FOLIA OECOLOGICA – vol. 53, no. 2 (2026), doi: 10.2478/foecol-2026-0010

A probabilistic habitat–suitability overlap framework (*HMI*) reveals spatial bias in MaxEnt models of West African forest butterflies

Fabio Petrozzi[1], **Luca Luiselli**[2,3,4,*]

[1]OrViCa international, Rome, Italy
[2]Department of Environmental and Applied Biology, Rivers State University, Port Harcourt, Nigeria
[3]Laboratory of Ecology and Ecotoxicology, Faculty of Sciences, University of Lome, Lomé, Togo
[4]Institute for Development, Ecology, Conservation and Cooperation, Rome, Italy

Abstract

Petrozzi, F., Luiselli,L., 2026. A probabilistic habitat–suitability overlap framework (*HMI*) reveals spatial bias in MaxEnt models of West African forest butterflies. *Folia Oecologica*, 53 (2): 108–122.

Species distribution models (SDMs) are widely used to infer potential species ranges, yet their ecological reliability in rapidly transformed landscapes remains poorly understood. We applied Maximum Entropy (MaxEnt) modelling to estimate the potential distribution of twelve forest-specialist butterfly species of high commercial value across West Africa, with a focus on Côte d'Ivoire, Ghana and Liberia, based on bioclimatic covariates. Models showed good predicted performance (mean jackknife AUC = 0.854, significantly better than random; null model test $p < 0.01$), driven by both precipitation-related variables (i.e. precipitation seasonality; 57.4% of explained variance in model predictions) and temperature-related variables (annual mean temperature, 33%). However, spatial predictions showed limited concordance with contemporary forest cover: 45.23% in Liberia but only 4.69% in Cote d'Ivoire and 6.42% in Ghana. Key refugia such as Taï Forest (Cote d'Ivoire) were partially underpredicted. To evaluate whether the observed agreement between predicted habitat suitability and forest cover reflects genuine ecological associations or is instead driven by spatial bias, we developed a probabilistic framework in which the expected overlap is defined as the product of forest availability and suitability probability. Building on this null expectation, we derived a Habitat Matching Index (*HMI*), expressed as the ratio between observed and expected overlap, allowing us to quantify deviations from random spatial association and to distinguish ecological signal from patterns generated by habitat availability alone. *HMI* values differed markedly among countries (Côte d'Ivoire: 0.319; Ghana: 0.570; Liberia: 1.407), demonstrating that apparent model performance is strongly influenced by forest extent rather than ecological fidelity alone. These findings reveal a fundamental limitation of climate-based SDMs: high statistical accuracy does not necessarily imply ecological realism. Integrating habitat availability into model evaluation provides a critical correction, improving inference in heterogeneous and rapidly changing tropical landscapes. We advocate combining climatic, land-cover, and species-specific data to enhance SDM reliability in conservation biogeography.

Keywords

habitat matching index, probabilistic overlap model, spatial bias, species distribution modelling, tropical forests

Introduction

Species distribution modelling (SDM) provides a robust analytical framework for estimating potential species ranges using limited occurrence data (Elith and Leathwick, 2009). SDMs are sufficiently flexible to model both individual species distributions and entire communities. This latter approach, which has been used for instance

---

*Corresponding author:
e-mail: l.luiselli@ideccngo.org

Institute of Forest Ecology
Slovak Academy of Sciences

by Ferrier and Guisan (2006) and others, highlights the capacity of SDMs to capture broader ecological patterns. Among available SDM approaches, the Maximum Entropy algorithm (MaxEnt) (Phillips et al., 2017; Phillips and Dudík, 2008) is widely regarded for its robustness and high predictive performance, particularly when applied to presence-background data and small sample sizes (Merow et al., 2013). Owing to its ability to integrate species occurrence records with multiple environmental predictors, MaxEnt has become a standard tool in ecological research for estimating habitat suitability across heterogeneous landscapes. Moreover, it is extensively used to forecast potential range shifts under future climate-change scenarios (Khanum et al., 2013; Li et al., 2020; Zhao et al., 2021). Model performance in MaxEnt is typically assessed using statistical metrics such as the Area Under the Receiver Operating Characteristic Curve (AUC), the True Skill Statistic (TSS), omission rates, and cross-validation procedures (Fourcade et al., 2014). Although these metrics are statistical, their interpretation often extends into the biological domain. High discriminatory ability and low omission rates suggest that the selected environmental predictors capture ecologically meaningful constraints on species' realized niches (Warren and Seifert, 2011). Additionally, response curves, permutation importance, and jackknife tests enable inference regarding the relative influence of environmental variables. Consequently, MaxEnt evaluation occupies a hybrid methodological space in which statistical diagnostics both quantify predictive accuracy and inform ecological and biogeographical hypotheses. Nevertheless, the reliability of predictions at fine spatial scales remains debated, particularly in rapidly changing landscapes where presence-only data may be sparse, outdated, or spatially biased, thereby compromising model performance.

For rare, declining, or highly vulnerable species, accurate SDMs are essential for effective conservation planning. Even minor predictive errors may have disproportionately large consequences when species occupy restricted or fragmented habitats. Overestimation of suitable habitat (false positives) may result in inefficient allocation of conservation resources, whereas underestimation (false negatives) risks leaving critical habitats unprotected (Guisan et al., 2013). Narrow-range, micro-endemic, and habitat-specialist species are especially sensitive to such errors, as coarse-resolution environmental data and limited occurrence records increase predictive uncertainty (Giachello et al., 2025; Tessarolo et al., 2021). Because conservation decisions for these taxa are often irreversible, high predictive precision is essential.

Importantly, SDM outputs may be influenced not only by ecological relationships but also by the spatial availability of suitable habitat. In highly heterogeneous landscapes, the probability that predicted high-suitability pixels overlap with actual habitat (e.g., forest) may depend on the relative abundance of that habitat within the study region. This introduces a potential bias whereby model "performance," when evaluated spatially, may partially reflect habitat availability rather than ecological accuracy. To address this issue, we introduce a probabilistic framework linking habitat availability and predicted suitability overlap. Specifically, if forest cover occupies a proportion F of the landscape, and high-suitability pixels are distributed independently of current forest cover, the expected overlap between suitability and forest is proportional to F. Deviations from this expectation can therefore be interpreted as evidence of ecological signal versus spatial randomness.

Butterflies play ecologically and economically important roles in tropical ecosystems, functioning as pollinators, bioindicators (Tsingalla et al., 2026), and components of the international entomological trade (Bonebrake et al., 2010; Small, 2007). In West Africa, butterfly diversity is particularly high in forest ecosystems (Larsen, 2008; Sáfián et al., 2020); however, ecological data remain scarce, especially in the Upper Guinean forest block (Fermon, 2002; Fermon et al., 2000), where rapid habitat loss poses a severe threat to biodiversity. Côte d'Ivoire, Ghana and Liberia were selected for this study based on our long-term field research on rainforest reptiles and insects (e.g. Ajong et al., 2024; Luiselli et al., 2025), which has provided detailed knowledge of forest habitat conditions. We focused on forest-specialist butterfly species that are apparently rare based on their high commercial value (Slone et al., 1997; Wang et al., 2023). In fact, in the absence of comprehensive population censuses, market value can serve as a proxy for rarity and conservation concern (New and Collins, 1991), although morphological traits such as wingspan, shape, and coloration explain approximately 49% of the variance in aesthetic ranking, which in turn predicts sales volume (Wang et al., 2023).

Building on this framework, the present study addresses five main research questions:

1. *Distribution:* What is the potential geographic distribution of twelve forest-specialist butterfly species in Côte d'Ivoire, Ghana and Liberia?
2. *Model performance:* How well do climate-driven MaxEnt models perform at landscape spatial scales, as evaluated by AUC, TSS, omission rates, and cross-validation?
3. *Ecological relevance and limitations:* To what extent do model predictions reflect ecological constraints, and what are the limitations of MaxEnt at fine spatial scales in rapidly changing habitats, particularly for African rainforest butterflies?
4. *Spatial bias assessment (probabilistic framework):* To what extent can the observed overlap between predicted suitability and forest cover be explained by habitat availability alone?
5. *Conservation implications:* How can integrating model predictions with species-specific environmental sensitivity inform conservation assessments for commercially valuable forest butterflies?

Materials and methods

Study area

Although the focal species occur in Côte d'Ivoire, Libe

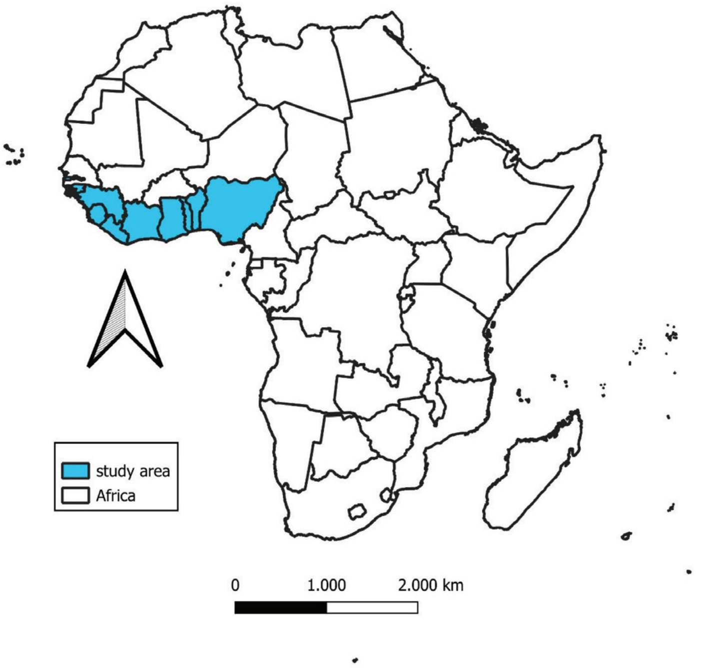

Fig. 1. Study area.

ria, and Ghana, MaxEnt models were developed across the broader West African region, extending from Guinea to Nigeria and including Côte d'Ivoire, Ghana, Liberia, and Sierra Leone, (Fig. 1). This approach enables a more comprehensive representation of environmental gradients, improving the accuracy of predicted distributions within the focal countries. The study area includes ecosystems ranging from coastal rainforests to savannah mosaics, characterized by pronounced seasonal variation in rainfall and temperature.

Target species

To ensure that only currently occurring species were included, we first reviewed recent scientific literature for West Africa. In particular, we relied on a doctoral thesis on Liberian butterflies (Sáfián, 2021), which provides detailed distributional and ecological data. Species reported from the study area were then cross-referenced with their market values (Table S1) derived from specialized online platforms. A total of 50 species of butterflies distributed in forest environments of West Africa were searched on the eBay platform and other sites specialized in sales for entomological collections (research carried out in November 2024), selecting for our analyses only those with a price > 60 US dollars at the search time. A total of twelve forest-associated butterfly species of high commercial value were selected (Table S1).

For each selected species, we compiled information on habitat, data availability, and ecological notes (Table 1). All species occupy broadly similar habitats, and differences in habitat descriptions largely reflect variation in terminology among sources rather than true ecological divergence. Indeed, all records refer to moist rainforest patches within the Upper Guinean forest block (Chatelain et al., 2004; Freeman et al., 2021) (Fig. S1).

On this basis, pooling species into a single modelling framework is justified (see below). Table 1 also indicates the expected reliability of MaxEnt predictions, distinguishing between species with well-documented ecological profiles and those with limited ecological information. Species such as *Charaxes fournierae*, *Charaxes smaragdalis*, and *Papilio zalmoxis* are considered robust candidates for modelling, whereas species such as *Papilio horribilis*, *Abantis lucretia*, *Apallaga ankasa*, and *Pteroteinon pruna* require cautious interpretation (Table 1). This classification informed model parameterization, evaluation, and interpretation. By explicitly incorporating variability in data quality, we aimed to account for uncertainty and ensure that predictions were appropriately contextualized.

Data collection and optimized MaxEnt modelling

Occurrence data for the various species (n = 72 records) were obtained from the Global Biodiversity Information Facility (GBIF) (Telenius, 2011), limited to records post-1980 to ensure recent distributional relevance, and from field verified sources. Using the "coordinateUncertaintyinMeters" in R, we calculated the error in the coordinates with a 1,000 m threshold, and only 2.8% of the records were outside this threshold.

To minimize potential taxonomic errors, only records identified to the species level using accepted scientific names were retained. A conservative data selection strategy was adopted, including exclusively records associated with specimens housed in curated collections, where taxonomic verification is possible, while excluding other sources of occurrence data (https://training.gbif.org/en/data-use/taxonomic-uncertainty, last accessed on 12th De-

Table 1. Summary of the twelve forest associated butterfly species included in this study. For each species, the number of used records (indicated after the scientific name), known habitat, data availability, and relevant notes are provided. The "Modeling Implications" column indicates the expected reliability of MaxEnt predictions based on the quality of existing ecological data, highlighting which species are robust candidates for species distribution modeling and which are considered less known, requiring cautious interpretation and potential field validation.

Species	Known Habitat	Data Availability	Notes	Modeling Implications
Charaxes fournierae (n = 2)	Lowland tropical moist broadleaf forests	Good	Forest-dependent; sensitive to deforestation (Henning, 1988)	Suitable for MaxEnt modeling; predictions likely reliable
Charaxes mycerina (n = 2)	Likely forest habitats (inferred from Poor genus)	Poor	Ecological profile entative; requires field verification (Henning, 1988)	Predictions uncertain; results should be interpreted cautiously
Charaxes nichetes (n = 8)	Forests, woodland savanna, riparian areas	Moderate	Occurs near watercourses; forest-associated) (Henning, 1988)	Good confidence in model outputs but field validation recommended
Charaxes petersi (n = 6)	Likely forest-associated	Moderate	Occurs near watercourses; forest-associated (Henning, 1988)	Good confidence in model outputs but field validation recommended
Charaxes smaragdalis (n = 5)	Lowland rainforest, gallery forest, forest patches	Good	Forest-dependent; sensitive to habitat loss (Henning, 1988)	Suitable for MaxEnt modeling; predictions likely reliable
Charaxes subornatus (n = 5)	Likely evergreen or moist forest	Poor	Habitat tentative; data limited (Henning, 1988)	Good confidence in model outputs but field validation recommended
Graphium latreillianus (n = 17)	Forested zones; primary forests and wet-forest mosaics	Poor	Specific habitat information scarce (https://www.metamorphosis.org.za/articlesPDF/1040/014%20Genus%20Graphium%20Scopoli.pdf; accessed on 30th November 2025)	Predictions uncertain; field validation recommended
Papilio zalmoxis (n = 3)	Humid tropical forests	Good	Large forest swallowtail; forest-dependent (West and Hazel, 1979)	Suitable for MaxEnt modeling; predictions likely reliable
Papilio horribilis (n = 19)	Upper Guinean endemic; wet forest	Poor	Wet forest species (https://abdb-africa.org/species/Papilio_horribilis; accessed on 30th November 2025), but with no detailed habitat information available	Modeling uncertain; results should be interpreted cautiously
Abantis lucretia (n = 1)	Unspecified forest	Poor	Forest species, but with no detailed habitat information available (https://web.archive.org/web/20131203030759/http://atbutterflies.com/downloads/hesperiidae_pyrginae.doc, accessed 30th November 2025)	Modeling uncertain; results should be interpreted cautiously

Table 1 – Continued

Species	Known Habitat	Data Availability	Notes	Modeling Implications
pallaga ankasa (n = 2)	Unspecified forest	Poor	Forest species, but with no detailed habitat information available (Sáfián , 2021; Sáfián et al., 2011)	Modeling uncertain; results should be interpreted cautiously
teroteinon pruna (n = 1)	Riverine forest	Poor	Riverine forest species (https://en.wikipedia.org/wiki/Pteroteinon_pruna#:~:text=Pteroteinon%20pruna%2C%20commonly%20known%20as%20Evans%27%20red-eye%2C%20is,Uganda.%20%5B3%5D%20The%20habitat%20consists%20of%20riverine%20forests; accessed on 30th November 2025).	Modeling uncertain; results should be interpreted cautiously

cember 2025). The resulting species-specific datasets were then merged into a consolidated .csv file, standardizing the species variable as "community" (see Supplementary dataset 1). The aggregation of occurrence records from multiple target species to develop a single, comprehensive MaxEnt model is justified on the following grounds: (i) all species occupy the same habitat type and display broadly congruent spatial distributions (Carcasson, 2011); and (ii) this "aggregated" modelling strategy has been previously implemented and validated in comparable studies (Doser et al., 2023). By using RStudio (R version 4.3.1) dismo package (Hijmans et al., 2017) the initial dataset was also inspected to identify spatial duplicates.

To ensure data quality and reduce spatial sampling bias, we visually inspected occurrence data for clustering and assessed spatial autocorrelation. Many species displayed clusters of identical or near-identical coordinates, indicative of sampling concentration at specific localities. To address this, we implemented spatial thinning using the spThin R package (Aiello-Lammens et al., 2015) (Table S2), applying a minimum nearest-neighbor distance threshold of 5 km. This procedure removed spatially redundant points while preserving geographic representativeness, thereby minimizing potential model overfitting due to spatial autocorrelation. To further verify the spatial distribution of occurrences and the absence of sampling bias, we mapped all filtered presence points against the study area boundary and environmental layers. These distribution maps confirmed adequate spatial coverage and informed subsequent ecological niche modeling. The filtering and visualization steps ensured that the occurrence data met best practices for species distribution modeling (Boria et al., 2014; Fourcade et al., 2014). After removing spatial duplicates and applying spatial thinning 58 unique, spatially validated occurrence points remained (Table S2). Background points were likewise restricted to the study area.

The environmental variables (initially n = 19) were obtained from WorldClim (Fick and Hijmans, 2017) at a landscape to regional spatial scale with a spatial resolution of 2.5 arc-minutes (Table S3) prior to applying the niche modelling procedure. We used WorldClim climate layers due to their widespread application in species distribution modelling, which ensures methodological comparability and reproducibility across studies, while maintaining consistency with the broader ecological literature.

All environmental layers were clipped to the African continent and resampled to a uniform spatial extent. After processing the bioclimatic raster layers and clipping them to the African continent, discrepancies in spatial resolution and extent could arise as a result of the clipping procedure. To ensure consistency among variables, all layers were therefore resampled to a common spatial resolution and extent. Multicollinearity among variables was evaluated through Pearson correlation analysis ($|r| > 0.7$ threshold; Bombi et al., 2011). Seven variables were retained for model development. A sequential selection procedure was then applied, with a maximum of five predictors retained. Because the number of uncorrelated variables exceeded this threshold, further filtering was performed based on percentage contribution derived from a preliminary MaxEnt model (Phillips et al., 2006), trained using all presence points and 1,000 randomly sampled background points. The five variables with the highest contributions were selected for subsequent analyses.

The optimized MaxEnt model (version 3.4.1; Phillips et al., 2017) was implemented in R (R Core Team, 2026) using the ENMeval package (Muscarella et al., 2014). Model tuning involved testing multiple combinations of regularization multipliers (0.5, 1, 1.5, 2, 3, 4) and feature classes (L, LQ, H, LQH, LQHP). For each model run, 10,000 background points were randomly sampled within the study area. Given the relatively high number of occurrence records (n = 58), a jackknife (leave-one-out cross-validation) approach was applied. The optimal model was selected based on AUC and the minimum AICc (Akaike Information Criterion corrected for small sample sizes), and its performance was summarized using the mean jackknife AUC. To assess whether the observed AUC significantly exceeded random expectations, a null model approach was implemented (Raes and Ter Steege, 2007). One hundred random pseudo-presence datasets (each with the same number of points as the number of records) were generated within the same study area and environmental space. For each replicate, a Maxent model was trained using the same regularization multiplier (L) and feature class (LQ) identified from the observed data, and evaluated via jackknife cross validation.

ENFA modelling

As an alternative approach based on presence data alone, an Ecological Niche Factor Analysis (ENFA) (Hirzel et al., 2002) was performed using the "ecospat" package (Di Cola et al., 2017). The records and covariates used were those of the MaxEnt model. The marginality and specialization factors were extracted from the five environmental variables and used to generate a suitability map. The suitability map was generated using the first marginality factor (which represents the deviation of presence conditions from the environmental mean of the study area). The predictive quality of the ENFA was assessed using: (i) AUC (comparing presence values with 1,000 random background points); and (ii) Corrected Boyce Index (Hirzel et al., 2006), calculated with ecospat.boyce, with 10 classes.

Forest Overlap

By using R Studio, we clipped both the MaxEnt model raster and the JRC Global Map of Forest Cover 2020 raster (Bourgoin et al., 2025), which represents forest cover, along the borders of Liberia, Côte d'Ivoire, and Ghana. We divided habitat suitability values into four classes: low (0.2–0.4), medium (0.4–0.6), high (0.6–0.8), and very high (0.8–1.0). Accordingly, when selecting areas with the highest likelihood of species' presence, we focused on those with suitability values equal to or greater than 0.8. Henceforth, for each country, we overlaid the two different rasters to see the percentage overlap between forest cover and very high suitability (>0.8) and the overlap between highly suitable areas and total forest extent. Lastly, we found out what the percentage of forests was in relation to the territory of each of the nations studied.

Quantifying habitat–suitability overlap: a probabilistic model

To evaluate whether the spatial agreement between predicted high suitability and forest cover reflects ecological processes or is driven by habitat availability alone, we developed a probabilistic model (Fig. 2). Let: (i) A = total number of pixels in a given country; (ii) F = number of forest pixels; (iii) S = number of pixels classified as highly suitable (suitability ≥ 0.8); and (iv) O = number of pixels where forest and high suitability overlap, we define:

- Forest proportion:

$$P_F = F/A$$

- Suitability proportion:

$$P_S = \mathrm{S/A}$$

- Observed overlap:

$$P_O = O/A.$$

Under the null hypothesis that suitability is spatially independent of forest cover (i.e., random with respect to habitat), the expected overlap is:

$$E_{[PO]} = P_F \cdot P_S.$$

To assess deviation from randomness, we define a Habitat Matching Index (*HMI*):

$$HMI = \frac{PO}{PS * PF}\,.$$

In this case the interpretation is as follows: $HMI = 1$: overlap explained by random habitat availability; $HMI > 1$: ecological association (model captures habitat dependence); $HMI < 1$: mismatch (model predicts suitability outside habitat).In order to compute *HMI*, we use: O/S = overlap between forest and highly suitable areas (our first

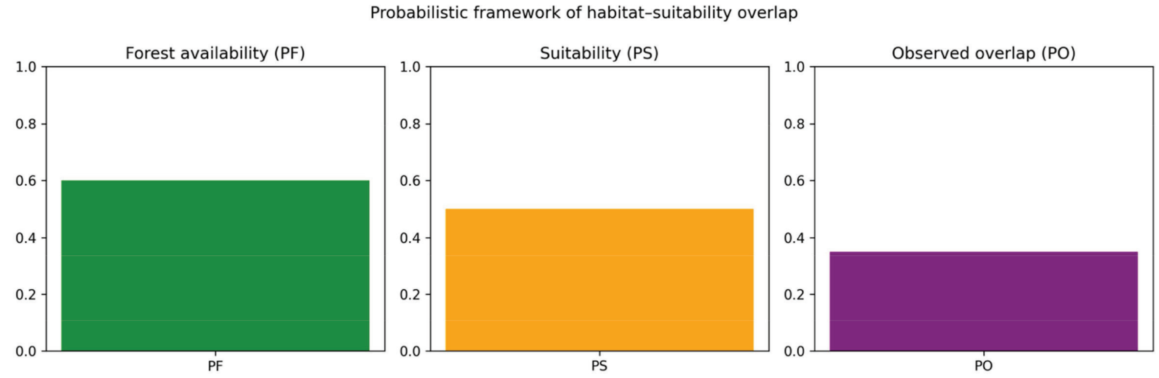

Fig. 2. Probabilistic framework for habitat–suitability overlap. Conceptual representation of the probabilistic model linking forest availability (P_F), habitat suitability probability (P_S), and observed overlap (P_O). Under a null assumption of spatial independence between forest cover and climatic suitability, expected overlap is given by $E[P_O] = P_F \cdot P_S$. Deviations from this expectation form the basis for evaluating non-random spatial association between predicted suitability and habitat distribution.

metric), $P_F = F/A$ = forest proportion (our third metric). From the model:

$$HMI = \frac{PO}{PS*PF}$$

But since: $$O/S = P_O/P_S$$

we can rewrite: $$HMI = \frac{O/S}{PF}.$$

Additionally, to explicitly test the hypothesis that higher overlap in each country of the study area is driven by greater forest availability, we model:

$$P_O = \beta_0 + \beta_1 P_F + \epsilon,$$

where β_1 captures the dependence of overlap on forest availability. This formulation allows generalization across regions and provides a diagnostic tool for SDM users to disentangle ecological signal from spatial sampling effects.

Results

Of the various environmental variables, the following bioclimatic variables were selected: bio 1 (Annual Mean Temperature), bio 2 (Mean Diurnal Range), bio 12 (Annual Precipitation), bio 14 (Precipitation of Driest Month), and bio 15 (Precipitation Seasonality). Following spatial thinning, the dataset was reduced by 35.6% in the number of occurrence records, thereby minimizing spatial autocorrelation while maintaining broad geographic coverage.

Despite strong statistical performance, ecological realism was limited. On average, only 18.78% of highly suitable areas (≥0.8) currently retain forest cover when the three countries were pooled. Notably, less than the 50% of the Taï Forest National Park—one of the most intact and the largest rainforest system in West Africa—was predicted as highly suitable (i.e. 47.69%), while heavily deforested regions were.

Optimized MaxEnt model

The optimal MaxEnt model, selected based on minimum AICc, showed good predictive performance (Swets, 1988): global ROC AUC = 0.826, and mean jackknife AUC = 0.854. The null model test (100 randomizations, $p < 0.001$) confirmed performance significantly better than random (mean AUC of null models = 0.747). The major predictor among covariates was precipitation seasonality (bio 15, with a value of 57.4% of the explained variance), followed by annual mean temperature (bio 1, 33%); the other covariates had a practically negligible impact. The response curves (Fig. 3) revealed a negative relationship between occurrence probability and temperature, and a unimodal response to precipitation of the driest month (bio 14), with optimal values around approximately 40–50 mm. Mean diurnal range (bio 2) showed an approximately linear negative relationship, while annual precipitation (bio 12) had no significant effect. The jackknife confirmed the dominant role of a limited number of predictors; in fact, the full model achieved a jackknife AUC = 0.921, with only a slight performance decrease when a single variable was excluded (mean AUC = 0.913). Conversely, models built using a single variable showed a more pronounced reduction in performance (mean AUC = 0.827), although still comparable to the global model AUC.

ENFA model

The ENFA model yielded an AUC of 0.804 and a non-sig-

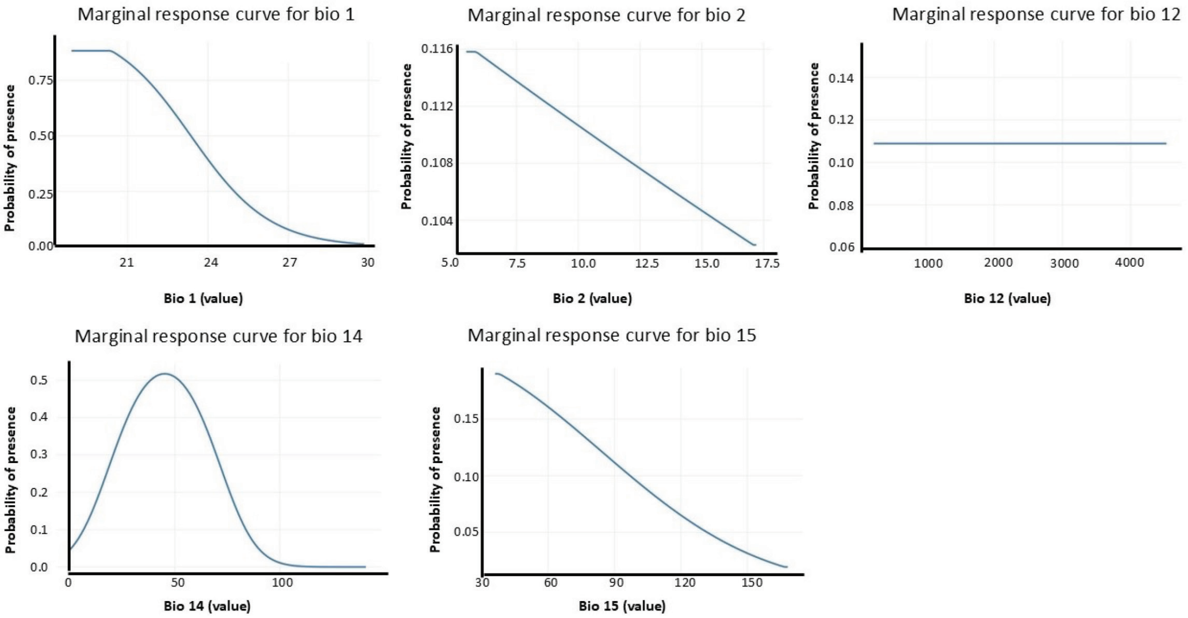

Fig. 3. Marginal response curves of the selected bioclimatic predictors used in the MaxEnt models. The plots illustrate the effect of each environmental variable on predicted probability of species presence when all other variables are held at their average sample value. Variables include Bio 1 (annual mean temperature), Bio 2 (mean diurnal range), Bio 12 (annual precipitation), Bio 14 (precipitation of the driest month), and Bio 15 (precipitation seasonality). Response shapes highlight both linear and non-linear relationships between climatic gradients and habitat suitability across the study region.

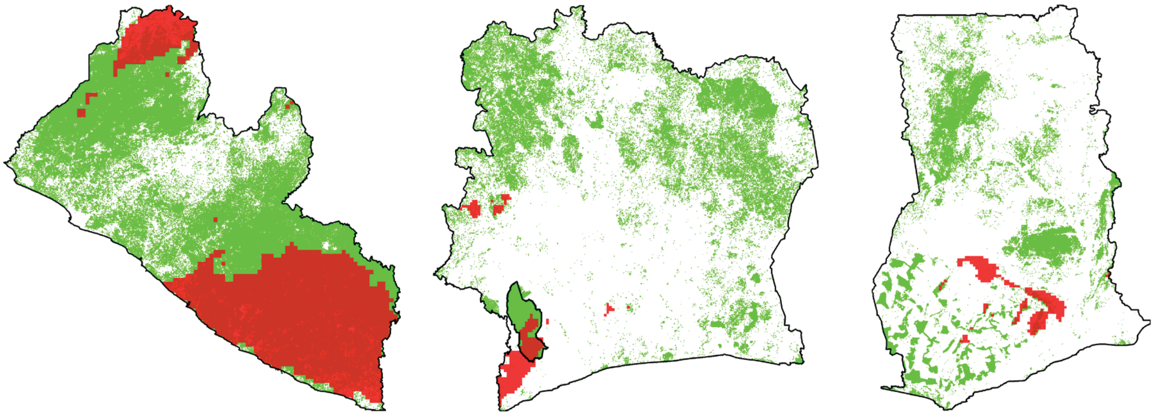

Fig. 4. Overlap between areas of maximum suitability (in red) and forest distribution (in green) in (from left to right) Liberia, Cote d'Ivoire and Ghana. On the map of Ivory Coast, the perimeter of the Tai forest is highlighted by the thicker black line.

nificant Boyce index (0.167), indicating lower predictive performance relative to MaxEnt.

Forest overlap

The overlap between forest cover and very high habitat suitability (>0.8) showed the following percentages: in Côte d'Ivoire, only 4.69% of forested areas fall within high-suitability zones, while in Ghana this value is 6.42%, and in Liberia it reaches 45.23% (Fig. 4). Forest cover as a proportion of national area is 14.69% in Côte d'Ivoire, 11.26% in Ghana, and 32.13% in Liberia.

Habitat Matching Index (HMI) and evaluation of the probabilistic overlap model

Applying the proposed framework, the observed overlap P_O increases markedly from Côte d'Ivoire and Ghana to Liberia (Fig. 5). This pattern is consistent with the higher forest proportion P_F observed in Liberia compared to Cote d'Ivoire and Ghana. Under the null expectation $E[PO] = P_F \cdot P_S$, the substantially higher overlap observed in Liberia suggests that a large component of model–habitat agreement can be explained by greater habitat availability, rather than improved ecological prediction alone. Conversely, the very low overlap values observed in Côte d'Ivoire (4.69%) and Ghana (6.42%), despite non-negligible forest presence, indicate a clear mismatch between predicted suitability and current habitat distribution. These results support the hypothesis that:

- Spatial overlap between suitability and habitat is strongly influenced by the proportion of available habitat (PF) (Fig. 6).
- Higher apparent model "fit" in Liberia is at least partially driven by greater forest extent, increasing the probability of random overlap.
- In highly deforested landscapes (e.g., Côte d'Ivoire), SDM predictions may become decoupled from current habitat, leading to low HMI values and reduced ecological realism.

To account for the effect of habitat availability on model–landscape agreement, we computed the Habitat Matching Index (*HMI*), defined as the ratio between the observed overlap of highly suitable areas with forest cover (*O/S*) and the proportion of forest within each country (P_F). The resulting *HMI* values were: Côte d'Ivoire: 0.319; Ghana: 0.570; and Liberia: 1.407. These values reveal strong differences among countries. Liberia exhibited an $HMI > 1$, indicating that the overlap between predicted suitability and forest cover exceeds that expected under random spatial association. In contrast, Côte d'Ivoire and Ghana showed *HMI* values well below 1, indicating that highly suitable areas are disproportionately located outside forested habitats. Overall, the probabilistic model highlights that SDM outputs should be interpreted relative to habitat availability, and that raw overlap metrics alone may overestimate ecological accuracy in regions with extensive habitat cover.

Discussion

Potential distribution and macroecological signal

Addressing the first research question, our MaxEnt models identify a coherent macroecological pattern in the potential distribution of forest-specialist butterflies across West Africa. At the regional scale, areas of high climatic suitability broadly coincide with the Upper Guinean forest block, extending from Liberia through western Côte d'Ivoire into Ghana and western Togo. This result confirms that the model effectively captures large-scale environmental gradients structuring rainforest ecosystems, particularly precipitation regimes, which emerged as the dominant predictors. Such patterns are consistent with the well-established coupling between rainfall and tropical forest distribution (Miller and Gosling, 2014) and suggest that, at coarse spatial resolutions, climatic variables remain strong determinants of potential species ranges. However, when projections are restricted to the national scale, particularly in Côte d'Ivoire and Ghana, this macroecological coherence breaks down. Predicted high-suitability areas extend

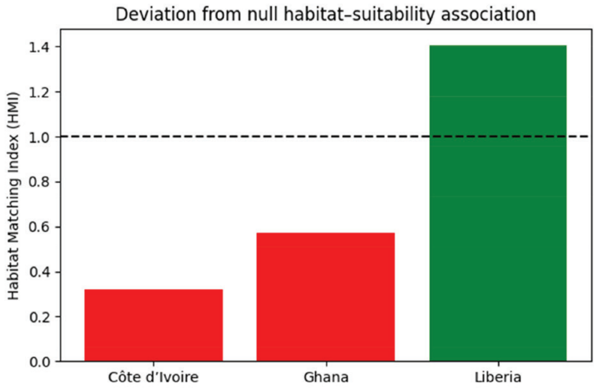

Fig. 5. Habitat Matching Index (HMI) across West African countries. Habitat Matching Index (HMI) values for Côte d'Ivoire, Ghana, and Liberia. HMI quantifies deviations between observed and expected habitat–suitability overlap under a null model of spatial independence. Values < 1 indicate lower-than-expected overlap (model–habitat mismatch), whereas values > 1 indicate higher-than-expected overlap (positive association potentially driven by habitat availability or ecological coupling).

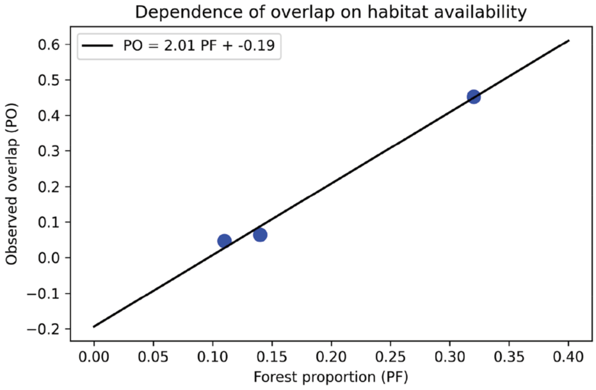

Fig. 6. Dependence of observed overlap on forest availability. Linear relationship between observed habitat–suitability overlap (P_O) and forest proportion (P_F) across study countries. The regression illustrates that spatial overlap increases with forest extent, supporting the hypothesis that part of the apparent model performance is structurally driven by habitat availability rather than ecological niche accuracy alone ($PO = \beta_0 + \beta_1 P_F + \varepsilon$).

into regions that nowadays are not forested, while major forest refugia are not consistently identified. This divergence between regional consistency and local inaccuracy highlights a fundamental scale-dependence in SDM performance and sets the stage for evaluating model reliability beyond purely statistical criteria.

Model performance and the limits of statistical validation

With respect to the second research question, optimized MaxEnt models exhibited high statistical performance with a strong predictive ability (AUC = 0.854), confirmed by null model testing ($p < 0.001$). These values fall within ranges typically interpreted as excellent to good diagnostic performance (Swets, 1988), reinforcing the perception of model robustness. Nevertheless, our results clearly demonstrate that the standard validation metric AUC primarily assesses discrimination between presence and background conditions rather than ecological validity. High AUC values can arise even when predictions are ecologically misleading, particularly when environmental gradients are strong and presence data are spatially structured (Merow et al., 2013). In this study, the strong climatic signal associated with precipitation likely facilitates high statistical performance, but this does not guarantee accurate representation of current habitat suitability. The discrepancy between statistical and ecological performance is particularly evident in the low concordance between predicted suitability and forest cover (mean overlap = 18.78%), indicating that model accuracy in a statistical sense does not translate into biologically meaningful predictions. This finding reinforces a growing body of literature cautioning against overreliance on AUC-based evaluation in SDMs (Warren and Seifert, 2011).

Ecological realism and limitations in dynamic landscapes

The third research question concerns the ecological relevance of model predictions. Our results reveal a pronounced decoupling between climatic suitability and actual habitat conditions, particularly in highly deforested landscapes. The failure to predict high suitability in Taï Forest—one of the most intact rainforest systems in West Africa (Koné et al., 2024; Månefjord et al., 2025) —represents a critical limitation, as this area is known to harbour high biodiversity and function as a major refugium. In fact, the climate model, if it works to predict a species' ecological niche, should be able to represent a proxy for the type of habitat, especially in a very large and stable area like the Taï forest. If this does not happen, it obviously implies that the climate model alone cannot be considered a good proxy for the actual ecological distribution of the target species. Conversely, the model predicts high suitability in regions where forest has been almost entirely removed. This pattern likely reflects the persistence of climatic conditions associated with historical forest cover, rather than current ecological viability. Such temporal mismatches are inherent to SDMs that rely on occurrence data spanning multiple decades and do not incorporate dynamic land-use change.

Additional factors further constrain ecological realism. First, sampling bias in occurrence data, particularly from GBIF, may inflate suitability in historically surveyed but degraded areas (Beck et al., 2014; Qian et al., 2022). Second, the ecological specialization of the target species—likely dependent on microhabitat features such as canopy structure, host plants, and microclimatic buffering (Fermon, 2002)—cannot be captured by coarse-resolution climatic predictors. Fortunately, the available data quantity and quality is increasing and therefore this issue would be less relevant in the years to come (Vance et al., 2024). This limitation is well documented in tropical systems, where microclimatic heterogeneity plays a key role in species persistence (Montejo-Kovacevich et al., 2020). Together, these factors indicate that MaxEnt models, when based solely on climatic variables, are inherently limited in their ability to predict fine-scale distributions of habitat-specialist insects in rapidly changing environments (Lissovsky et al., 2021). However, these models can be used to assess the causes of extinction although our used data do not provide any information from reports of local extinctions (Fois et al., 2018).

Spatial bias and the role of habitat availability: insights from the HMI framework

The fourth research question directly addresses whether the observed overlap between predicted suitability and forest cover reflects ecological processes or is simply driven by habitat availability. The probabilistic framework developed in this study provides a formal basis for disentangling these effects. Raw overlap values show a strong gradient among countries, with Liberia exhibiting substantially higher overlap (45.23%) compared to Côte d'Ivoire (4.69%) and Ghana (6.42%). At face value, this could be interpreted as evidence of better model performance in Liberia. However, Liberia also has a much higher proportion of forest cover (32.13%), increasing the baseline probability of overlap under random spatial association.

The Habitat Matching Index (*HMI*) corrects for this effect by normalizing overlap against habitat availability. The resulting values—Côte d'Ivoire (0.319), Ghana (0.570), and Liberia (1.407)—fundamentally alter the interpretation. While Liberia still shows a positive deviation from random expectation ($HMI > 1$), indicating some degree of ecological signal, this effect is moderate rather than overwhelming. In contrast, Côte d'Ivoire and Ghana exhibit strong negative deviations ($HMI < 1$), demonstrating that predicted suitability is disproportionately located outside forested habitats. These results confirm that a substantial component of model–habitat agreement is driven by habitat availability itself, rather than by true ecological accuracy. In other words, the higher apparent "fit" in Liberia is partly a probabilistic artifact of greater forest extent. This finding is further supported by the conceptual regression model ($P_O = \beta_0 + \beta_1 P_F + \epsilon$), in which the positive relationship between overlap and forest proportion reflects a structural dependency of SDM outputs on habitat distribution.

From a theoretical perspective, this represents a critical insight: SDM validation must account for the spatial distribution of habitat to avoid conflating ecological signal with geometric probability. The HMI provides a simple yet generalizable metric to achieve this, with direct applicability to a wide range of modelling frameworks, including MaxEnt and other presence-only approaches.

Conservation implications and applied significance

The fifth research question concerns the implications of these findings for conservation. The combination of low forest overlap (18.78%) and low *HMI* values in deforested countries indicates a severe contraction of ecologically functional habitat. For high-value species such as *Charaxes fournierae* and *Papilio zalmoxis*, it is highly unlikely to make an assessment of extinction risk based solely on climate models. The only estimate that could be highlighted from the current evidence is to see how many suitable areas have been deforested and then assess the reduction in range for each species – something that climate-based models, by failing to capture habitat availability, simply cannot do.

Moreover, the strong spatial mismatch identified here implies that conservation strategies based solely on climatic suitability could misdirect resources toward areas that are no longer ecologically viable, while partially overlooking critical refugia such as Taï Forest. In fact, precipitation variables, while often predictive at broad scales, may function simply as proxies for forest cover when natural vegetation persists; this can lead to erroneous predictions in landscapes where the original forest–climate equilibrium has been disrupted by anthropogenic land conversion and deforestation. This risk is particularly acute for poorly known species (e.g., *Apallaga ankasa*), where limited ecological data further amplify uncertainty. For these cases, MaxEnt outputs are inadequate without field validation (Fois et al., 2018; Luiselli and Pacini, 2025). In such cases, independent field data are essential to ground-truth model predictions, as they provide direct evidence of species presence, habitat use, and local population conditions that cannot be inferred from environmental variables alone. Systematic surveys and ecological observations are therefore necessary to reduce uncertainty, refine model outputs, and ensure that predictions are biologically meaningful and relevant for conservation assessment.

On the other hand, these models can support decisions on where ecological restoration measures can be planned (Sahana et al., 2022). A more robust conservation framework therefore requires integrating multiple layers of information: (i) climatic suitability as a measure of macro-scale potential, (ii) habitat availability and integrity as meso-scale constraints, and (iii) species-specific ecological requirements at fine scales. The HMI framework contributes to this integration by providing a diagnostic tool to assess whether model predictions are ecologically meaningful or driven by spatial bias.

General implications for species distribution modelling

Beyond the specific case study, this work has broader implications for SDM theory. It demonstrates that model outputs are jointly determined by environmental gradients and the spatial configuration of habitat, and that failure to account for the latter can lead to systematic misinterpretation of results. In highly heterogeneous and rapidly changing landscapes, such as tropical forest regions, this issue becomes particularly acute. Here, the decoupling between climate and land cover undermines the fundamental assumption that current distributions reflect equilibrium with environmental conditions. As a result, SDMs may accurately reconstruct historical climatic niches while failing to predict contemporary ecological distributions. The integration of probabilistic null models, such as the

HMI framework, represents a necessary step toward more rigorous model evaluation. By explicitly quantifying the expected overlap under random conditions, this approach allows researchers to distinguish between genuine ecological signal and artifacts of spatial structure.

Acknowledgements

We thank the Global Biodiversity Information Facility (GBIF) for data access. We are indebted with Prof. Julia E. Fa (Manchester Metropolitan University) for having revised the English style, and to two anonymous referees for helpful comments on the submitted draft.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary dataset

The Supplementary dataset for this article can be found online at: https://osf.io/xhscz/

References

Aiello-Lammens, M. E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5): 541–545. https://doi.org/10.1111/ecog.01132

Ajong, S.N., Luiselli, L., Lado, T.F., Demaya, G.S., Eniang, E.A., Segniagbeto, G.H., Ketoh, G., Le Duc, O., De Palma, M., Amori, G., Fa, J., Behangana, M., Hema, E., Dendi, D., 2024. Living with hyraxes: biogeography and comparative ecology of West African Fornasinius beetles. *African Journal of Ecology*, 62 (3): e13325. https://doi.org/10.1111/aje.13325

Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its *effect* on modeling species' geographic distributions. *Ecological Informatics*, 19: 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Bombi, P., Akani, G.C., Ebere, N., Luiselli, L., 2011. Potential effects of climate change on high-and low-abudance populations of the Gaboon viper (Bitis gabonica) and the nose-horned viper (B. nasicornis) in southern Nigeria. *The Herpetological Journal*, 21 (1): 59–64.

Bonebrake, T.C., Ponisio, L.C., Boggs, C.L., Ehrlich, P.R., 2010. More than just indicators: a review of tropical butterfly ecology and conservation. *Biological Conservation*, 143: 1831–1841. https://doi.org/10.1016/j.biocon.2010.04.044

Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275: 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Bourgoin, C., Verhegghen, A., Carboni, S., Degreve, L., Ameztoy Aramendi, I., Ceccherini, G., Colditz, R., Achard, F., 2025. *Global Forest Maps for the Year 2020 to Support the EU Regulation on Deforestation-free Supply Chains*. Luxembourg: Publications Office of the European Union. https://data.europa.eu/doi/10.2760/1975879, JRC141702

Cabette, H.S., Souza, J.R., Shimano, Y., Juen, L., 2017. Effects of changes in the riparian forest on the butterfly community (Insecta: Lepidoptera) in Cerrado areas. *Revista Brasileira de Entomologia*, 61 (1): 43–50. https://doi.org/10.1016/j.rbe.2016.10.004

Carcasson, R.H., 2011. *Handguide to the butterflies of Africa*. London, U.K.: Wm Collins & Sons & Co. 216 p.

Chatelain, C., Dao, H., Gautier, L., Spichiger, R., 2004. Forest cover changes in Côte d'Ivoire and Upper Guinea. In *Biodiversity of West African forests: an ecological atlas of woody plant species*. Wallingford UK: CABI Pulishing, p. 15–32.

Di cola, v., Broennimann, o., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordjiik, W., Salamin, N., Guisan, A., 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40 (6): 774–787. https://doi.org/10.1111/ecog.02671

Doser, J.W., Finley, A.O., Banerjee, S., 2023. Joint species distribution models with imperfect detection for highdmensional spatial data. *Ecology*, 104 (9): e4137. https://doi.org/10.1002/ecy.4137

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. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fermon, H., 2002. *The butterfly community of a managed West African rainforest: patterns of habitat specificity, diversity, stratification and movement*. Unpublished dissertation. Georg-August-Universitaet Goettingen, Goettingen,Germany.

Fermon, H., Waltert, M., Larsen, T.B., Dall'Asta, U., Muehlenberg, M., 2000. Effects of forest management on diversity and abundance of fruit-feeding nymphalid butterflies in south-eastern Côte d'Ivoire. *Journal of Insect Conservation*, 4 (3): 173–188. https://doi.org/10.1023/a:1009638808635

Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43 (3): 393–404. https://doi.org/10.1023/a:1021374009951

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 (122): 4302–4315. https://doi.org/10.1002/joc.5086

Fois, M., Cuena-Lombraña, A., Fenu, G., Bacchetta, G., 2018. Using species distribution models at local scale to guide the search of poorly known species: review, methodological issues and future directions. *Ecological Modelling*, 385: 124–132.

Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS One*, 9 (5): e97122. https://doi.org/10.1371/journal.pone.0097122

Freeman, B., Roehrdanz, P.R., Peterson, A.T., 2021. Species distributions and forest connectivity across the humid Upper Guinea lowland rainforest of West Africa. Biodiversity and Conservation, 28: 671–685. https://doi.org/10.1007/s10531-018-01684-6

Giachello, S., Lefosse, S., Simoncini, A., Bonato, L., 2025. Species distribution models for the conservation of a micro-endemic animal: the contribution of regional land cover. *Biodiversity and Conservation*, 34: 1689–1707. https://doi.org/10.1007/s10531-025-03036-7

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan,T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecology Letters*, 16 (12): 1424–1435. https://doi.org/10.1111/ddi.12066

Fois, M., Bacchetta, G., Cuena-Lombraña, A., Cogoni, D., Pinna, M.S., Sulis, E., Fenu, G., 2018. Using extinctions in species distribution models to evaluate and predict threats: a contribution to plant conservation planning on the island of Sardinia. *Environmental Conservation*, 45 (1): 11–19.

Henning, S.F., 1988. *The Charaxinae butterflies of Africa*. Johannesburg: Aloe Books. 457 p.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2017. *Dismo: Species Distribution Modeling*. R package version 1.1-4.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199 (2): 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83 (7): 2027–2036. https://doi.org/10.2307/3071784

Khanum, R., Mumtaz, A.S., Kumar, S., 2013. Predicting impacts of climate change on medicinal asclepiads of Pakistan using Maxent modeling. *Acta Oecologica*, 49: 23–31. https://doi.org/10.1016/j.actao.2013.02.007

Koné, D., Diarrassouba, A., Diagne, T., Luiselli, L., 2024. On the occurrence of Pelusios gabonensis in the Taï National Park (Côte d'Ivoire). *African Journal of Ecology*, 62 (3): e13320. https://doi.org/10.1111/aje.13320

Larsen, T.B., 2008. Forest butterflies in West Africa have resisted extinction… so far (Lepidoptera: Papilionoidea and Hesperioidea). *Biodiversity and Conservation*, 17 (12): 2833–2847. https://doi.org/10.1007/s10531-008-9399-z

Li, Y., Li, M., Li, C., Liu, Z., 2020. Optimized maxent model predictions of climate change impacts on the suitable distribution of Cunninghamia lanceolata in China. *Forests*, 11 (3): 302. https://doi.org/10.3390/f11030302

Lissovsky, A.A., Dudov, S.V., Obolenskaya, E.V., 2021. Species-distribution modeling: advantages and limitations of its application. 1. General approaches. *Biology Bulletin Reviews*, 11 (3): 254–264.

Luiselli, L., Fa, J. E., Le Duc, O., Eniang, E.A., Gonedele-Bi, S., Segniagbeto, G. H., Behangana, M., Ajong, S.N., Hema, E.M., Koné, D., Demaya, G.S., Morani, M., Amori, G., Malonza, P.K., Mutua, M.N., Ngalieu, M.M.K., Ofori-Boateng, C., Amponsah, C.K., Opoku-Kwarteng, C., Togba, S.O., Mensah, M., De Palma, M., Pacini, N., Dendi, D., 2025. Red Listing African Goliath beetles: assessing threats and conservation needs. *African Journal of Ecology*, 63 (1): e70018. https://doi.org/10.1111/aje.70018

Luiselli, L., Pacini, N., 2025. Remote sensing that makes sense in ecological research—from pixels to conservation. *African Journal of Ecology*, 63 (1): e70002. https://doi.org/10.1111/aje.70002

Månefjord, H., Yamoa, A.S.D., Gbogbo, Y.A., Müller, L., Runemark, A., Kouakou, B.K., Boateng, R., Huzortey, A.A., Badu, I.K., Wahlberg, N., Brydegaard, M., Zoueu, J.T., Anderson, B., Li, M., 2025. Stratification of insect diversity and daily activity patterns in the West African virgin forest Taï assessed by entomological Lidar. *Scientific Reports*, 15 (1): 25663. https://doi.org/10.1038/s41598-025-05200-z

Mena, S., Kozak, K.M., Cárdenas, R.E., Checa, M.F., 2020. Forest stratification shapes allometry and flight morphology of tropical butterflies. *Proceedings of the Royal Society B*, 287 (1937): 20201071. https://doi.org/10.1098/rspb.2020.1071

Merow, C., Smith, M.J., Silander Jr, 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. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Miller, C.S., Gosling, W.D., 2014. Quaternary forest associations in lowland tropical West Africa. *Quaternary Science Reviews*, 84: 7–25. https://doi.org/10.1016/j.quascirev.2013.10.027

Montejo-Kovacevich, G., Martin, S.H., Meier, J.I., Bacquet, C.N., Monllor, M., Jiggins, C.D., Nadeau, N.J., 2020. Microclimate buffering and thermal tolerance across elevations in a tropical butterfly. *Journal of Experimental Biology*, 223 (8): jeb220426. https://doi.org/10.1242/jeb.220426

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. https://doi.org/10.1111/2041-210x.12261

New, T.R., Collins, N.M., 1991. *Swallowtail Butterflies: An Action Plan for Their Conservation*. Gland: IUCN/SSC Lepidoptera Specialist Group. 37 p.

Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E. 2017. Opening the black box: an open-source elease of Maxent. *Ecography*, 40 (7): 887–93. https://doi.org/10.1111/ecog.03049

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31: 161–175. https://doi.org/10.1111/j.0906-7590.2007.5203.x

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190 (3-4): 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Qian, H., Zhang, J., Jiang, M.C., 2022. Global patterns of fern species diversity: an evaluation of fern datain GBIF. *Plant Diversity,* 44 (2): 135–140. https://doi.org/10.1016/j.pld.2021.10.001

R Core Team, 2026. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raes, N., Ter Steege, H., 2007. A null model for significance testing of presence only species distribution models. *Ecography*, 30 (5): 727–736. https://doi.org/10.1111/j.2007.0906-7590.05041.x

Sáfián, S., 2021. *The butterfly fauna (Lepidoptera: Papilionoidea) of Liberia*. Doctoral dissertation. Roth Gyula Doctoral School of Forestry and Game Management Diversity and Ecology of Forest Ecosystems, Faculty of Forestry, University of Sopron, Sopron, Hungary.

Sáfián, S., Csontos, G., Winkler, D., 2011. Butterfly community recovery in degraded rainforest habitats in the Upper Guinean Forest Zone (Kakum forest, Ghana). *Journal of Insect Conservation*, 15 (1): 351–359. https://doi.org/10.1007/s10841-010-9343-x

Sáfián, S., Koïvogui, S., Simonics, G., Florczyk, K.J., 2020. Butterfly diversity (Lepidoptera: Papilionoidea) in the Ziama Massif in Guinea and the adjacent Wonegizi and Wologizi Mountains in Liberia (West Africa): a transboundary conservation approach. *Metamorphosis*, 31 (1): 104–128. https://doi.org/10.4314/met.v31i1.21

Sahana, M., Areendran, G., Sajjad, H., 2022. Assessment of suitable habitat of mangrove species for prioritizing restoration in coastal ecosystem of Sundarban Biosphere Reserve, India. *Scientific Reports*, 12 (1): 20997. https://doi.org/10.1038/s41598-022-24953-5

Seyed, M-R. A., Mostafa, T., Hossein, B., 2020. MaxEnt modeling for predicting suitable habitats and identifying the effects of climate change on a threatened species, Daphne mucronata, in central Iran. *Ecological Informatics*, 43: 116–123. https://doi.org/10.1016/j.ecoinf.2017.10.002

Slone, T.H., Orsak, L.J., Malver, O., 1997. A comparison of price, rarity and cost of butterfly specimens: Implications for the insect trade and for habitat conservation. *Ecological Economics*, 21 (1): 77–85. https://doi.org/10.1016/s0921-8009(96)00096-1

Small, R.D.S., 2007. Becoming unsustainable? Recent trends in the formal sector of insect trading in Papua New Guinea. *Oryx*, 41: 386–389.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science*, 240 (4857): 1285–1293. https://doi.org/10.1126/science.3287615

Telenius, A., 2011. Biodiversity information goes public: GBIF at your service. *Bioscience*, 61 (6): 448–450. https://doi.org/10.1111/j.1756-1051.2011.01167.x

Tessarolo, G., Lobo, J.M., Rangel, T.F., Hortal, J., 2021. High uncertainty in the effects of data characteristics on the performance of species distribution models. *Ecological Indicators*, 121: 107147. https://doi.org/10.1016/j.ecolind.2020.107147

Tsingalia, M.H., Kipngetich, J.M., Oroto, B.A., 2026. Charaxes butterflies as bioindicators of forest integrity: conservation priorities for Kenya's threatened Nandi Forests. *African Journal of Ecology*, 64 (4): e70176.

Vance, T. C., Huang, T., Butler, K. A., 2024. Big data in Earth science: emerging practice and promise. *Science*, 383 (6688): eadh9607.

Wang, Z., Chan, W.-P., Pham, N.T., Zeng, J., Pierce, N.E., Lohman, D.J., Meng, W., 2023. One in five butterfly species sold online across borders. Biological Conservation, 283: 110092. https://doi.org/10.1016/j.biocon.2023.110092

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21 (2): 335–342. https://doi.org/10.1890/10-1171.1

West, D.A., Hazel, W.N., 1979. Natural pupation sites of swallowtail butterflies (Lepidoptera: Papilioninae): Papilio polyxenes Fabr., P. glaucus L. and Battus philenor (L.). *Ecological Entomology*, 4 (4): 387–392.

Zhao, G., Cui, X., Sun, J., Li, T., Wang, Q. I.., Ye, X., Fan, B., 2021. Analysis of the distribution pattern of Chinese Ziziphus jujuba under climate change based on optimized biomod2 and MaxEnt models. *Ecological Indicators*, 132: 108256. https://doi.org/10.1016/j.ecolind.2021.108256

Received January 9, 2026
Accepted April 18, 2026

Supplementary materials

Table S1. Detailed pricing information per specimen for the target butterfly species identified on eBay using the methods described in this study

Species	Family	Price ($)	Average price ($)
Charaxes fournierae	Nymphalidae	450–1,904	1,177
Charaxes mycerina	Nymphalidae	50–150	100
Charaxes nichetes	Nymphalidae	35–155	95
Charaxes petersi	Nymphalidae	50–119	84.5
Charaxes smaragdalis	Nymphalidae	52–102	77
Charaxes subornatus	Nymphalidae	30–60	45
Graphium latreillianus	Papilionidae	76–108	92
Papilio horribilis	Papilionidae	20–250	135
Papilio zalmoxis	Papilionidae	183–2,129	1,156
Pteroteinon pruna	Hesperiidae	15–70	35

Table S2. Number of occurrence records per species beforeand after applying spatial thinning with a minimum distance of 5 km. The thinning procedure reduced the total number of points by 35.6%, removing spatial clusters to minimize spatial autocorrelation and sampling bias, while maintaining a representative geographic coverage of the study area.

Species	Records before thinning	Records after thinning	% Reduction
Abantis lucretia	1	1	0%
Apallaga ankasa	2	2	0%
Charaxes fournierae	2	2	0%
Charaxes mycerina	2	2	0%
Charaxes nichetes	8	6	25%
Charaxes petersi	7	5	28.6%
Charaxes smaragdalis	6	4	33.3%
Charaxes subornatus	5	3	40%
Graphium latreillianus	17	10	41.2%
Papilio horribilis	17	11	35.3%
Papilio zalmoxis	3	3	0%
Pteroteinon pruna	2	2	0%
Total	90	58	35.6%

Table S3. Bioclimatic variables used for the MaxEnt model, after deleting the collinear variables (see Dormann et al., 2013 for the employed methodology)

Code	Name	Unit of measurement
BIO1	Annual Mean Temperature	°C × 10
BIO2	Mean Diurnal Range	°C × 10
BIO3	Isothermality	%
BIO4	Temperature Seasonality	%
BIO5	Max Temperature of Warmest Month	°C × 10
BIO6	Min Temperature of Coldest Month	°C × 10
BIO7	Temperature Annual Range	°C × 10
BIO8	Mean Temperature of Wettest Quarter	°C × 10
BIO9	Mean Temperature of Driest Quarter	°C × 10
BIO10	Mean Temperature of Warmest Quarter	°C × 10
BIO11	Mean Temperature of Coldest Quarter	°C × 10
BIO12	Annual Precipitation	mm
BIO13	Precipitation of Wettest Month	mm
BIO14	Precipitation of Driest Month	mm
BIO15	Precipitation Seasonality	%
BIO16	Precipitation of Wettest Quarter	mm
BIO17	Precipitation of Driest Quarter	mm
BIO18	Precipitation of Warmest Quarter	mm
BIO19	Precipitation of Coldest Quarter	mm

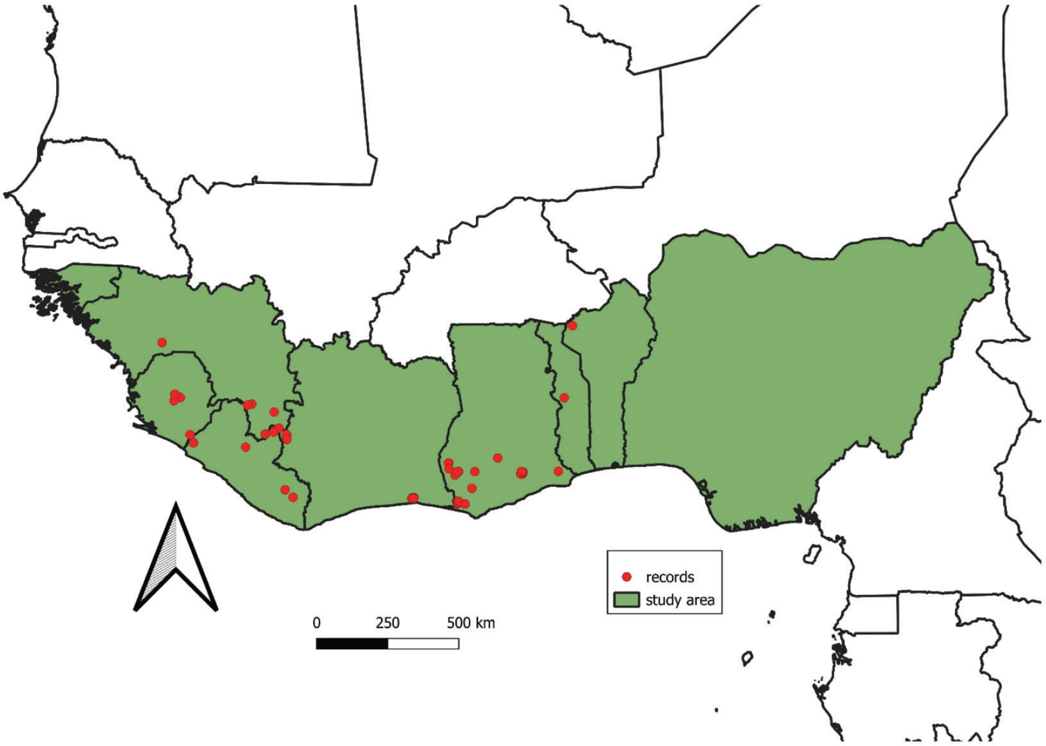

Fig. S1. Distribution records of the various butterfly species (pooled) in the study area.