International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 09, Issue – 09, September 2024, PP – 01-11

Application of Species Distribution Models to Predict the Distribution Model of *Cinnamomum Balansae* Lecomte in Ta Đung National Park, Đak Nong, Viet Nam

Phan Thi Hang[1], Nguyen Thi Thanh Huong[2]

[1,2]*Faculty of Agriculture and Forestry, Tay Nguyen University, VietNam*

Abstract: Cinnamomum balansae Lecomte is a rare and endangered tree species in Vietnam in general and the Central Highlands in particular, with high ecological and economic value. Information on the species' distribution plays a crucial role in conservation efforts, especially in the context of current climate change. This study aims to apply the species distribution models (SDMs) to predict suitable habitats and potential areas where Cinnamomum balansae Lecomte appears in Ta Dung National Park, Dak Nong, Vietnam. The results recorded 72 distribution points within the study area, with 48 points used to build the model and 24 points used to test the model's accuracy. The study utilized Pearson correlation coefficients, PCA analysis, and the JackKnife method to select 8 important variables from the initial 25 variables affecting the habitat of Cinnamomum balansae. These variables include Bio04 (temperature seasonality), Bio01 (annual mean temperature), Bio07 (annual temperature range), Bio13 (precipitation of the wettest month), current forest cover status, NDVI, slope, and elevation. The MaxEnt model yielded an AUC score of 86%, indicating good performance in predicting the current and potential distribution of the species. This model provides a solid scientific basis for identifying priority areas for habitat conservation and management, ensuring the long-term survival of the species under increasingly severe climate change conditions

Keywords: SDM, MaxEnt, Cinnamomum balansae, GIS, Habitat suitability

I. Introduction

In recent years, Species Distribution Models (SDMs) have emerged as an effective tool in the fields of spatial ecology, conservation, and natural resource management. These models provide quantitative predictions of species' spatial distributions based on the relationship between observed data and environmental variables. The continuous development of SDMs has expanded their applicability across various fields, from assessing the impacts of climate change to managing biodiversity (Franklin, 2010; Guisan *et al.,* 2017; Thuiller *et al.,* 2019; Araújo *et al.,* 2020). Most SDMs have been primarily developed for common and economically valuable species. However, studies on SDMs for rare and valuable tree species remain relatively limited, despite their crucial role in biodiversity conservation. Rare tree species often have narrow geographic distributions, small population sizes, and are vulnerable to environmental changes and human impacts. Therefore, research and application of SDMs for this group of species are necessary to support conservation efforts, particularly for managers and policymakers in prioritizing monitoring, species conservation, and sustainable ecosystem management activities (Márcia *et al.,* 2020; Brown *et al.,* 2022).

Ta Dung National Park, Đak Nong, Vietnam, boasts a diverse and rich flora, including 1,406 vascular plant species belonging to 760 genera and 192 families across six plant divisions. The national park is characterized by diverse ecosystem types, including 90 rare plant species, among which are 21 rare tree species such as *Cinnamomum balansae*, *Aquilaria crassna, Dalbergia oliveri*, and *Sindora siamensis* (Biodiversity Report of Tà Đùng National Park, 2022). *Cinnamomum balansae* Lecomte (C. balansae) is a high-value tree species known for its distinctive fragrance, economic value from its wood, and particularly the medicinal value of the essential oil extracted from its bark and trunk (Duy *et al.,* 2021; Hai *et al.,* 2016). In recent years, the population of C. balansae has significantly declined due to overexploitation and poor natural regeneration, making it one of the rare tree species in Vietnam. C. balansae is currently classified under group IIA according to Decree 06/2019/NĐ-CP on the management of endangered, precious, and rare forest plants in Vietnam, listed as 'Critically Endangered' (CR) in the Vietnam Red Book and 'Endangered' (EN) according to the IUCN classification (2023) (IUCN, 2023). However, there has been no comprehensive survey or study on this species in Tà Đùng National Park to date. Therefore, Therefore, predicting the distribution model of the C. balansae is essential to support conservation efforts for this species.

In this study, we applied SDMs to predict suitable habitats and potential locations for the occurrence of C. balansae. The use of SDM plays a crucial role in identifying potential distribution areas, thereby providing a basis for determining priority areas for conservation and ecological restoration efforts (Guisan et al., 2017; Zurell et al., 2020; Liu et al., 2022)."

II. Material and Method

Research Area

Ta Dung National Park is located on the administrative boundary of Dak Som commune, Dak G'long district, Dak Nong province, Viet Nam.

Coordinates: From 11047'27" to 11059'20" North latitude; From 107053'10" to 10806'32" East longitude.

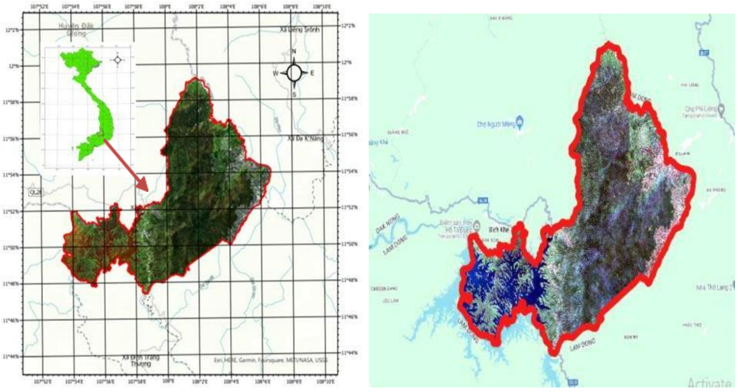


Figure 1: Location of Ta Dung National Park area

Material

The study utilized a dataset of 19 climate variables from the period 1970-2000, downloaded from WorldClim version 2.1 (Fick & Hijmans, 2017). The topographic data, including slope, aspect, and elevation, were also obtained from the Shuttle Radar Topography Mission (SRTM) data on WorldClim to maintain resolution consistency with the climate variables. Soil pH data were retrieved from Google Earth Engine (GEE), a powerful geospatial analysis tool that allows access to and processing of large-scale environmental data (Gorelick *et al.,* 2017). Information on the current forest cover status in Tà Đùng National Park was downloaded, processed, corrected, and classified using high-resolution SE2A imagery, enabling detailed assessment of forest changes and status (Hansen *et al.,* 2013). All these variables were standardized to a spatial resolution of 30s, approximately 1km x 1km, to match the resolution of the climate data (Karger *et al.,* 2017; Xu *et al.,* 2022) with the UTM WGS 84 zone 48 coordinate system.

In addition, the Normalized Difference Vegetation Index (NDVI) variable derived from SE2A imagery was also used in this study. The NDVI is based on the near-infrared (Visible and Near Infrared - VNIR) light that vegetation strongly reflects and the Red light that vegetation absorbs (Li *et al.,* 2023). In Sentinel-2 imagery, the VNIR band is in band B8, and the Red band is in band B4. The formula used to calculate the NDVI is as follows

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} = \frac{\mathrm{B8} - \mathrm{B4}}{\mathrm{B8} + \mathrm{B4}} \ (1)$$

During the survey at Ta Dung National Park, we established 900m² standard plots and conducted transect surveys to record the distribution points of the species. A total of 72 distribution points of the species C. balansae were recorded. Among these, 48 points were randomly selected to build the species distribution model, while the remaining 24 points were used to validate the model quality (Merow *et al.,* 2013). The recorded points were carefully chosen to represent the species' habitat diversity within the study area. To ensure accuracy, only C. balansae trees with a diameter at breast height greater than 6 cm were recorded and included in the analysis, to eliminate noise from young trees and increase the stability of the distribution model (Wisz *et al.,* 2020). This method improves the model's predictive capability and provides deeper insights into the environmental factors affecting the species' distribution (Franklin *et al.,* 2023).

Data Analyzis

In this study, we examined the collinearity of a series of data variables, including 19 climatic variables, topographic data, NDVI index, forest cover status, and soil pH. To assess the collinearity between variables, we calculated the Pearson correlation coefficient (r) for each pair of variables (Zuur *et al.,* 2020). Testing and treating collinearity among independent variables are a crucial step to limit multicollinearity, ensuring that the input variables in the species distribution model do not excessively depend on each other. This contributes to improving the accuracy and predictive efficiency of the model (Dormann *et al.*, 2018; Liu *et al.*, 2022).

Variables with high correlation coefficients (r > |0.85|) were considered for removal or transformation to minimize the negative impact of multicollinearity on the model's accuracy (Kuhn & Johnson, 2019; Mansour *et al.*, 2023). Additionally, Principal Component Analysis (PCA) is a powerful tool for dimensionality reduction, eliminating unnecessary noise while retaining the most important variables, thereby significantly improving the analysis and computation process (Jolliffe & Cadima, 2016). In our study, after analysis, 16 variables were retained for model development, including 10 climatic variables (Bio19, Bio16, Bio14, Bio13, Bio12, Bio07, Bio04, Bio03, Bio02, Bio01); 3 topographic variables (slope, elevation, aspect); and 3 other variables (soil pH; forest cover status, and NDVI).

Subsequently, we continued to use the Jackknife method to test the contribution of the selected variables to the construction of the SDMs Model. The Jackknife analysis results indicated that the variables Bio01, Bio04, Bio07, Bio13, NDVI, forest cover status, slope, and elevation significantly contributed to the simulation of the distribution area of the species C. balansae. Therefore, these variables were selected to run the distribution simulation, serving as a basis for identifying potential distribution areas and supporting effective species conservation strategies (Valavi *et al.*, 2022).

The study utilized the MaxEnt model running in R Studio 4.4.1 to predict suitable habitats and potential distribution of the species C. balansae), serving as a basis for identifying priority areas for conservation and ecological restoration. The results from this model will support effective conservation planning, ensuring the protection and maintenance of this rare species in the wild (Phillips *et al.*, 2017; Zhu & Peterson, 2017.

The Area Under the Curve (AUC) is used to evaluate the performance of the model. The AUC value ranges from 0 to 1, with a value of 0.5 indicating that the model has no better predictive ability than random guessing. Conversely, the closer the AUC value is to 1, the higher the model's accuracy and its ability to clearly distinguish between areas with and without the species' presence, reflecting the good predictive performance of the distribution model (Phillips *et al.*, 2006; Peterson *et al.*, 2011; Yang *et al.*, 2023).

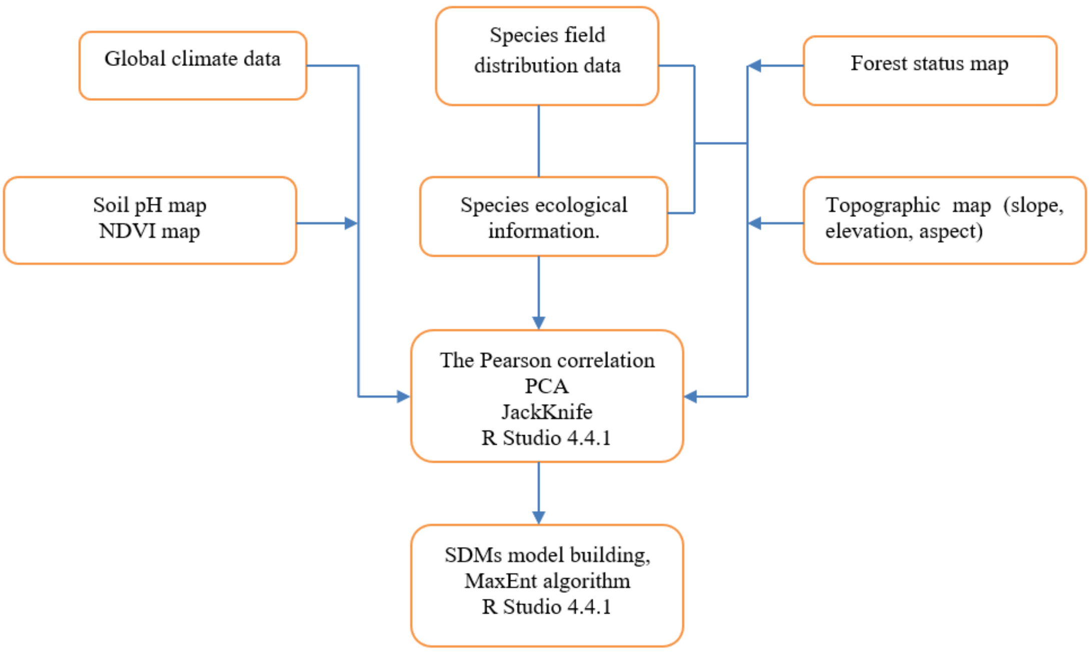


Figure 2: Summary of the approach to modeling Species Distribution Models (SDMs)

III. Results and Discussion

3.1. Data Set Construction

Topographic data were downloaded from WorldClim with a resolution of 30s (~1 km²), compatible with other climate and weather variables. These topographic variables were derived from the SRTM digital elevation model (Fick & Hijmans, 2017). The link to download the SRTM elevation data is:
https://geodata.ucdavis.edu/climate/worldclim/2_1/base/wc2.1_30s_elev.zip

The NDVI index is calculated from Sentinel-2A satellite images and adjusted to a 30s resolution to match climate variables. NDVI is a crucial index in assessing forest cover and ecosystem stability, providing information on forest density changes. This index helps in monitoring changes in vegetation density, which is vital for constructing species distribution models for rare woody species (Zhu *et al.*, 2022; Zeng *et al.*, 2020).

Soil pH and forest cover status are indeed critical factors in species distribution modeling (SDM). Soil pH, which can be downloaded and processed using Google Earth Engine (GEE), directly influences plant growth and species distribution models. Similarly, forest cover status, derived from high-resolution Sentinel-2A satellite imagery, provides essential data for predicting species distribution based on habitat presence and variability (Heng *et al.*, 2017).

The synthesized results of topographic variables, NDVI, soil pH, and the current forest cover status of Tà Đùng National Park are shown in **Figure 3**.

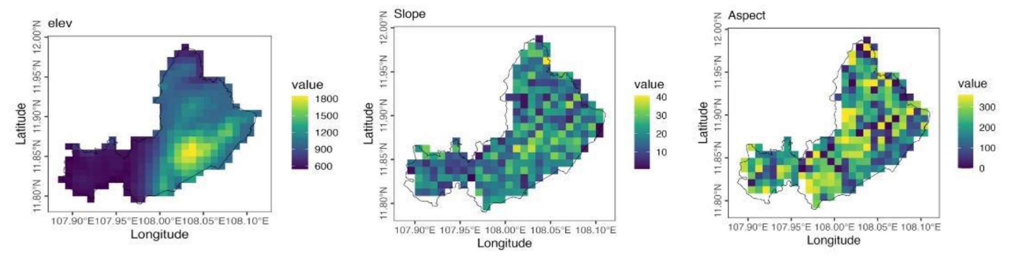
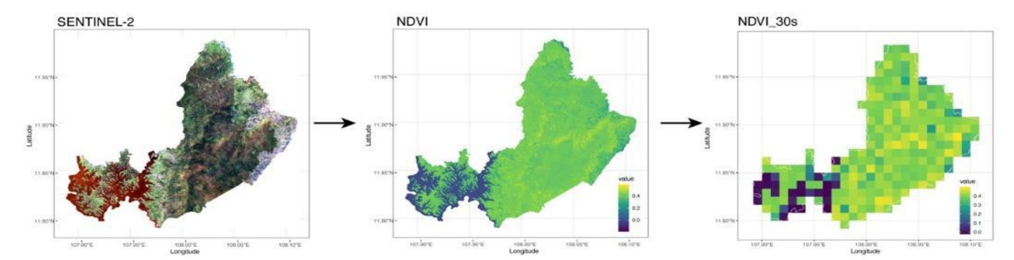
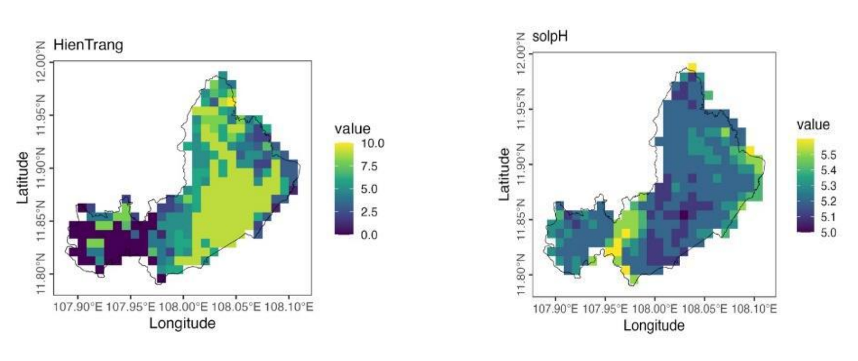
Topographic data (slope, elevation, aspect)	
NDVI map	
Soil pH map and forest cover status map (Hientrang)	

Figure 3: Data on topographic variables, NDVI, soil pH, and current forest cover status.

Data on 19 climatic variables, including temperature, precipitation, humidity, and other factors, were downloaded from WorldClim version 2.1, a widely used global climate data source in ecological and geographical research (Fick & Hijmans, 2017). This data represents the period from 1970 to 2000 and has a spatial resolution of 30 seconds (~1 km²). These climatic variables provide essential information for species distribution modeling based on climatic conditions in different areas. **Figures 4 and 5** illustrate the 19 climatic variables for the Ta Dung National Park area. This dataset can be downloaded from the following web address: https://geodata.ucdavis.edu/climate/worldclim/2_1/base/wc2.1_30s_bio.zip

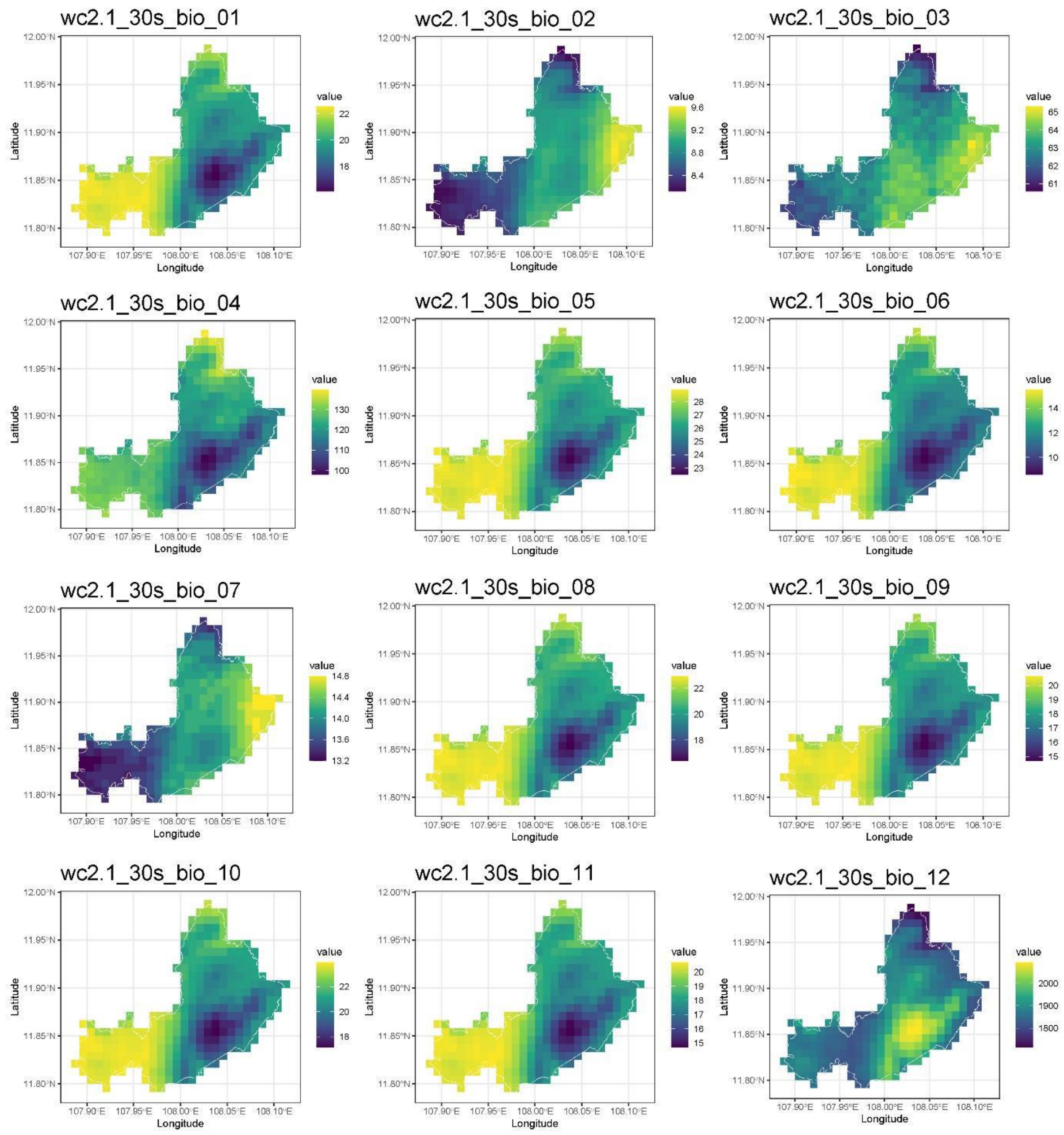


Figure 4: Climate and Weather Data from WorldClim at Ta Dung National Park (Bio1 to Bio12)

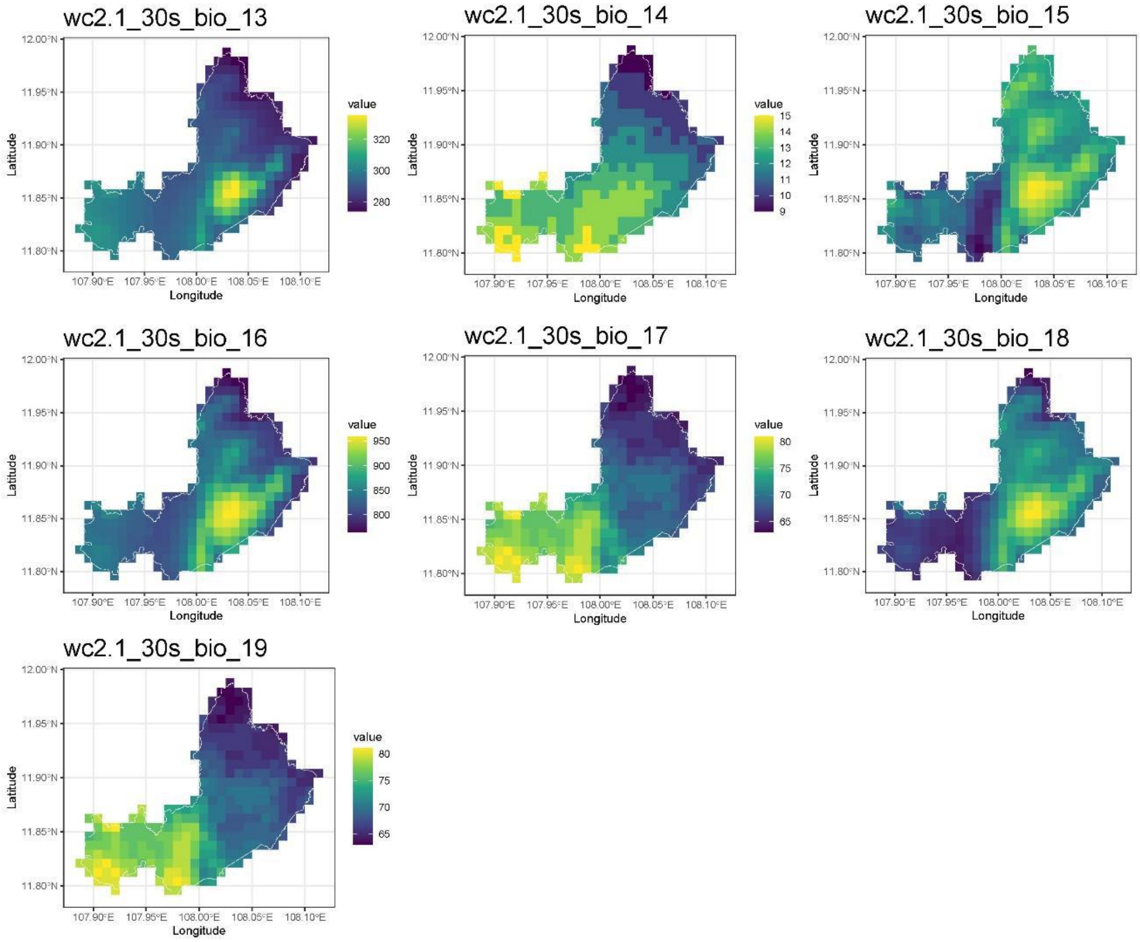


Figure 5: Climate and Weather Data from WorldClim at Ta Dung National Park (Bio13 to Bio19)

Thus, a total of 25 data layers were downloaded and processed to support the construction of distribution SDMs. The combination of these variables provides a comprehensive view of species habitats, thereby improving the accuracy of species distribution models. Additionally, it enhances the ability to predict ecological distributions based on different environmental conditions (Elith & Leathwick, 2009; Franklin, 2010).

3.2. Pearson correlation matrix, Principal Component Analysis (PCA) among variables, and assessment of the contribution of variables to the construction of species distribution models

In the field of GIS and remote sensing research, data often includes multiple variables with complex interrelationships. Therefore, identifying and controlling the correlations between variables is a crucial step to ensure the accuracy of subsequent analyses. In this study, Pearson correlation analysis is used to assess the linear correlation between variables, helping to identify and eliminate multicollinearity, thereby ensuring the stability and reliability of the model (Sedgwick, 2022).

Additionally, Principal Component Analysis (PCA) is a powerful tool for dimensionality reduction, eliminating unnecessary noise while retaining the most important variables, thereby significantly improving the analysis and computation process (Jolliffe & Cadima, 2016). PCA allows for the optimization of input data by extracting the principal components that contain the most information, enhancing the efficiency of complex multivariate analyses in GIS and remote sensing (Abdi & Williams, 2010; Ringnér, 2021).

Both methods, Pearson correlation analysis and PCA, play essential roles in reducing the complexity of spatial data, helping researchers build effective predictive models and draw more accurate scientific conclusions from multivariate datasets (Schober *et al.*, 2018; Jolliffe & Cadima, 2016; Ringnér, 2021).

The study conducted Pearson correlation and PCA analyses in R Studio 4.4.1, with the results presented in **Figure 6**.

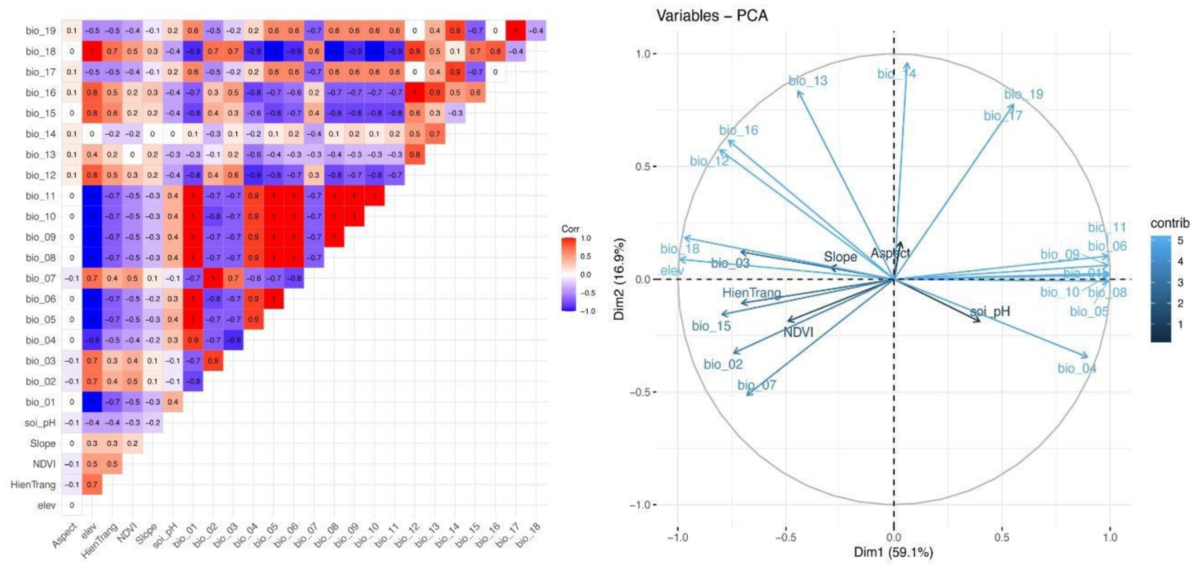


Figure 6: Results of Pearson Correlation and PCA Analysis

Thus, from the initial 25 variables, after conducting Pearson correlation and PCA analyses, the study selected 16 variables to assess their contribution to the model. These variables include Climate variables (Bio19, Bio16, Bio14, Bio13, Bio12, Bio07, Bio04, Bio03, Bio02, Bio01); topographic variables (slope, elevation, aspect); soil pH; forest cover status; and NDVI.

Assessing the contribution of variables is a crucial factor in optimizing the performance and reliability of the model. This process helps identify variables with high statistical significance and substantial impact on prediction outcomes, while eliminating less important or low-contributing variables. This not only reduces complexity but also enhances the accuracy and generalizability of the SDM, thereby improving computational efficiency (Dormann *et al.*, 2013).

In this study, the Jackknife method was used to assess the importance of each variable. Specifically, this method examines the increase in model value when a single variable is added and evaluates the decrease in significance when that variable is removed. Jackknife analysis helps clarify the impact of each variable on overall performance, while optimizing key factors to enhance the accuracy and predictive capability of the SDMs (Phillips *et al.*, 2006). The detailed results of the Jackknife analysis were calculated using R Studio 4.4.1 and are presented in **Table 1**.

Table 1: Analysis results of the impact of model variables on the distribution of *Cinnamomum balansae*

Variable	**PC%**	**PI%**	**AUC INCREASES**	**AUC DECREASES**
Bio_04	38,7	56,2	U	
Aspect	16,7	1,8		
HienTrang	11,0	0,0		
Bio_16	6,5	0,0		
Bio_07	5,6	17,2		
Bio_13	4,9	20,9		**D**
Slope	4,6	2,5		
Bio_12	3,9	0,0		
Bio_03	3,8	0,0		
Bio_01	2,8	0,1		
NDVI	0,9	0,0		
Soil pH	0,5	1,3		
Bio_02	0,0	0,0		
Bio_14	0,0	0,0		
Bio_19	0,0	0,0		
Elevation	0,0	0,0		

Note: PC: Percent Contribution; PI: Permutation Importance; U: AUC increases; D: AUC decreases; Hientrang: Current Forest cover status

Among the variables used, Bio04 (temperature seasonality) is the main contributor to the species distribution model, accounting for 38.7% of the total contribution. This is followed by Aspect and Forest Cover Status, with contributions of 16.7% and 11%, respectively. The permutation importance analysis shows that Bio04 has the greatest impact, accounting for 56% of the total importance, followed by Bio13 (precipitation of the wettest month) with 21%, and Bio07 (annual temperature range) with 17%.

The analysis results indicate that eight variables play a crucial role in determining the distribution of C. balansae in the study area. These variables include: Bio01 (annual mean temperature), Bio04, Bio07, Bio13, NDVI, forest cover status, slope, and elevation. These variables were selected as inputs for running the spatial distribution simulation of the species.

3.3. Developing current and potential distribution maps of *Cinnamomum balansae* at Ta Dung National Park

This study applied a SDMs to predict the distribution of *Cinnamomum balansae*, based on 70% of the current distribution points to train the model with eight key variables. These include four climatic variables, two topographic variables (slope and elevation), and two variables related to forest cover status and NDVI. The model was trained through 700 iterations and used cross-validation for the remaining 30% of distribution points. The study also established 900m² sample plots and recorded a total of 72 occurrence points of *Cinnamomum balansae* during field surveys in the study area. Of these, 48 points were used to train the model and 24 points to test and evaluate the model. This data division ensures the objectivity and accuracy of the predictions while improving the model's generalization capability (Merow *et al.*, 2013; Muscarella *et al.*, 2014).

The predicted distribution of C. balansae at Ta Dung National Park, using the MaxEnt algorithm in R Studio 4.4.1, is presented in **Figure 7**. This figure illustrates the current and potential distribution areas of C. balansae based on the MaxEnt model. The prediction values range from 0 to 1, with colors from light green to dark green indicating the potential occurrence of the species. Areas with higher prediction values (dark green) indicate a higher likelihood of species presence. The MaxEnt model achieved high accuracy with an AUC (Area Under the Curve) value of 0.86 **(Figure 8**), confirming the model's reliable classification capability (Merow *et al.*, 2013; Muscarella *et al.*, 2014). Although *Cinnamomum balansae* is a rare and valuable timber species with limited natural presence, the MaxEnt model still demonstrated high predictive performance with an accuracy of 86%. This result is consistent with previous studies, including those by Bente Støa and Rune Halvorsen (2021), who also recognized MaxEnt as a powerful tool for predicting the distribution of rare timber species.

Additionally, the results of this study demonstrate that the MaxEnt model can perform effectively even in limited study areas, such as a national park. This is consistent with previous studies, including those by Catherine E. Burns, 2003, Pradeep Adhikari *et al.*,2018), where the authors recognized the capability of MaxEnt in predicting the distribution of rare species in protected areas**.**

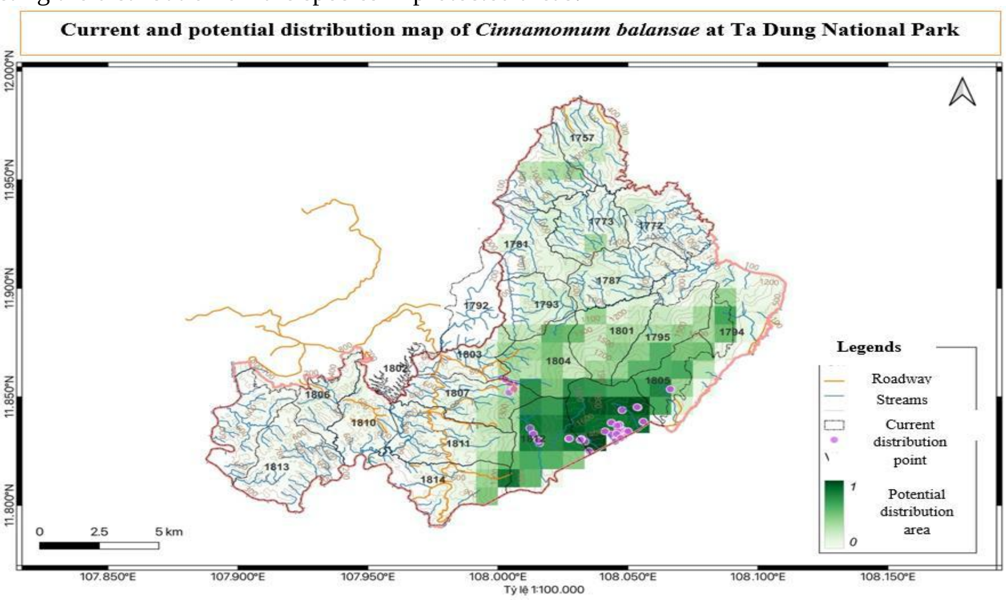


Figure 7: Current and potential distribution map of *Cinnamomum balansae* in Ta Dung National Park

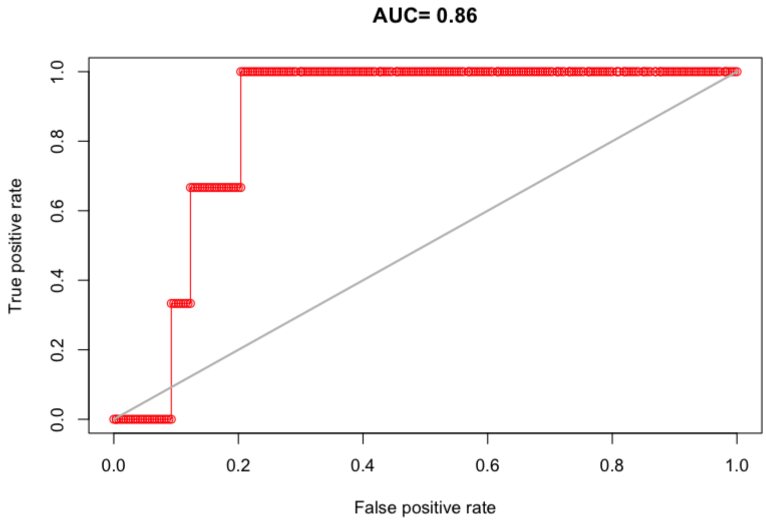


Figure 8: AUC Value

IV. Conclusion

This study indicates that the environmental adaptation of *Cinnamomum balansae* is significantly influenced by climatic factors such as Bio04 (temperature seasonality), Bio01 (annual mean temperature), Bio07 (annual temperature range), and Bio13 (precipitation of the wettest month). Additionally, ecological factors such as forest cover status, NDVI index, slope, and elevation also play crucial roles in determining the species' distribution. These factors interact in complex ways, directly impacting the species' habitat and affecting their survival and development in the wild (Elith & Leathwick, 2009; Franklin, 2010). Further research is needed to explore the effects of temperature and ecological factors on the growth and development of *Cinnamomum balansae* at different growth stages.

In Ta Dung National Park, Dak Nong, Vietnam, *Cinnamomum balansae* primarily distributes in evergreen forests at elevations above 1,400 meters. The current and potential distribution model of the species, constructed using the MaxEnt algorithm with an AUC value of 86%, provides a solid scientific basis for identifying priority areas for conservation and habitat management. Based on the model results, areas with high potential distribution should be prioritized for the establishment of strict conservation zones to protect the species from negative environmental impacts and human activities (Merow *et al.*, 2013; Zhu *et al.*, 2022).

Future studies can expand the application of the MaxEnt model to evaluate the accuracy of predicting the distribution of other rare timber species in Ta Dung National Park. This not only helps verify the model's applicability but also contributes to improving conservation strategies for valuable plant species. Furthermore, results from SDMs in this study can be used to analyze and predict habitat changes for *Cinnamomum balansae* under future climate change scenarios, aiding in the development of adaptation and protection measures for the species in the increasingly complex context of climate change (Elith *et al.*, 2022; Qiao *et al.*, 2023).

The application of remote sensing technology combined with field observations will enable monitoring of habitat changes and the impact of climate change on *Cinnamomum balansae* populations. Additionally, the involvement of local communities in habitat protection and management activities will play a crucial role in conservation efforts. Collaboration with international conservation organizations and scientific research will also contribute to enhancing the effectiveness of species conservation. These factors form the basis for developing effective and sustainable conservation measures in the future (Hao *et al.*, 2023; Muscarella *et al.*, 2014)

Acknowledgements

The author expresses gratitude to the leadership of Tay Nguyen University, the leadership of the Faculty of Agriculture and Forestry, and especially the RIM research group at Tay Nguyen University for their support and assistance during this research. Furthermore, the author extends special thanks to the individuals and Ta Dung National Park for their support during the field surveys and the execution of this research.

The author also expresses sincere gratitude to the Ministry of Education and Training of Vietnam for funding this project.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of interest: The authors declare no conflict of interest

REFERENCES

[1]. Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: *Computational Statistics*, 2(4), 433–459.

[2]. Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., et al. (2020). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaa6498.

[3]. Burns, C. E., Johnston, K. M., & Schmitz, O. J. (2003). Global climate change and mammalian species diversity in U.S. national parks. *Proceedings of the National Academy of Sciences*, 100(16), 10345-10350.

[4]. Duy, V. D., et al. (2021). Evaluation genetic diversity and a bottleneck of Cinnamomum balansae H. Lecomte populations in Phu Tho province by microsatellite (SSR) markers. *Journal of Forestry Science and Technology*, 1, 3–9.

[5]. Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, *Evolution, and Systematics*, 40, 677-697.

[6]. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2022). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43-57.

[7]. Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302-4315.

[8]. Franklin, J. (2010). Mapping species distributions: Spatial inference and prediction. *Cambridge University Press.*

[9]. Franklin, J., Miller, J. A., Menzel, A., & Thorne, J. H. (2023). Advances in species distribution modeling. *Annual Review of Environment and Resources*, 48, 365-391.

[10]. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.

[11]. Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat suitability and distribution models: With applications in R. *Cambridge University Press*.

[12]. Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850-853.

[13]. Hai, T. N., Nghi, D. H., Phuong, L. D., & Hoang, T. V. (2016). Study on the sylviculture characters of Cinnamomum balansae Lecomte at Ben En national park. *Journal of Forestry Science and Technology*, 6, 176–185.

[14]. Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., & Shangguan, W. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748.

[15]. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965-1978.

[16]. IUCN (2023). Cinnamomum balansae. *The IUCN Red List of Threatened Species*, 2023.

[17]. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, *Physical and Engineering Sciences*, 374(2065), 20150202.

[18]. Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.

[19]. Li, X., Shen, H., & Zhang, L. (2023). Advanced vegetation indices for remote sensing applications: A review of concepts and new developments. Remote Sensing of Environment, 289, 113052.

[20]. Liu, C., Newell, G., & White, M. (2022). On the selection of species distribution modeling algorithms for ecological niche modeling. *Ecological Modelling*, 472, 110052.

[21]. Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.

[22]. Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

[23]. Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., et al. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecology*, 100(3), e02558.
[24]. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887-893.
[25]. Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259.
[26]. Ringnér, M. (2021). What is principal component analysis? *Nature Biotechnology*, 40, 1404–1406.
[27]. Sedgwick, P. (2022). Pearson's correlation coefficient. *BMJ*, 377, o1538.
[28]. Støa, B., Halvorsen, R., et al. (2021). Evaluating the performance of species distribution models for rare species. *Ecological Modelling*, 456, 109667
[29]. Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6(1), 35–39.
[30]. Thuiller, W., et al. (2019). Ensemble forecasting of species distributions: The consensus approach. *Ecography*, 32(3), 369-373.
[31]. Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2022). A comparison of MaxEnt and boosted regression trees for modelling species distribution using presence-only data. *Methods in Ecology and Evolution*, 13(1), 37-49.
[32]. Wis, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., et al. (2020). The role of biotic interactions in shaping distributions and realized assemblages of species: Implications for species distribution modelling. *Biological Reviews*, 95(4), 1157-1173.
[33]. Xu, T., Liu, Z., Sun, Z., Zhang, S., & Wang, Y. (2022). Recent developments in high-resolution global land cover mapping: A review. *Remote Sensing*, 14(15), 3572.
[34]. Zeng, Y., Gao, S., & Liu, Y. (2020). Assessment of NDVI time-series data for monitoring vegetation dynamics and land degradation in arid regions. *Sustainability*, 12(4), 1650.
[35]. Zhang, Z., Tang, H., & Zhu, X. (2022). Integrating remote sensing and machine learning to predict potential distributions of endangered plant species in a changing climate. *Remote Sensing*, 14(3), 712.
[36]. Zhu, G., & Peterson, A. T. (2017). Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. *Biological Invasions*, 19, 2519-2532.
[37]. Zhu, X., Liu, D., & Tang, H. (2022). Monitoring forest dynamics using multi-source remote sensing data: Applications of NDVI and machine learning. *Remote Sensing*, 14(2), 415.
[38]. Zurell, D., Franklin, J., König, C., Bouchet, P. J., Serrano, D., Cabral, J. S., et al. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261-1277.
[39]. Qiao, H., Soberón, J., & Peterson, A. T. (2023). No silver bullets in correlative ecological niche modeling: insights from testing among many potential algorithms for niche estimation. *Global Ecology and Biogeography*, 32(2), 228-241.

