ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202510.00097

Driving Mechanisms and Nonlinear Threshold Identification of Vegetation in China: Based on Causal Inference and Machine Learning (Post-print)

Authors: ZHANG Houtian, Shidong Wang, DING Junjie, WANG Shidong

Date: 2025-10-22T00:00:00+00:00

Abstract

Climate change significantly affects vegetation dynamics. Thus, understanding interactions between vegetation and climatic factors is essential for ecological management. This study utilized kernel Normalized Difference Vegetation Index (kNDVI) and climatic data (temperature, precipitation, humidity, and vapor pressure deficit (VPD)) for China from 2000 to 2022, integrating Geographic Convergent Cross Mapping (GCCM) causal modeling, Extreme Gradient Boosting-Shapley Additive Explanations (XGBoost-SHAP) nonlinear threshold identification, and Geographical Simulation and Optimization Systems-Future Land Use Simulation (GeoSOS-FLUS) spatial prediction modeling to investigate vegetation spatio-temporal characteristics, driving mechanisms, nonlinear thresholds, and future spatial patterns. Results indicated that from 2000 to 2022, China's kNDVI showed an overall increasing trend (annual average ranging from 0.29 to 0.33) with distinct spatial differentiation: 52.77% of areas (located in agricultural and ecological restoration regions in the central-eastern plain) experienced vegetation improvement, whereas 2.68% of areas (located in the southeastern coastal urbanized regions and the Yangtze River Delta) experienced vegetation degradation. The coefficient of variation (CV) of kNDVI at 0.30–0.40 (accounting for 10.61%) was significantly higher than that of NDVI (accounting for 1.80%). Climate-driven mechanisms exhibited notable library length (L) dependence. At short-term scales (L<50), vegetation-driven transpiration regulated local microclimate, with a causal strength from kNDVI to temperature of 0.04–0.15; at long-term scales (L>100), cumulative temperature effects dominated vegetation dynamics, with a causal strength from temperature to kNDVI of 0.33. Humidity and kNDVI formed bidirectional positive feedback at long-term scales (L=210, causal strength>0.70), whereas the long-term suppressive effect of VPD was particularly pronounced (causal strength=0.21)

in arid areas. The optimal threshold intervals identified were temperature at –12.18°C–0.67°C, precipitation at 24.00–159.74 mm, humidity lower than 22.00%, and VPD of <0.07, 0.17–0.24, and >0.30 kPa; notably, the lower precipitation threshold (24.00 mm) represented the minimum water requirements for vegetation recovery in arid areas. Future kNDVI spatial patterns are projected to continue the trend of "southeastern optimization and northwestern delay" from 2025 to 2040: the area proportion of high kNDVI values (>0.50) will rise from 40.43% to 41.85%, concentrated in the Sichuan Basin and the southern hills; meanwhile, the proportion of low-value areas of kNDVI (0.00–0.10) in the arid northwestern areas will decline by only 1.25%, constrained by sustained temperature and VPD stress. This study provides a scientific basis for vegetation dynamic regulation and sustainable development under climate change.

Full Text

Preamble

Journal of Arid Land (2025) 17(10): 1341–1360
doi: 10.1007/s40333-025-0110-y; CSTR: 32276.14.JAL.0250110y
Science Press & Springer-Verlag

Driving Mechanism and Nonlinear Threshold Identification of Vegetation in China: Based on Causal Inference and Machine Learning

ZHANG Houtian[1], WANG Shidong[2]*, DING Junjie[3]
[1] School of Intelligent Construction and Civil Engineering, Luoyang Institute of Science and Technology, Luoyang 471023, China
[2] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China
[3] Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450003, China

---

Abstract

Climate change significantly affects vegetation dynamics, making it essential to understand interactions between vegetation and climatic factors for effective ecological management. This study utilized kernel Normalized Difference Vegetation Index (kNDVI) and climatic data (temperature, precipitation, humidity, and vapor pressure deficit (VPD)) for China from 2000 to 2022, integrating Geographic Convergent Cross Mapping (GCCM) causal modeling, Extreme Gradient Boosting-Shapley Additive Explanations (XGBoost-SHAP) nonlinear threshold identification, and Geographical Simulation and Optimization Systems-Future Land Use Simulation (GeoSOS-FLUS) spatial prediction modeling to investigate vegetation spatiotemporal characteristics, driving mechanisms, nonlinear thresholds, and future spatial patterns. Results indicated that from 2000 to 2022, China's kNDVI showed an overall increasing trend (annual average ranging from 0.29 to 0.33) with distinct spatial differentiation:

52.77% of areas (located in agricultural and ecological restoration regions in the central-eastern plain) experienced vegetation improvement, whereas 2.68% of areas (located in the southeastern coastal urbanized regions and the Yangtze River Delta) experienced vegetation degradation. The coefficient of variation (CV) of kNDVI at 0.30–0.40 (accounting for 10.61%) was significantly higher than that of NDVI (accounting for 1.80%).

Climate-driven mechanisms exhibited notable library length (L) dependence. At short-term scales ($L < 50$), vegetation-driven transpiration regulated local microclimate, with a causal strength from kNDVI to temperature of 0.04–0.15; at long-term scales ($L > 100$), cumulative temperature effects dominated vegetation dynamics, with a causal strength from temperature to kNDVI of 0.33. Humidity and kNDVI formed bidirectional positive feedback at long-term scales ($L = 210$, causal strength > 0.70), whereas the long-term suppressive effect of VPD was particularly pronounced (causal strength = 0.21) in arid areas.

The optimal threshold intervals identified were temperature at −12.18°C–0.67°C, precipitation at 24.00–159.74 mm, humidity lower than 22.00%, and VPD of < 0.07, 0.17–0.24, and > 0.30 kPa; notably, the lower precipitation threshold (24.00 mm) represented the minimum water requirements for vegetation recovery in arid areas. Future kNDVI spatial patterns are projected to continue the trend of "southeastern optimization and northwestern delay" from 2025 to 2040: the area proportion of high kNDVI values (> 0.50) will rise from 40.43% to 41.85%, concentrated in the Sichuan Basin and the southern hills; meanwhile, the proportion of low-value areas of kNDVI (0.00–0.10) in the arid northwestern areas will decline by only 1.25%, constrained by sustained temperature and VPD stress. This study provides a scientific basis for vegetation dynamic regulation and sustainable development under climate change.

Keywords: kernel Normalized Difference Vegetation Index (kNDVI); climate drivers; machine learning; Geographic Convergent Cross Mapping (GCCM); Extreme Gradient Boosting-Shapley Additive Explanations (XGBoost-SHAP); Geographical Simulation and Optimization Systems-Future Land Use Simulation (GeoSOS-FLUS) model

1 Introduction

Vegetation indices are vital tools for assessing ecosystem dynamics, with the traditional Normalized Difference Vegetation Index (NDVI) being the most widely applied [**?**, **?**]. However, NDVI suffers from a saturation effect under high vegetation biomass conditions, reducing its sensitivity to dense vegetation and leading to the underestimation of vegetation cover and productivity. To address this issue, researchers have proposed several improved indices, such as the Enhanced Vegetation Index (EVI), yet the saturation problem remains unresolved [**?**]. Recently, the kernel Normalized Difference Vegetation Index (kNDVI) has

emerged, recalibrating the relationship between red and near-infrared reflectance using nonlinear methods, substantially mitigating the saturation limitations of NDVI [?]. Empirical studies have shown that kNDVI outperforms other indices in regions with dense vegetation, demonstrates stronger correlations with indicators such as primary vegetation productivity, and exhibits greater robustness to noise and saturation. Therefore, kNDVI provides a more accurate and reliable alternative for vegetation monitoring, particularly in areas with extremely dense vegetation.

Although kNDVI effectively addresses the saturation problem of traditional NDVI, existing studies have mainly explored its spatiotemporal variation characteristics and its relationship with climatic and environmental factors, for example, through trend analysis and lag correlation to reveal the influence of climate variables on kNDVI [?]. However, most current analyses remain at the level of correlation, lacking an in-depth understanding of the causal direction and strength between vegetation and meteorological factors [?]. Thus, it is urgent to introduce new approaches to clarify the causal direction between vegetation and climatic elements [?, ?]. In recent years, commonly used causal inference models in ecology mainly include Granger causality tests, Bayesian networks, and structural equation models. Granger causality test infers causal direction through lag effects in time series but assumes linearity and struggles to address spatial dependencies [?, ?]. Bayesian networks use probabilistic graphical models to infer conditional dependencies among variables [?] but face high computational complexity in high-dimensional data and often overlook spatial heterogeneity [?]. Structural equation models quantify direct and indirect effects among multiple variables through path analysis but rely on prior assumptions about path structure and cannot fully capture dynamic spatiotemporal interactions [?]. To overcome these limitations, the Geographic Convergent Cross Mapping (GCCM) approach innovatively integrates spatial lag covariates and Dynamic Time Warping (DTW) algorithms. GCCM incorporates spatial autocorrelation and temporal lag effects, employs Convergent Cross Mapping (CCM) to validate the robustness of causal relationships, and avoids spurious correlations [?]. It also supports the quantification of nonlinear causal directions, making it particularly suitable for threshold response analysis in ecosystems [?]. These features make GCCM a powerful tool for uncovering the complex causal networks between vegetation and climate.

In recent years, machine learning models have demonstrated unique value, yet the limitations of different methods still need to be addressed. For example, Random Forest identifies key driving factors through feature importance ranking but cannot quantify the direction of influence and is sensitive to high-dimensional sparse data [?]. Support Vector Machines use kernel functions to handle nonlinear relationships but have poor interpretability, making it difficult to reveal interaction mechanisms [?]. Deep learning captures complex nonlinear patterns through neural networks but its “black box” nature complicates ecological interpretation and requires massive training data [?, ?]. In this regard, the integrated application of the Extreme Gradient Boosting-Shapley Additive Explanations

(XGBoost-SHAP) framework offers an innovative solution. Extreme Gradient Boosting (XGBoost) improves the modeling efficiency of high-dimensional climate data through regularization strategies and parallel computing [**?**]. Shapley Additive Explanations (SHAP) values use game theory to fairly allocate feature contributions and quantify both the nonlinear direction and threshold mutation points of individual factors [**?**]. Interaction SHAP value visualization reveals synergistic and antagonistic effects between factors, breaking the interpretability bottleneck of traditional machine learning models.

Future predictions of vegetation dynamics require coupling of natural processes with human disturbances. The Geographical Simulation and Optimization Systems-Future Land Use Simulation (GeoSOS-FLUS) model has shown significant advantages in this field. By integrating system dynamics, cellular automata, and multi-objective optimization algorithms, the model accurately simulates the spatial game process of land use transformation [**?**]. Its adaptive inertia competition mechanism effectively addresses the sensitivity to initial conditions of traditional models. By coupling climate scenarios and socio-economic pathways, GeoSOS-FLUS model enables multi-scale ecological-economic coevolution simulations and uses Markov Chain Monte Carlo (MCMC) parameter optimization to reduce model uncertainty [**?**]. Compared with traditional models such as Conversion of Land Use and Its Effects at Small Regional Extent (CLUE-S) and Cellular Automata-Markov (CA-Markov), it improves the Kappa coefficient by 0.15–0.28, making it especially suitable for vegetation cover scenario prediction in rapidly urbanizing regions of China and providing dynamic decision support for multi-objective spatial planning [**?**, **?**].

Despite the growing number of studies on kNDVI and its relationship with climate, three key gaps persist: (1) most researches focus on trend or correlation analyses, without explicitly identifying bidirectional causal relationships between kNDVI and climatic factors; (2) existing threshold identification methods usually adopt single-factor or linear assumptions, lacking deeper exploration of nonlinear threshold responses and interactions among multiple climate drivers; and (3) scenario projections often remain disconnected from causal mechanism analyses, failing to establish a comprehensive and integrated analytical framework from driver diagnosis to spatial prediction.

To address these gaps, this study developed an innovative analytical framework that integrated causal inference, machine learning-based threshold detection, and scenario-based spatial prediction, using kNDVI and climate data in China from 2000 to 2022. Specifically, the aims of the study are to: (1) apply GCCM to explicitly quantify the bidirectional causal relationships between vegetation dynamics (kNDVI) and key climatic factors, clarifying driving directions and intensities; (2) employ XGBoost-SHAP approach to uncover nonlinear threshold response curves and critical tipping points of multiple climatic drivers, overcoming the limitations of conventional methods; and (3) integrate the above causal and threshold analyses into GeoSOS-FLUS model to spatially project kNDVI dynamics for the period from 2025 to 2040. This comprehensive framework

provides a robust scientific basis for vegetation monitoring, environmental management, and adaptive ecological planning, thereby contributing to sustainable regional ecological development.

2.1 Study Area

China is located at the junction of Asia and Europe, covering a land area of 9.6×10^6 km^2 and comprising 34 provincial-level administrative regions (Fig. 1). Most of southeastern China has semi-humid to humid climates, while the northwestern region is predominantly arid and semi-arid. The eastern region experiences monsoonal climates, the northwestern region has a temperate continental climate, and the southwestern region also shows temperate continental characteristics. Parts of the Qinghai-Xizang Plateau are classified as high mountain environments. A clear climatic gradient extends from humid climate in the east to arid climate in the west. Overall, China's climate forms a three-step topography characterized by a west-to-east gradient in increasing precipitation and temperature [**?**].

Fig. 1 Demarcation of seven geographic regions in China. Note that the map is based on the standard map (GS(2022)0438) of the Map Service System (http://bzdt.ch.mnr.gov.cn/) marked by the Ministry of Natural Resources of the People's Republic of China, and the boundary of base map has not been modified. The demarcation line of the seven geographic regions was referred to Tong et al. (2022).

2.2.1 NDVI

The NDVI data used in this study were obtained from the "China Annual Vegetation Index Spatial Distribution Dataset" (http://www.resdc.cn). This dataset is derived from 10-d, 1-km vegetation index data acquired by the SPOT-VEGETATION and PROBA-V sensors, and it employs the Maximum Value Composite (MVC) method to produce a relatively clean spatiotemporal distribution of vegetation index. To address NDVI saturation in densely vegetated areas, we applied the kernel method directly to this original NDVI dataset to derive the corresponding kNDVI values. Both the original NDVI and the derived kNDVI underwent preprocessing steps—including interpolation, clipping, and projection transformation—to ensure consistency in spatial extent, resolution, and coordinate reference systems, thus providing a robust foundation for subsequent analyses of vegetation dynamics and their climatic drivers.

2.2.2 Meteorological Data

The monthly climatic data used in this study included near-surface temperature, precipitation, humidity, and VPD, sourced from the TerraClimate dataset (https://www.climatologylab.org) covering the period from 2000 to 2022 [**?**]. TerraClimate integrates multi-source observations and reanalysis data, offering wide spatial coverage and high spatiotemporal resolution, and can effectively

reflect regional-scale climate change features. To ensure consistency with other data in the study area, we first clipped the original data using the international administrative boundary of China. The clipped data were then processed through resampling and projection transformation, using bilinear interpolation for resampling to preserve relative data fidelity and applying projection transformation to unify the coordinate system. This resulted in a climatic dataset consistent with kNDVI data in both spatial resolution and projection coordinates, providing reliable input for subsequent spatial trend analysis, partial correlation analysis, and future scenario simulation using GeoSOS-FLUS model.

2.3.1 Construction of kNDVI Dataset

In this study, the calculation formula for NDVI is as follows [**?**]:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}, \tag{1}$$

where NIR is the reflectance in the near-infrared band; and RED is the reflectance in the visible red band. The calculation formula of kNDVI derived from NDVI is as follows [**?**]:

$$\mathrm{kNDVI} = \tanh(\mathrm{NDVI}^2). \tag{2}$$

2.3.2 Univariate Linear Regression Trend Analysis

This study employed a univariate linear trend regression analysis model to quantitatively analyze the changing trend of kNDVI in China from 2000 to 2022. The calculation formula is as follows [**?**]:

$$\mathrm{Slope} = \frac{n \times \sum_{i=1}^{n} i \times N_i - \sum_{i=1}^{n} i \sum_{i=1}^{n} N_i}{n \times \sum_{i=1}^{n} i^2 - (\sum_{i=1}^{n} i)^2}, \tag{3}$$

where Slope is the annual rate of change in kNDVI for each pixel; n is the total length of the time series ($n = 23$ in this study); and N_i is the kNDVI value recorded for the same pixel in the ith year. A positive slope (slope $>$ 0) indicates an upward vegetation trend, whereas a negative slope (slope $<$ 0) indicates a decline trend. After calculating slope for every pixel, a two-tailed Student's t-test ($\alpha = 0.05$) was applied to gauge statistical significance. The resulting trends were grouped into four classes—significantly decreasing (slope < 0, $P < 0.05$), not significantly decreasing (slope < 0, $P > 0.05$), not significantly increasing (slope > 0, $P > 0.05$), and significantly increasing (slope > 0, $P < 0.05$)—thereby quantifying both the magnitude and confidence level of vegetation change across China from 2000 to 2022.

2.3.3 GCCM

GCCM was employed to quantify bidirectional causality between vegetation dynamics (Y: kNDVI) and each climatic driver (X: temperature, precipitation, humidity, and VPD) by reconstructing state-space manifolds for the paired variables and performing cross mapping under a common dynamical-systems assumption. If X and Y arise from the same underlying system, their shadow manifolds M_X and M_Y are topologically linked, allowing the state of one to be locally inferred from neighborhoods on the other; the strength and direction of causation are then assessed via the convergence of cross-map skill as the library length (L) increases. Let $\psi(x, s)$ denote the delay-embedded state vector of X at sample index s with embedding dimension E and delay τ. On the shadow manifold M_X, identify the $E + 1$ nearest neighbors $\psi(x, s_i)$ of $\psi(x, s)$ (ordered by increasing distance). The corresponding observations $Y(s_i)$ on M_Y are combined with normalized weights to produce a local prediction $\hat{Y}(s)$ of $Y(s)$ [**?**]:

$$\hat{Y}(s) = \sum_{i=1}^{E+1} w_i(s) Y(s_i), \tag{4}$$

where the weights reflect proximity on M_X and are constrained to sum to one [**?**]:

$$\sum_{i=1}^{E+1} w_i(s) = 1. \tag{5}$$

Proximity is converted to weights by an exponential kernel acting on the distance between state vectors; closer neighbors receive larger weights. With $d_i(s)$ a distance function on the embedded states and $d_{\min}(s)$ the distance to the nearest neighbor [**?**, **?**]:

$$w_i(s) = \frac{\exp(-d_i(s)/d_{\min}(s))}{\sum_{j=1}^{E+1} \exp(-d_j(s)/d_{\min}(s))}. \tag{6}$$

This kernel amplifies local information while damping distant states, yielding a stable and locally weighted mapping from M_X to M_Y, thereby exhibiting cross-map skill and convergence.

Causal strength from $X \to Y$ is quantified by the Pearson correlation ρ between predictions $\hat{Y}(s)$ and observations Y, computed over a library of size L and evaluated as L increases [**?**]:

$$\rho_{X \to Y}(L) = \frac{\mathrm{Cov}(\hat{Y}, Y)}{\sqrt{\mathrm{Var}(\hat{Y})\mathrm{Var}(Y)}}. \tag{7}$$

A monotonic or asymptotically increasing ρ indicates that information in X reconstructs Y, consistent with $X \to Y$ causality; the reverse direction $Y \to X$ is assessed analogously. Statistical significance is evaluated with phase-randomized surrogate series (permutations $B = 1000$), yielding empirical two-tailed P-values; combinatorial counts may involve the factorial $n!$ (i.e., $n = 1\times2\times\cdots\times n$), which appears only in the permutation bookkeeping. Convergence is declared when the terminal-segment slope of ρ falls below a small tolerance (e.g., 10^{-3}) and $P < 0.05$. Embedding parameters E and τ were selected by false-nearest-neighbors and average-mutual-information criteria, respectively.

2.3.4 XGBoost-SHAP

To quantify nonlinear responses of vegetation to climate while retaining interpretability, we coupled Extreme Gradient Boosting (XGBoost) with Shapley Additive Explanations (SHAP). XGBoost fits kNDVI with an additive ensemble of regression trees and learns each new tree by minimizing a regularized objective that balances empirical loss and model complexity. After training, SHAP attributes the model output to individual predictors under a cooperative-game framework, allowing us to read threshold-like behavior and interactions directly from attribution curves.

At the mth boosting step, the model adds a regression tree f_m and optimizes the regularized objective [**?**]:

$$\mathcal{L}^{(m)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(m-1)} + f_m(x_i)) + \Omega(f_m) + C, \tag{8}$$

where $\mathcal{L}^{(m)}$ is the total objective at step m; $l(\cdot)$ is the empirical loss (e.g., squared error) between observation y_i and the updated prediction $\hat{y}_i^{(m-1)} + f_m(x_i)$; $f_m(x_i)$ is the output of the current tree; $\Omega(f_m)$ penalizes tree complexity (e.g., node/leaf penalties and leaf-weight shrinkage); C is a constant that does not affect optimization; and n is the number of training samples. Thus, Equation 8 makes each new tree explain the residuals from the previous iteration while controlling overfitting through Ω function. After model fitting, the contribution of feature i to the prediction for an instance x is measured by its SHAP value [**?**]:

$$\phi_i = \sum_{R \subseteq N\,\{i\}} \frac{|R|!(|N| - |R| - 1)!}{|N|!} [f_x(R \cup \{i\}) - f_x(R)], \tag{9}$$

where ϕ_i is the SHAP attribution for feature i; N is the set of all features; R is any subset that excludes feature i; $|R|$ denotes the cardinality (size) of subset R; $|N|$ is the number of all features; $f_x(R)$ is the model's expected output when the features in R are fixed to their values in instance x and all remaining features are marginalized over a background distribution; and $f_x(R \cup \{i\})$ is

defined analogously but with i included. The weights $\frac{|R|!(|N|-|R|-1)!}{|N|!}$ average the marginal contribution of feature i across all possible feature orderings; here the exclamation mark "!" is the factorial operator, e.g., $k! = 1 \times 2 \times \cdots \times k$, which counts permutations and thus encodes the game-theoretic fairness axiom in SHAP.

Throughout, kNDVI served as the response and the four climatic factors as explanatory variables in XGBoost model; SHAP analysis was then used to identify sensitive regions and critical intervals of vegetation response.

2.3.5 GeoSOS-FLUS Model

GeoSOS-FLUS model is based on the traditional cellular automata framework and integrates an improved roulette wheel algorithm to allow more flexible simulation of future spatiotemporal patterns [**?**, **?**, **?**]. In the model, the neural network-based driving module is first trained using sample data from historical periods and integrates factors such as temperature, precipitation, humidity, and VPD to estimate the "suitability probability" of each pixel. In this study, pixel-wise kNDVI values were grouped into five numerical intervals: 0.00–0.10, 0.10–0.20, 0.20–0.30, 0.30–0.50, and > 0.50. By inferring the transition probabilities among these groups, the model dynamically simulated the succession and expansion processes of kNDVI types of China from 2025 to 2040 at spatiotemporal scale.

3.1.1 Temporal Dynamics of kNDVI

This study analyzed kNDVI change trend in China from 2000 to 2022 (Fig. 2). The results showed that kNDVI had a clear overall upward trend during the study period, accompanied by some interannual fluctuations. In 2000, kNDVI reached its lowest value of 0.29 and then gradually increased, reaching a maximum of 0.35 in 2021 before slightly declining to 0.33 in 2022. From 2000 to 2005, kNDVI exhibited a continuous upward trend, rising from 0.29 to 0.32, indicating a steady enhancement in vegetation vitality during this period. From 2006 to 2011, kNDVI fluctuated slightly but remained relatively stable, ranging from 0.31 to 0.32. After 2012, the upward trend became more pronounced, particularly from 2017 to 2021, when the upward trend was significant and reached the maximum of 0.34 in 2021. In 2022, kNDVI dropped slightly to 0.33 compared with 2021, a change likely associated with localized extreme weather events or short-term vegetation fluctuations.

3.1.2 Spatial Dynamics of kNDVI

This study revealed the spatial distribution pattern and the spatial heterogeneity of change trends in kNDVI across China from 2000 to 2022 (Fig. 3). The results showed that regional average kNDVI values ranged from 0.00 to 0.71. High-value areas (kNDVI > 0.50) were mainly concentrated in the southwestern

mountainous regions, northeastern forest areas, and the eastern Qinghai-Xizang Plateau. Low-value areas (kNDVI < 0.20) were widely distributed in the arid desert areas of the northwest and the alpine meadow areas in the northern Qinghai-Xizang Plateau.

Trend analysis showed that kNDVI changes exhibited significant spatial differentiation: 52.77% of the area had a significant increasing trend ($P < 0.05$), mainly in the central and eastern plains' agricultural regions, southern hilly areas, and regions where ecological restoration projects have been implemented (such as areas of returning farmland to forest). A total of 33.12% of the area showed a non-significant increasing trend, mainly at the edges of arid deserts in the northwest and north, and the Qinghai-Xizang Plateau, and the Sichuan Basin. Areas with a significant decreasing trend accounted for only 2.68%, sporadically distributed in southeastern coastal urban expansion zones, the Yangtze River Delta, and some edge areas of arid zones in the northwest, suggesting that land use conversion and urban expansion have exerted a suppressive effect on vegetation growth. In addition, 11.42% of the area showed a non-significant decreasing trend, mainly in ecologically fragile alpine meadows and sparse shrub regions.

The kNDVI values presented in Figures 2 and 3 were derived directly from the publicly available NDVI dataset using kernel methods, without additional masking of non-vegetation areas. Consequently, low kNDVI values in non-vegetation areas reflected their actual land cover conditions.

3.1.3 Comparison of Spatial Variability Between NDVI and kNDVI

A comparison of the coefficients of variation (CV) for NDVI and kNDVI from 2000 to 2022 revealed pronounced differences in temporal stability (Fig. 4). NDVI exhibited extremely low variability ($0.00 \leq \mathrm{CV} < 0.10$) across 64.58% of China, mainly in the eastern and northeastern plains, the southern hilly-forested belt, and parts of the Qinghai-Xizang Plateau, reflecting regulation by humid and stable climates or long-term agricultural and forestry management. NDVI exhibited moderate variability ($0.10 \leq \mathrm{CV} < 0.20$) across the Loess Plateau and several ecotones, accounting for 22.66%, whereas high variability ($0.40 \leq \mathrm{CV} < 1.00$) was limited to desert-oasis fringes and isolated mountains in Northwest China, accounting for 2.17%.

kNDVI showed a much smaller low variability ($0.00 \leq \mathrm{CV} < 0.10$), restricted to the southeast coast, the North China Plain, northeastern agroforestry mosaics, and the southern Qinghai-Xizang Plateau, accounting for 41.29%. The percentage of moderate variability ($0.10 \leq \mathrm{CV} < 0.20$) accounted for 23.41%, and the percentage of medium-to-high variability ($0.20 \leq \mathrm{CV} < 0.30$) accounted for 13.18% (compared with only 8.79% for NDVI). Area of kNDVI with CV of 0.30–0.40 (10.61%) occurred in the Inner Mongolian Plateau and the northern Qinghai-Xizang alpine meadows; it was much larger than the area of NDVI with CV of 0.30–0.40 (1.80%), which was dispersed in the western Kunlun Moun-

tains, the northern Tianshan Mountains, and the southeastern Qinghai-Xizang Plateau. Almost 11.51% of the land—mainly in the central-western Inner Mongolia Plateau and arid desert margins—exhibited CV value of over 0.40.

These differences arose from the intrinsic design of the two indices. NDVI saturated in densely vegetated or intensively managed areas, producing uniformly low CV in humid croplands and subtropical forests. In contrast, kNDVI employed a kernel transformation that enhances anomaly signals and suppresses background noise, reducing saturation effects and increasing sensitivity to interannual extremes. Consequently, climatically volatile or geomorphically transitional regions (e.g., plateau–desert interfaces and mountain ecotones) displayed higher temporal variability in kNDVI. Highly irrigated or canopy-closed stable zones remained low in both metrics, but their proportion decreased in the kNDVI spatial distribution because subtle phenological shifts were detected.

3.2.1 Spatial Pattern of Climatic Factors

From 2000 to 2022, climatic factors in China exhibited significant geographical differentiation (Fig. 5). Temperature showed a clear north-south gradient, with the highest annual averages ($> 20.00^{\circ}$C) in the southern regions, peaking in the Pearl River Delta, the southern coastal areas, and the Hainan Island (up to 29.70°C). The lowest temperatures occurred mainly on the Qinghai-Xizang Plateau, where some areas fell below -19.40°C, reflecting the significant regulatory effect of high-altitude topography on temperature distribution.

Precipitation displayed a spatial differentiation pattern of humid southeast and arid northwest. The highest values (2959.20 mm) appeared in the southeastern regions, including the Pearl River Delta and the southeastern coastal zones. In contrast, precipitation in the arid inland northwest approached 9.80 mm, highlighting the strong controlling effect of southeastern monsoon circulation and topographic barriers on moisture transport.

Humidity patterns closely matched the distribution of precipitation. High humidity areas ($> 80.00\%$) occurred mainly in the southeastern monsoon region south of the Yangtze River, the Sichuan Basin, the Hengduan Mountains, and the southern coastal areas. In contrast, the arid northwest and the Inner Mongolian Plateau generally exhibited low humidity levels ($< 20.00\%$).

VPD, which represents atmospheric aridity, showed low values (< 0.30 kPa) in the Qinghai-Xizang Plateau, the western mountainous areas of the Sichuan Basin, and parts of southern China. High VPD values (1.90 kPa) occurred mainly in the northeastern region and part of the Inner Mongolian Plateau, corresponding to low atmospheric aridity levels linked to low temperatures and high altitudes.

3.2.2 GCCM-Based Analysis of kNDVI Driving Mechanism

GCCM revealed the spatiotemporal causal relationships between kNDVI and the four climatic factors in China from 2000 to 2022 (Fig. 6). The results indicated that the interactions between vegetation and the four climatic factors exhibited significant spatial scale dependence and bidirectional feedback mechanisms.

When the library length L was less than 50, kNDVI exerted a stronger influence on temperature, with causal strength increasing from 0.04 at $L = 10$ to 0.15 at $L = 50$, indicating that vegetation regulates the local microclimate through short-term transpiration effects. When $L > 100$, the driving effect of temperature on kNDVI significantly increased, with causal strength rising from 0.06 at $L = 50$ to 0.33 at $L = 210$. The relationship between kNDVI and precipitation showed clear asymmetry. When $L < 100$, precipitation had a relatively weak causal effect on kNDVI (0.06), but the effect increased to 0.48 at $L = 210$, reflecting the long-term cumulative influence of precipitation on vegetation recovery. At $L = 210$, the feedback effect of kNDVI on precipitation was also relatively strong. The interaction between kNDVI and humidity exhibited an evident bidirectional enhancement. At $L = 50$, the causal strength in both directions was less than 0.16. At $L = 210$, the causal strength of kNDVI on humidity reached 0.71, while that of humidity on kNDVI was 0.70, indicating a stable positive feedback mechanism between vegetation growth and humidity over longer time scales. A competitive relationship emerged between kNDVI and VPD. When $L < 100$, kNDVI more strongly affected VPD, with a causal strength of 0.06 at $L = 50$. At $L = 210$, the suppressive effect of VPD on kNDVI became dominant, reaching a causal strength of 0.21.

3.2.3 Identification of Nonlinear Threshold of kNDVI

Figure 7 presents the nonlinear responses and threshold ranges between kNDVI and the four climatic factors. All climatic factors exhibited evident nonlinear influences on kNDVI. Temperature displayed a typical "U"-shaped response, whereas precipitation, humidity, and VPD showed "S"-shaped response trends, indicating that their effects on kNDVI varied significantly across different value ranges. This study identified the optimal threshold ranges of each factor for kNDVI. The optimal temperature threshold range was −12.18°C–0.67°C, within which kNDVI changes were more pronounced. For precipitation, the optimal threshold range was 24.00–159.74 mm. The optimal humidity threshold was below 22.00%. For VPD, the optimal threshold included three parts: < 0.07, 0.17–0.24, and > 0.30 kPa. Overall, different climatic factors significantly influenced kNDVI trends within specific threshold ranges, revealing a clear nonlinear response of vegetation status to environmental conditions.

3.3.1 Accuracy Assessment

To evaluate the performance of the coupled GCCM–GeoSOS-FLUS framework, we conducted a temporal validation by using the kNDVI dataset from 2000 as

the baseline input and forecasting the vegetation distribution for 2010. The predicted 2010 kNDVI map was then compared with the actual kNDVI observations for the same year (Fig. 8). The simulation reproduced the spatial distribution of vegetation with a high level of consistency, achieving an overall accuracy of 0.92, which meets the reliability threshold of the Kappa coefficient test. This result confirmed that the integrated approach can accurately capture decadal vegetation dynamics and provided a robust basis for projecting future kNDVI spatial patterns across China.

3.3.2 Spatial Prediction of kNDVI

The spatial prediction results based on the GeoSOS-FLUS model indicated that from 2025 to 2040, the spatial distribution pattern of kNDVI in China will generally follow a gradual optimization trend, although regional heterogeneity will remain evident (Fig. 9). The proportion of high kNDVI values (> 0.50) will increase from 40.43% in 2025 to 41.85% in 2040, with a clear spatial concentration in the Sichuan Basin, the Yunnan-Guizhou Plateau, the middle and lower reaches of the Yangtze River Plain, and southern hilly regions. The stable expansion of vegetation cover in these regions reflects the long-term synergy between ecological protection measures and humid climatic conditions. The area proportion of kNDVI values in 0.30–0.50 will remain relatively stable at around 15.00% over the period from 2025 to 2040, concentrated in the southern Loess Plateau, the eastern part of the Qinghai-Xizang Plateau, the northern edge of the North China Plain, and parts of the agropastoral ecotone in Northeast China. This result demonstrates that in the future vegetation that gains from ecological restoration efforts can effectively offset the pressures of agriculture and grazing, maintaining moderate vegetation cover.

The proportion of low kNDVI values (0.00–0.10) will decline from 32.92% in 2025 to 31.67% in 2040, mainly in the Tarim Basin, the Qaidam Basin, and alpine desert areas of northwestern Qinghai-Xizang Plateau. These regions will continue to experience slow vegetation recovery under extremely arid and cold conditions. The proportion of kNDVI values of 0.10–0.30 will change little (11.65% in 2025 and 11.51% in 2040), with wide distribution in the central and western Inner Mongolia Plateau, the Hexi Corridor, and the central Qinghai-Xizang Plateau, affected by the combined impacts of grassland degradation, overgrazing, or urban expansion.

The increase in the proportion of high-value areas is limited, and its spatial expansion trend is consistent with the long-term positive feedback mechanism between humidity and vegetation revealed by the GCCM model. In contrast, the slow reduction of low kNDVI areas in the arid northwest corresponds with the long-term stress effects of rising temperature and increasing VPD on vegetation, as revealed by the earlier GCCM model analysis.

4.1 Causal Spatio-Temporal Prediction and Nonlinear Threshold Effects of Vegetation Dynamics

The interactions between vegetation dynamics and climatic factors remain a core scientific issue in understanding ecosystem responses to global change. Existing studies have often analyzed the relationship between climatic factors and vegetation indices under linear assumptions. However, growing evidence suggests that vegetation responses to climate change exhibit significant nonlinear characteristics and threshold effects [**?**]. By integrating GCCM-based causal inference and XGBoost-SHAP threshold identification, this study systematically uncovered the spatiotemporal driving mechanisms and nonlinear response patterns between kNDVI and climatic factors in China. This approach offers a new perspective for understanding the complex coupling relationships between vegetation and climate.

Regarding the spatial differentiation of climatic factors, the geographical patterns of temperature, precipitation, humidity, and VPD in China align closely with previous studies [**?**, **?**]. This study revealed that the Qinghai-Xizang Plateau and the western Sichuan Basin have significantly low VPD. These areas are influenced by the combined effects of moist monsoonal inflow and orographic uplift, forming a characteristic hydrothermal redistribution that suppresses VPD. By contrast, the highest VPD values are confined to the hyper-arid basins and leeward margins of Northwest China, where local values approach the maximum (1.90 kPa). In addition, high VPD is not solely driven by aridity but is also linked to topography-induced large diurnal temperature variations.

Regarding the causal mechanisms between vegetation and climatic factors, this study demonstrated the pronounced dependence of the inferred direction of influence on spatial and temporal scale. At short-term scales ($L < 50$), the dominant influence of kNDVI on temperature relates closely to the feedback mechanism of vegetation transpiration on the local microclimate [**?**]. This result supported the theory of a bidirectional interaction between vegetation and climate. Unlike earlier work emphasizing albedo-mediated effects, our analysis suggested that in humid regions (e.g., South China) and over short-time scales, vegetation regulates near-surface air temperature mainly via transpiration: by increasing air humidity, more energy is partitioned into latent heat and less into sensible heating, which cools the air [**?**]. At long-term scales ($L > 100$), the enhanced driving effect of temperature on kNDVI may reflect the nonlinear decline in the vegetation carbon sink function caused by accumulated temperature effects. When temperature exceeds the regional optimal threshold, the efficiency of vegetation photosynthesis decreases [**?**]. The optimal temperature threshold identified in this study (-12.18°C–0.67°C) overlaps with the climatic transition zone from boreal coniferous forests to temperate deciduous forests, suggesting that vegetation type conversion may be an important mechanism driving threshold effects.

The long-term cumulative effect of precipitation on kNDVI (causal strength

reaching 0.48 at $L = 210$) underscores its role as a major constraint on vegetation recovery in arid and semi-arid areas. Although short-term precipitation fluctuations exert limited influence on vegetation growth, prolonged precipitation deficits can cause irreversible losses in soil moisture storage capacity [**?**], which aligns with the lower precipitation threshold (24.00 mm) identified in this study. When precipitation falls below this threshold, vegetation cannot restore productivity through short-term water supply. Notably, the feedback strength of kNDVI on precipitation increases significantly at long-term scales, potentially due to hydrological improvements such as reduced runoff and enhanced soil water-holding capacity, driven by increased vegetation cover [**?**]. This study quantified this lag effect at regional scale, offering a reference point for assessing long-term benefits of ecological projects.

The bidirectional enhancement effect between humidity and kNDVI (causal strengths > 0.70 at $L = 210$) reflects the close coupling between vegetation and the atmospheric water cycle. High humidity environments promote vegetation photosynthesis by reducing stomatal resistance [**?**], while vegetation transpiration further increases air humidity, forming a self-sustaining positive feedback loop. This mechanism is particularly evident in subtropical evergreen broadleaf forest regions but may also elevate the ecological risk of extreme wet events. When humidity exceeds 22.00%, this study observed a significant decline in the response probability of kNDVI, which may relate to pest outbreaks or root hypoxia resulting from prolonged high humidity [**?**].

The suppressive effect of VPD on kNDVI (causal strength reaching 0.21 at $L = 210$) highlights the profound impact of atmospheric aridity on vegetation growth. At short-term scales, vegetation can partially offset the increased transpiration pressure from rising VPD through stomatal regulation [**?**]. However, long-term high VPD environments can lead to the collapse of plant hydraulic systems [**?**]. The VPD thresholds identified in this study (< 0.07, 0.17–0.24, and > 0.30 kPa) closely align with the hydraulic safety margins of different vegetation types: the low VPD threshold (< 0.07 kPa) may reflect the adaptive limits of tundra vegetation to extremely humid conditions, while the high threshold (> 0.30 kPa) may correspond to the water-use strategies of desert shrubs. These findings indicated that, under future climate change, increasing atmospheric aridity may preferentially threaten the stability of ecotone.

The nonlinear response mechanisms between vegetation and climate revealed in this study hold important theoretical and applied value. They enhance understanding of the regulatory mechanisms of vegetation dynamics in China and offer a scientific basis for ecosystem management under climate change.

Moreover, the observed spatiotemporal variability in vegetation dynamics not only arises from climatic factors but also from anthropogenic influences, such as ecological restoration policies (e.g., returning farmland to forest and grassland projects), urbanization, and land use conversion [**?**]. For instance, vegetation improvements in the central-eastern agricultural and ecological restoration zones largely reflect the success of sustained ecological interventions and enhanced

agricultural management practices [**?**]. Conversely, vegetation degradation in the southeastern coastal urban areas and parts of the arid northwest is primarily linked to rapid urban expansion, intensified land-use changes, and pressures from overgrazing or excessive resource exploitation [**?**]. Therefore, fully understanding vegetation spatiotemporal dynamics requires integrating both climatic drivers and socio-economic factors to develop a comprehensive ecosystem management strategy.

4.2 Challenges and Strategy Optimization for Future Vegetation Restoration

From 2025 to 2040, vegetation cover improvement in China will follow a spatially heterogeneous pattern characterized by "southeastern optimization and northwestern delay". The proportion of kNDVI values over 0.50 will gradually increase to 41.85%, concentrated in the Sichuan Basin, the Yunnan-Guizhou Plateau, and southern China. This is closely related to the long-term positive feedback between humidity and vegetation, the sustained cumulative effect of precipitation, and the implementation of ecological protection measures. However, the growing risk of extreme precipitation events in the southeastern regions may threaten local vegetation stability and should be addressed in future ecological protection policies [**?**].

In contrast, the proportion of areas with low kNDVI values (0.00–0.10) in the arid northwest will decline slowly to 31.67%, yet ecological restoration remains far behind the southeast, largely due to the enduring stress from high temperatures and VPD. Enhancing vegetation adaptability to climate change in this region will require cross-regional ecological compensation, promotion of water-saving technologies, and introduction of drought-tolerant species. The proportion of kNDVI values of 0.10–0.30 remains stable at 11.51% in regions such as the Inner Mongolian Plateau, the edge of the North China Plain, and the Hexi Corridor, where urban expansion and grassland degradation exert combined pressures. Therefore, strict ecological regulation and optimization of the agricultural-ecological land use structure should be implemented in land use planning, and differentiated strategies should be adopted to effectively balance economic development and ecological protection needs.

Compared with the existing "Ecological Conservation Redline" policy, the optimization strategies proposed in this study emphasize differentiated measures for different regions: in the humid southeastern region, maintaining self-sustaining ecological mechanisms should remain the priority [**?**]; in the arid northwest, enhancing the climatic resilience of vegetation should be prioritized; and in the agropastoral ecotone and urban expansion zones, optimizing land use structure is essential to promote the coordinated and sustainable development of both ecology and agriculture.

4.3 Limitation and Prospect

Although this study integrated multi-scale data and advanced modeling approaches, several limitations should be acknowledged. First, the spatial resolution constraints of the datasets used may underestimate the effects of localized human activities, such as urban heat islands and intensive agricultural practices. Consequently, critical ecological processes occurring at finer spatial scales might not have been fully captured, potentially limiting the accuracy of vegetation change assessments in highly heterogeneous landscapes [**?**]. Second, while GCCM effectively identified causal relationships between vegetation and climatic factors, it didn't fully quantify the interactive contributions of natural variability and anthropogenic interventions. Therefore, the mechanisms driving vegetation dynamics under complex human-natural interactions remain partly unexplored [**?**]. Third, the vegetation projections do not account for variations associated with extreme climatic events or differentiated climate change scenarios. Given the increasing frequency of extreme weather conditions under global climate change, neglecting these scenarios may increase uncertainty in predicting vegetation responses, especially in ecologically fragile regions [**?**].

To address these limitations, future research could incorporate high-resolution remote sensing datasets (e.g., Sentinel series and unmanned aerial vehicle (UAV)-based imagery) to more precisely capture localized ecological processes. Integrating dynamic vegetation models (DVMs) or process-based ecological models would further improve understanding of vegetation-climate interactions at fine scales. Additionally, incorporating comprehensive socio-economic development scenarios, such as the Shared Socio-economic Pathways (SSP) combined with Representative Concentration Pathways (RCP), would allow more robust evaluation of how human development trajectories and climate policies jointly shape vegetation dynamics. Incorporating these elements will significantly enhance the robustness, accuracy, and practical applicability of vegetation prediction models, thus providing more reliable guidance for ecological management and climate change adaptation strategies.

5 Conclusions

This study analyzed the spatiotemporal changes, driving mechanisms, and nonlinear thresholds of vegetation in China from 2000 to 2022 by integrating kNDVI and climate data within a multi-model coupling framework (GCCM–XGBoost-SHAP–GeoSOS-FLUS). The results showed that vegetation exhibited an overall increasing trend, accompanied by significant spatial heterogeneity: notable improvement occurred in the central-eastern agricultural and ecological restoration zones, whereas degradation was evident in the southeastern coastal urban areas and the northwestern arid margins. Vegetation-climate interactions showed clear spatiotemporal scale dependence, with temperature regulating vegetation dynamics at long-term scales and humidity forming strong bidirectional feedback. Nonlinear responses revealed distinct climatic thresholds—temperature (-12.18°C–0.67°C), precipitation (24.00–159.74 mm), humidity ($< 22.00\%$), and

VPD (< 0.07, 0.17–0.24, and > 0.30 kPa). Future projections suggest persistent spatial differentiation, with enhanced vegetation recovery in the southeast contrasting with slower restoration in the northwest. These findings provide critical insights for regional ecological management, helping prioritize restoration efforts and design adaptation strategies under future climatic scenarios.

Conflicts of Interest

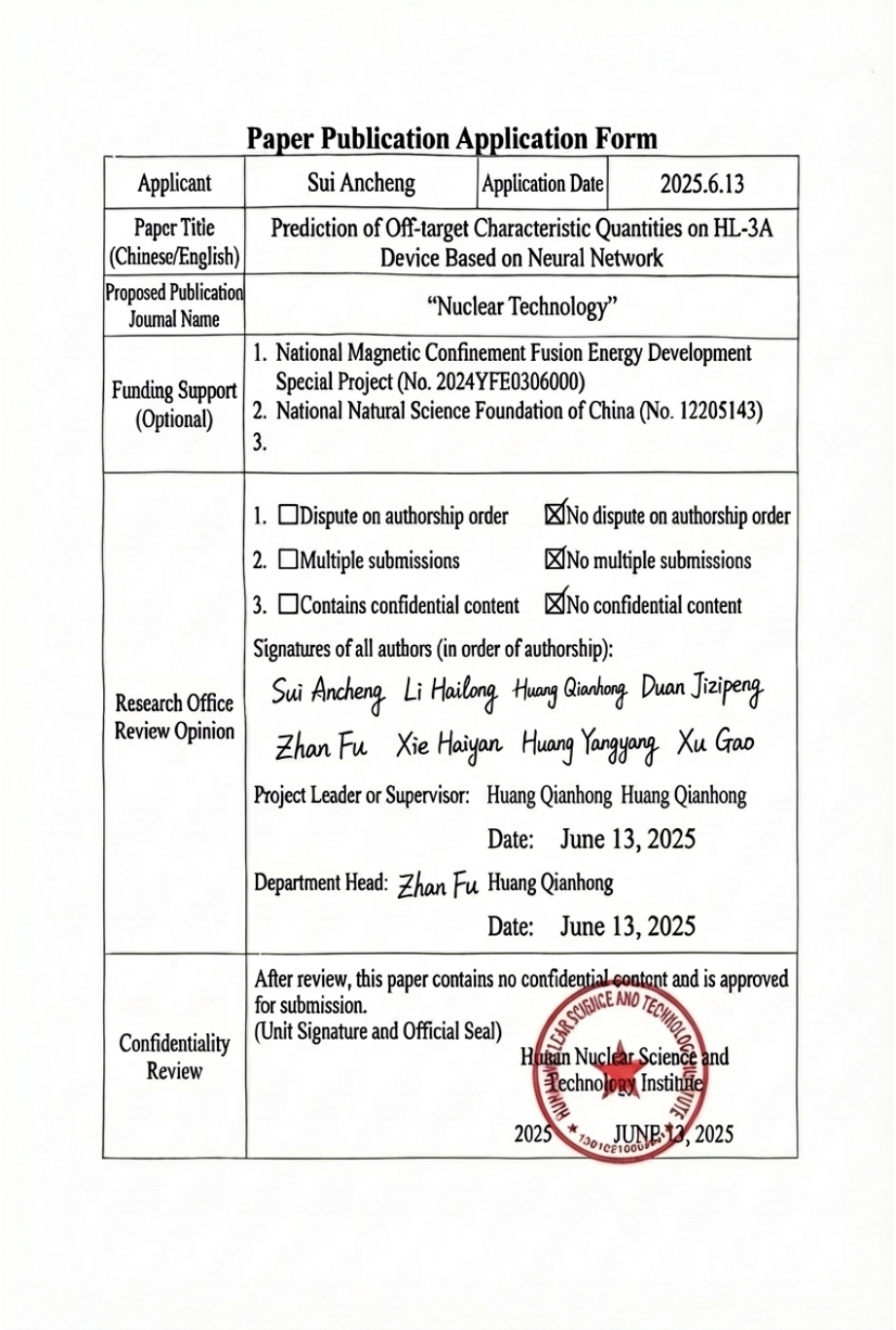
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the Key Science and Technology Research Projects of Henan Province (252102320).

Figures

Source: ChinaXiv — Machine translation. Verify with original.



Paper Publication Application Form

Applicant	Sui Ancheng	Application Date	2025.6.13
Paper Title (Chinese/English)	Prediction of Off-target Characteristic Quantities on HL-3A Device Based on Neural Network		
Proposed Publication Journal Name	"Nuclear Technology"		
Funding Support (Optional)	1. National Magnetic Confinement Fusion Energy Development Special Project (No. 2024YFE0306000) 2. National Natural Science Foundation of China (No. 12205143) 3.		
Research Office Review Opinion	1. ☐Dispute on authorship order ☒No dispute on authorship order 2. ☐Multiple submissions ☒No multiple submissions 3. ☐Contains confidential content ☒No confidential content Signatures of all authors (in order of authorship): Sui Ancheng Li Hailong Huang Qianhong Duan Jizipeng Zhan Fu Xie Haiyan Huang Yangyang Xu Gao Project Leader or Supervisor: Huang Qianhong Huang Qianhong Date: June 13, 2025 Department Head: Zhan Fu Huang Qianhong Date: June 13, 2025		
Confidentiality Review	After review, this paper contains no confidential content and is approved for submission. (Unit Signature and Official Seal) Hunan Nuclear Science and Technology Institute 2025 JUNE 13, 2025		

Figure 1: Figure 1

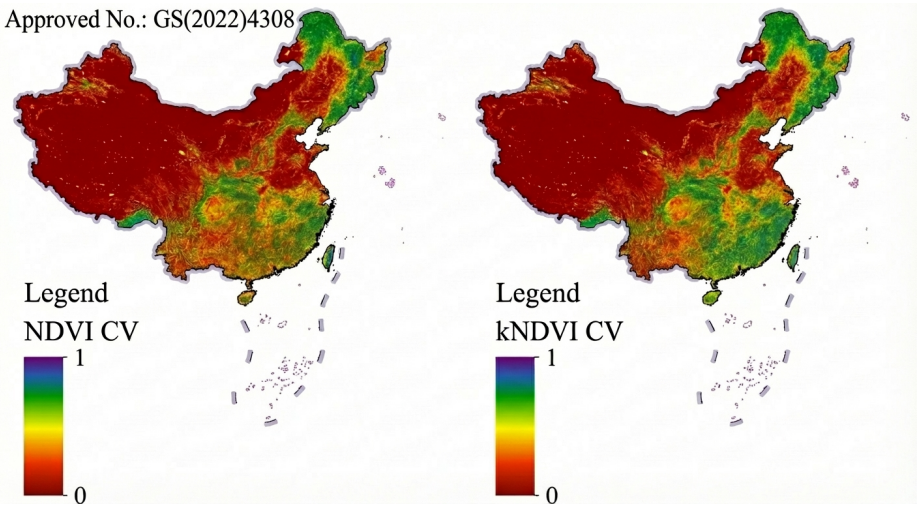


Figure 2: Figure 12

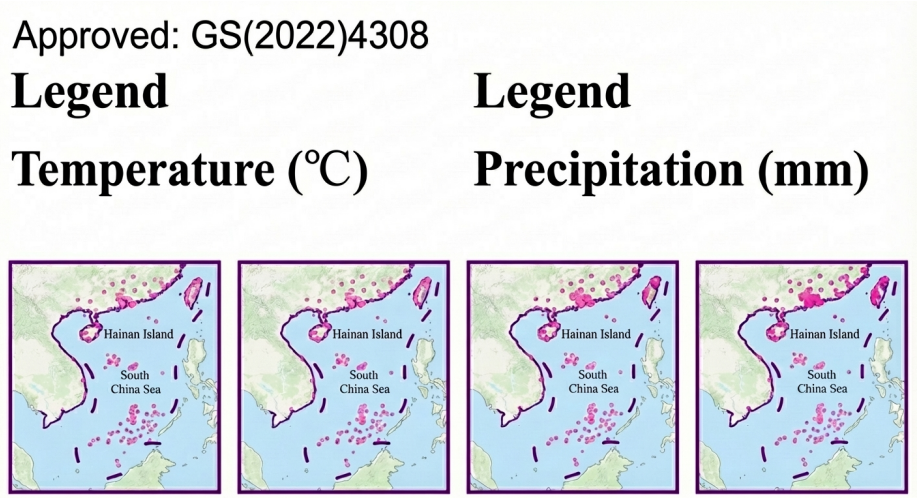


Figure 3: Figure 18

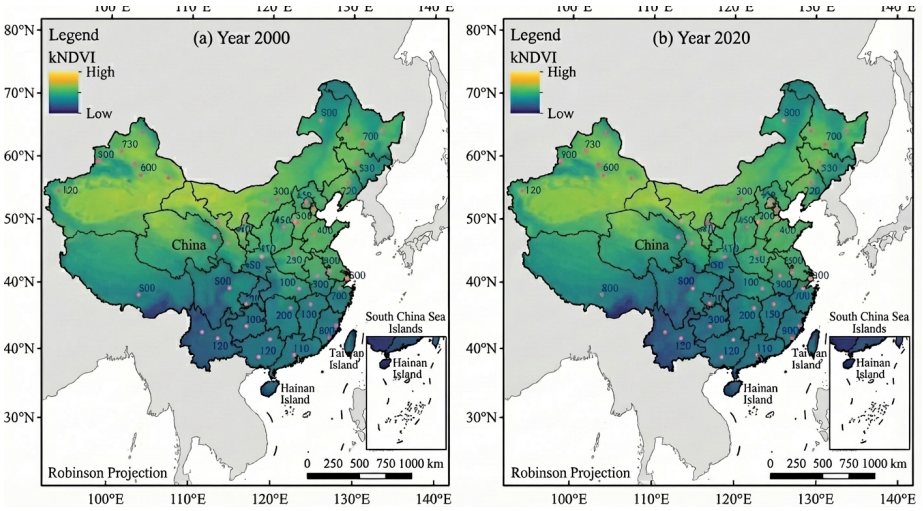


Figure 4: Figure 46