NONLINEAR EFFECTS OF ENVIRONMENTAL FACTORS ON SPECIES RICHNESS IN THE WESTERN TIEN SHAN USING A GENERALIZED ADDITIVE MODEL

Feruz Akbarov

*PhD; Sardor Poʻlatov, PhD; Ulugʻbek Qodirov, PhD Institute of Botany, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan *Corresponding author: Feruz Akbarov (feruzakbar88@gmail.com)*

Abstract: *This study evaluated the nonlinear effects of environmental factors on species richness in the Western Tien Shan using a Generalized Additive Model (GAM) with a Negative Binomial family. Species richness was analyzed at the 5 × 5 km grid-cell level, with sampling effort included as a control variable. The final model explained 97.9% of the total variation (Adjusted R^2 = 0.956), with sampling effort identified as the strongest predictor of species richness (edf = 3.88; $p < 0.001$). After correcting for the sampling effect, only annual precipitation (BIO12) retained a small but significant independent effect ($p = 0.039$). These findings show that sampling density is critical for interpreting real species richness patterns and must be explicitly considered in biodiversity hotspot assessment.*

Keywords: *species richness, sampling effort, GAM, biodiversity hotspot, Western Tien Shan*

INTRODUCTION

In recent years, the rapid growth of herbarium and occurrence databases has greatly expanded opportunities for large-scale biodiversity assessments. However, species richness patterns derived from such data are often strongly influenced by uneven sampling effort, which may obscure the true effects of environmental drivers on biodiversity patterns (Zizka et al., 2021; Engemann et al., 2015). Spatial sampling bias is now widely recognized as one of the major methodological challenges in biodiversity modeling, species distribution analysis, and richness mapping (Beck et al., 2014; Meyer et al., 2016). Mountain ecosystems such as the Western Tien Shan are characterized by steep environmental gradients and high floristic heterogeneity, making them ideal systems for evaluating the nonlinear responses of species richness to climatic, edaphic, and topographic factors. At the same time, herbarium collections from mountain regions are frequently concentrated in accessible valleys, roadsides, and well-surveyed floristic hotspots, which may artificially inflate richness estimates in heavily sampled locations (Zizka et al., 2021). To address this issue, Generalized Additive Models (GAMs) provide an effective framework for modeling nonlinear relationships between species richness and environmental predictors while simultaneously controlling for sampling intensity (Wood, 2017). Therefore, the aim of this study was to evaluate the nonlinear effects of environmental factors on species richness in the Western Tien Shan and to quantify the extent to which sampling effort acts as the dominant predictor of observed richness patterns.

Materials and Methods

Species richness was modeled using a Generalized Additive Model (GAM) to capture nonlinear relationships between richness and environmental predictors. Because the response variable consisted of overdispersed count data, a Negative Binomial distribution was used, which is widely recommended for ecological count datasets (Wood, 2017; Zuur et al., 2009).

The response variable was species richness calculated for 5 × 5 km grid cells across the Western Tien Shan. Predictor variables included climatic (BIO1, BIO12), edaphic (AWCh1, CECSOL, CRFVOL), and topographic (Elevation, slope classes) factors. To control for uneven observation intensity, sampling effort based on the number of herbarium records per grid cell was log-transformed and included in the model as a covariate, following current recommendations for correcting richness bias in biodiversity analyses.

Model performance was evaluated using adjusted R^2, estimated degrees of freedom (edf), and p-values. Smooth response curves were examined to interpret saturation patterns in species richness and to identify which environmental variables retained independent significance after controlling for sampling effort.

Results

Nonlinear effects of environmental factors on species richness

The relationships between species richness and environmental factors were evaluated using a Generalized Additive Model (GAM) with a Negative Binomial family, taking into account the characteristics of overdispersed count data (Fig. 1).

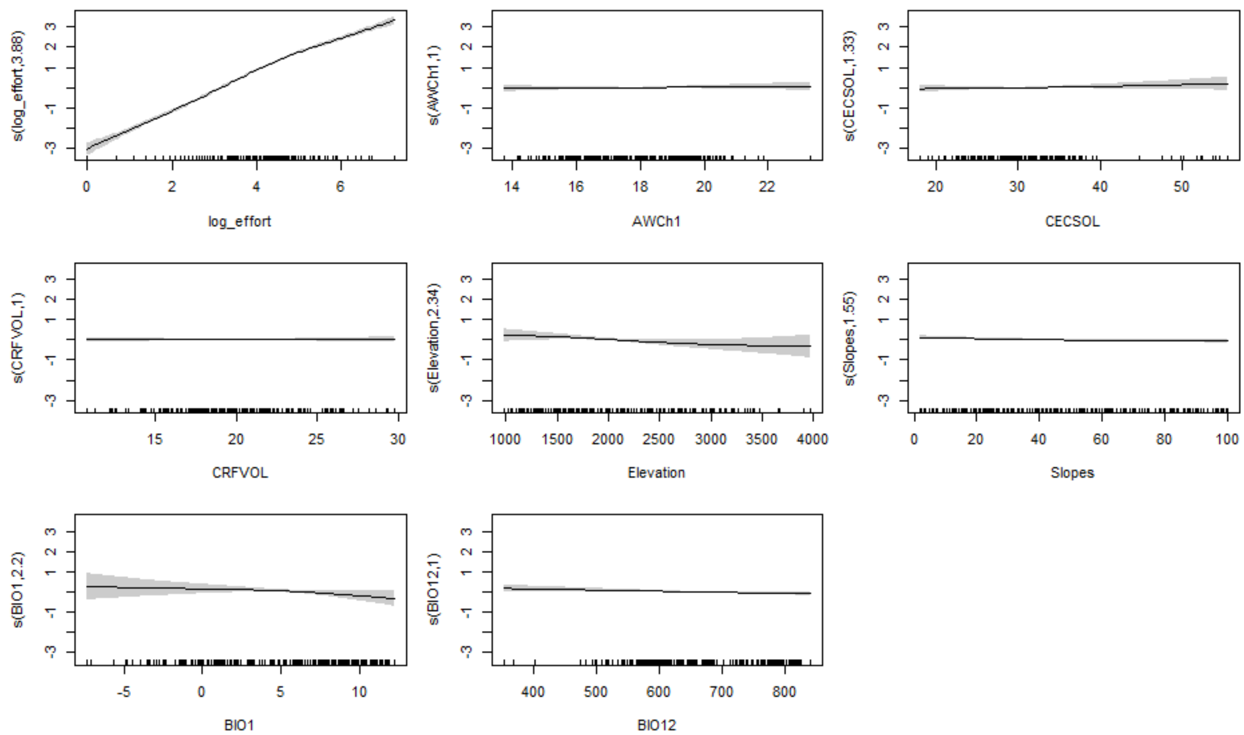

Fig. 1. Nonlinear effects of environmental factors on species richness.

Sampling effort (log-scaled) was included in the model as a control variable for observation intensity. The final GAM explained 97.9% of the total variation in species richness (Adjusted $R^2 = 0.956$). Observation intensity emerged as the strongest and most stable predictor of species richness (edf = 3.88; $p < 0.001$). The smooth curves showed that species richness increased sharply with increasing sampling effort and then approached a saturation phase, confirming the necessity of mandatory correction for sampling effects in order to isolate the true influence of environmental factors.

After controlling for sampling effort, only annual precipitation (BIO12) showed a statistically significant effect on species richness ($p = 0.039$). However, the amplitude of this effect was small, indicating that most of the regional variation in species richness is explained by observation intensity. Other climatic (BIO1), edaphic (AWCh1, CECSOL, CRFVOL), and topographic (Elevation, Slopes) variables did not show statistically significant independent effects in the model. Thus, while PCA and Z-score analyses demonstrated that ecological gradients-particularly topographic complexity and the elevation–temperature relationship-shape the spatial structure of the region, GAM results revealed that this signal is strongly modulated by sampling density, with only a weak but reliable independent effect retained for BIO12.

Discussion

The GAM results demonstrate that sampling effort is the strongest determinant of observed species richness patterns in the Western Tien Shan. The pronounced nonlinear saturation response indicates that species richness increases rapidly with additional herbarium records at low sampling levels and gradually approaches a plateau as sampling becomes more complete. This pattern reflects the classical effect of sampling completeness in biodiversity inventories and confirms that richness estimates derived from occurrence data may be strongly biased when observation intensity is not explicitly controlled (Beck et al., 2014; Zizka et al., 2021).

After accounting for sampling effort, only annual precipitation (BIO12) retained a statistically significant independent effect, although the amplitude of this response remained weak. This suggests that precipitation acts as a real ecological driver of species richness in the region, but its contribution is substantially smaller than the observational signal generated by collection density. Such a result is ecologically plausible for mountain ecosystems, where moisture availability is a key factor influencing floristic richness.

Overall, these findings highlight that correction for sampling effort is a methodological necessity in herbarium-based richness modeling. Without such correction, biodiversity patterns and hotspot delineation may largely reflect collection intensity rather than true ecological processes.

Conclusion

The GAM model results for species richness showed that the major part of the observed variation is explained by sampling effort, indicating that observation intensity is the strongest predictor of species richness. After controlling for the sampling effect, only annual precipitation (BIO12) showed a small but reliable independent effect among the environmental factors. These findings confirm that, although ecological gradients are strong at the regional scale, accounting for sampling density is of decisive importance for interpreting real species richness maps.

Acknowledgements

This research was carried out within the framework of the State Program of the Institute of Botany, Academy of Sciences of the Republic of Uzbekistan, entitled “Digital Nature: Development of a Digital Platform for the Flora of Central Uzbekistan” (2025–2029).

The study was also supported by the fundamental research projects "Taxonomic revision of polymorphic families in the flora of Uzbekistan" (A-FA-2021-427) and "Transboundary biodiversity hotspots of Uzbekistan and their current status" (FL-9024093685), as well as applied projects "Bioclimatic and geospatial analysis of medicinal plant resources of the Republic, patterns of adaptation and distribution (on the example of the genus Ferula L.)" (FL-9524115063) and "Preparation of the first Uzbek-language edition of the Flora of Uzbekistan" (AL-9224104450).

REFERENCES:

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on species distribution models. Ecological Informatics, 19, 10–15.

Engemann, K., Sandel, B., Boyle, B., Enquist, B. J., Jørgensen, P. M., Morueta-Holme, N., ... Svenning, J.-C. (2015). Limited sampling hampers big-data estimation of species richness in a tropical biodiversity hotspot. Ecology and Evolution, 5(3), 807–820.

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 19, 992–1006.

Smith, J. A. et al. (2024). Guidance on accounting for sampling effort in generalized models. EcoEvoRxiv.

Wood, S. N. (2017). Generalized Additive Models: An Introduction with R (2nd ed.). Chapman & Hall/CRC.

Zizka, A., Antonelli, A., & Silvestro, D. (2021). sampbias: a method for quantifying geographic sampling biases in species distribution data. Ecography, 44, 25–32.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed Effects Models and Extensions in Ecology with R. Springer.