on the loess plateau following implementation of the grain-to-green program[J]. Journal of Resources and Ecology, 2017, 8(4): 413-422.

[20] Zhao G, Mu X, Wen Z, et al. Soil erosion, conservation, and eco-environment changes in the Loess Plateau of China[J]. Land Degradation & Development, 2013, 24(5): 499-510.

[21] Wang Y, Shao M, Zhu Y, et al. Impacts of land use and plant characteristics on dried soil layers in different climatic regions on the Loess Plateau of China[J]. Agricultural and Forest Meteorology, 2011, 151(4): 437-448.

[24] Peng S, Ding Y, Liu W, et al. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017[J]. Earth System Science Data, 2019, 11(4): 1931-1946.

[25] Li X, Xiao J. A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data[J]. Remote Sensing, 2019, 11(5): 517-526.

[26] Su Y, Li J, Zhi H, et al. Projected drought conditions in Northwest China with CMIP6 models under combined SSPs and RCPs for 2015-2099[J]. Advances in Climate Change Research, 2020, 11(3): 210-217.

[27] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[29] Biau G, Scornet E. A random forest guided tour[J]. Test, 2016, 25(2): 197-227.

[30] Belgiu M, Dragut L. Random forest in remote sensing: A review of applications and future directions[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31.

[31] Ohlson J, Kim S. Linear valuation without OLS: The Theil-Sen estimation approach[J]. Review of Accounting Studies, 2015, 20(1): 395-435.

[32] Ullah S, You Q, Ali A, et al. Observed changes in maximum and minimum temperatures over China-Pakistan economic corridor during 1980-2016[J]. Atmospheric Research, 2019, 216: 37-51.

[33] Zeng N, Ren X, He H, et al. Estimating grassland aboveground biomass on the Tibetan Plateau using a random forest algorithm[J]. Ecological Indicators, 2019, 102: 479-487.

[35] Li C, Wang Y, Wu X, et al. Reducing human activity promotes environmental restoration in arid and semi-arid regions: A case study in Northwest

China[J]. Science of the Total Environment, 2021, 768: 144525.

[36] Li C, Dou T, Wang Y, et al. A method for quantifying the impacts of human activities on net primary production of grasslands in Northwest China[J]. Remote Sensing, 2021, 13(13): 2479-2497.

[37] Liu G, Shao Q, Fan J, et al. Change trend and restoration potential of vegetation net primary productivity in China over the past 20 years[J]. Remote Sensing, 2022, 14(7): 1634-1660.

[40] Gang C, Zhao W, Zhao T, et al. The impacts of land conversion and management measures on the grassland net primary productivity over the Loess Plateau, Northern China[J]. Science of the Total Environment, 2018, 645: 827-836.

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

[42] Gang C, Gao X, Peng S, et al. Satellite observations of the recovery of forests and grasslands in Western China[J]. Journal of Geophysical Research: Biogeosciences, 2019, 124(7): 1905-1922.

[43] Yu H, Wu Y, Niu L, et al. A method to avoid spatial overfitting in estimation of grassland aboveground biomass on the Tibetan Plateau[J]. Ecological Indicators, 2021, 125: 107450.

[44] Karlson M, Ostwald M, Reese H, et al. Mapping tree canopy cover and aboveground biomass in sudano-sahelian woodlands using Landsat 8 and random forest[J]. Remote Sensing, 2015, 7(8): 10017-10041.