Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Nested species distribution models highlight scale and climate-dataset effects on adaptation planning for sweet chestnut (Castanea sativa) forests in Türkiye

Huseyin Dogan
Istanbul University Cerrahpaşa
Osman Yalcın Yilmaz
yilmazy@iuc.edu.tr
Istanbul University Cerrahpaşa

Research Article

Keywords: Species Distribution Modeling, Multi-scale Modeling, Climate Change Adaptation, Forest Conservation, Topographic Heterogeneity, Micro-refugia, Castanea sativa

Posted Date: February 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8708047/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Projections of forest habitat suitability under climate change are highly sensitive to modeling scale and climate dataset selection, yet these sources of uncertainty are rarely quantified explicitly. Sweet chestnut (*Castanea sativa* Mill.) forests in Türkiye occupy climatically marginal zones where topographic heterogeneity strongly mediates climate impacts, providing an ideal system for evaluating how scale-dependent processes alter conservation priorities and refuge identification..

Methods

We developed a hierarchical SDM framework integrating national-scale ensemble models (~ 1 km resolution) with regional-scale models (~ 120 m resolution) incorporating downscaled climate, topography, and hydrological variables. Using WorldClim 2.1 and CHELSA 2.1 in parallel across three CMIP6 global circulation models under SSP1-2.6 and SSP5-8.5 scenarios, we explicitly quantified both scale- and dataset-driven uncertainty in current and future projections.

Results

Despite high predictive performance (AUC > 0.95; TSS > 0.75), national-scale models exhibited pronounced dataset-driven spatial divergence, with future projections revealing contrasting conservation narratives: WorldClim predicted severe habitat contraction (50–80% loss), while CHELSA indicated substantial persistence in high-elevation micro-refugia (10–20% loss under SSP1-2.6). Regional-scale integration of topographic and hydrological constraints reduced predicted suitable habitat by 60% relative to national projections, primarily by excluding climatically permissive but ecologically unsuitable lowland areas.

Conclusions

Hierarchical SDMs integrating macroclimatic projections with local environmental constraints substantially reduce prediction uncertainty and enhance ecological realism in topographically complex landscapes. Our results demonstrate that coarse-scale climatic models systematically misidentify conservation priorities by overlooking topographically-buffered micro-refugia, with direct implications for adaptive forest management and climate adaptation planning in mountain ecosystems globally.

1. INTRODUCTION

Forest ecosystems provide critical ecosystem services including carbon sequestration, biodiversity conservation, and socioeconomic benefits, yet face unprecedented pressure from accelerating climate change, which poses substantial threats to plant diversity across Europe (Thuiller et al., 2005; FAO, 2020; IPCC, 2021). Understanding how forest tree species respond to shifting climatic conditions is essential for developing effective adaptive management strategies and ensuring the long-term sustainability of forest resources (Lindner et al., 2010; Hanewinkel et al., 2013). Species distribution models (SDMs) are widely used to project habitat suitability under current and future climatic scenarios, offering quantitative support for conservation planning and silvicultural decision-making (Guisan & Zimmermann, 2000; Franklin, 2010).

However, SDM outputs are strongly influenced by spatial scale and the choice of environmental datasets, often leading to inflated or ecologically unrealistic predictions when models rely solely on coarse-resolution climate variables (Beaumont et al., 2005; Araújo & Guisan, 2006). For forest species with narrow ecological requirements and strong topographic associations, such scale-dependent uncertainty can undermine reliable habitat assessment and misdirect conservation and climate adaptation strategies.

Most SDM studies operate at a single spatial resolution, typically using bioclimatic datasets at kilometer-scale resolution. While these models capture broad climatic constraints, they often fail to represent local habitat heterogeneity driven by topography, hydrology, and microclimate—factors that are particularly important in mountainous forest landscapes (Guisan & Thuiller, 2005; Franklin et al., 2013). As a result, coarse-scale SDMs often overestimate suitable habitat, especially in lowland areas where macroclimate appears favorable but local conditions limit establishment or long-term persistence (Elith et al., 2010; Roberts et al., 2017).

Recent methodological advances highlight the potential of multi-scale and hierarchical SDM frameworks, in which broad-scale climatic suitability is refined using fine-scale environmental predictors (Mertes & Jetz, 2018; Mateo et al., 2019). Nested modeling approaches explicitly integrate macroclimatic constraints with local habitat characteristics, improving ecological realism and reducing extrapolation into unsuitable areas (Guisan et al., 2013). Despite growing recognition of their value, such approaches remain underutilized in applied forest ecology, and their ability to quantify uncertainty arising from different climate datasets has received limited empirical evaluation.

Sweet Chestnut (Castanea sativa Mill.) is an ecologically and economically important tree species across southern Europe and western Asia. In Türkiye, chestnut forests cover approximately 67,000 hectares and contribute to timber production, non-timber forest products, and ecosystem services (OGM, 2012). The species occurs at the eastern margin of its native range in this region, where strong climatic gradients and complex topography create a fragmented mosaic of suitable habitats (Zohary & Hopf, 1988; Villani et al., 1994; Conedera et al. 2004; Villani et al., 2004). These marginal populations are particularly sensitive to summer moisture deficits and climatic instability, making C. sativa an effective model system for evaluating how scale-dependent modeling choices influence habitat projections. Türkiye's steep elevation gradients—ranging from sea level to over 2000 m within short distances—further provide a natural laboratory for testing SDM performance in topographically complex forest landscapes.

In this study, we evaluate a nested, multi-scale species distribution modeling framework to evaluate how spatial scale and climate dataset choice jointly influence projections of forest habitat suitability. Using Sweet Chestnut (Castanea sativa Mill.) in Türkiye as a case study, we first compare national-scale ensemble SDMs constructed with two widely used climate datasets (WorldClim 2.1 and CHELSA 2.1) to quantify dataset-driven uncertainty in current and future projections under contrasting emission scenarios (SSP1–2.6 and SSP5–8.5). We then refine these projections through regional-scale hierarchical models that integrate downscaled climate variables with high-resolution topographic and hydrological predictors, explicitly testing how fine-scale environmental constraints modify coarse-scale climatic suitability.

Specifically, we address three questions relevant to forest adaptation planning: (1) to what extent do SDM outcomes and inferred conservation priorities depend on climate dataset choice at national scale; (2) how strongly do nested regional models reduce overprediction inherent in coarse-resolution climate-only models; and (3) how do these modeling choices affect estimates of climate-driven habitat loss and the identification of stable refugia in topographically complex landscapes. By explicitly linking modeling scale, dataset uncertainty, and ecological realism, this study provides a transferable framework for improving SDM-based decision support for climate-sensitive forest species beyond the focal system.

2. METHODS

2.1. Species, Study Area and Data Collection

Castanea sativa Mill. (Sweet Chestnut) is a deciduous broadleaved tree native to southern Europe and western Asia, with its eastern distribution extending through Türkiye into the Caucasus region, including Georgia and adjacent areas of western Transcaucasia (Zohary & Hopf, 1988; Villani et al., 2004; Conedera et al., 2016). This eastern range margin encompasses pronounced gradients in climate, elevation, and continentality associated with the transition between Mediterranean, Black Sea, and Caucasian climatic regimes. National-scale modeling was conducted across the entire territory of Türkiye, while regional-scale analyses focused on the Marmara region, a climatically transitional landscape where topographic heterogeneity is expected to strongly influence habitat suitability at fine spatial scales (Fig. 1).

Species occurrence data were compiled from multiple sources, including forest management inventories provided by the General Directorate of Forestry, herbarium records, and targeted field surveys conducted between 2018 and 2022. To ensure ecological relevance and minimize anthropogenic bias, only records representing naturally established, mature stands in which *C. sativa* was identified as the dominant species were retained. After data cleaning and removal of duplicate records, an initial dataset of 1,245 georeferenced presence points was obtained.

To reduce spatial autocorrelation and sampling bias, occurrence data were spatially thinned using the spThin package in R (Dormann et al., 2007; Aiello-Lammens et al., 2015). A minimum distance of 3 km between presence points was applied for national-scale models, resulting in 393 occurrences, while a 1 km threshold was used for regional-scale models, yielding 210 occurrences. These distances were selected to balance spatial independence with sufficient data density for model calibration at each scale.

Pseudo-absence data were generated to support presence–absence modeling algorithms following the approach of Barbet-Massin et al. (2012).. Pseudo-absence points were spatially buffered from presence locations and constrained within ecologically relevant background extents to reduce the risk of model extrapolation. Multiple pseudo-absence datasets were generated and later incorporated into ensemble modeling to account for uncertainty associated with background selection.

2.2. Environmental Variables

Climatic predictors for the baseline period were obtained from WorldClim 2.1 (1970–2000; Fick & Hijmans, 2017) and CHELSA 2.1 (1981–2010; Karger et al., 2017) at approximately 1 km spatial resolution. All 19 bioclimatic variables were initially considered for both datasets. Future climate projections were derived from three CMIP6 global circulation models (GFDL-ESM4, IPSL-CM6A-LR, and MPI-ESM1-2-HR) under two contrasting Shared Socioeconomic Pathways (SSP1–2.6 and SSP5–8.5), using mid-century time horizons consistent with each dataset (2021–2040 and 2041–2060 for WorldClim; 2011–2040 and 2041–2070 for CHELSA).

Prior to national-scale SDM calibration, multicollinearity among the 19 bioclimatic variables was assessed separately for each climate dataset using Variance Inflation Factor (VIF) analysis implemented in the *usdm* package in R (Naimi et al., 2014). Variables exceeding a VIF threshold of 4 were iteratively removed, resulting in a reduced set of weakly correlated climatic predictors for each dataset (O'Brien, 2007).

Regional-scale SDMs incorporated both climatic and fine-scale environmental predictors. Topographic and hydrological variables, including elevation, slope, and vertical distance to the channel network, were derived from the ASTER Global Digital Elevation Model (Version 2; NASA/METI, 2009) at 30 m resolution using SAGA GIS (Conrad et al., 2015) and resampled to a target resolution of 120 m.

To ensure consistency between national- and regional-scale models within a hierarchical framework, selected climatic variables from the national-scale SDMs (WorldClim 2.1) were downscaled from 1 km to 120 m resolution using Geographically Weighted Regression (GWR) in SAGA GIS (Brunsdon et al., 1996; see Appendix 8). Topographic covariates were included in the downscaling procedure to account for spatial non-stationarity in climate–environment relationships.

Regional-scale modeling followed a two-step nested approach (Mateo et al., 2019; Fig. 2). First, SDMs were calibrated using only the downscaled climatic variables to generate a climate-driven habitat suitability surface. Second, a set of 38 topographic and hydrological variables was screened for multicollinearity using VIF analysis, resulting in the selection of four ecologically relevant predictors: Direct Insolation, Slope, Minimal Curvature, and Vertical Distance to Channel Network. The climate-only habitat suitability raster produced in the first step was then included as an additional predictor alongside these four variables in the final regional-scale SDMs, allowing the integration of macroclimatic constraints with fine-scale environmental heterogeneity.

2.3. Multi-Scale SDMs and Model Evaluations

Species distribution models were developed at two spatial scales using a hierarchical framework that integrates national-scale climatic suitability with regional-scale environmental constraints. National-scale SDMs were calibrated at 1 km resolution across Türkiye using climate predictors only, while regional-scale SDMs were developed at 120 m resolution for the Marmara region by combining downscaled climatic variables with fine-scale topographic and hydrological predictors.

To limit extrapolation beyond the species' known environmental domain, national-scale models incorporated spatial masking based on the documented distribution of *Castanea sativa* in Türkiye (Fig. 1). This approach constrained model calibration and pseudo-absence selection to ecologically relevant areas. Spatial masking was not applied in the regional-scale models due to the smaller and more environmentally homogeneous extent of the Marmara region.

Models were implemented in R using the biomod2 package (version 4.2-6-2: Thuiller et al., 2009; Gueguen et al., 2025) and were calibrated using three algorithms: Random Forest (RF), Generalized Boosted Models (GBM), and Generalized Linear Models (GLM). For each climate dataset (WorldClim 2.1 and CHELSA 2.1), models were fitted using multiple pseudo-absence datasets. To improve predictive accuracy, reduce uncertainty, and enhance spatial and temporal transferability of predictions (Yilmaz et al., 2017; Yates et al., 2018; Sequeira et al., 2018), we implemented ensemble modeling based on model evaluation metrics. Only models exceeding a minimum performance threshold (TSS > 0.75) were retained and combined into ensemble predictions.

Model performance was evaluated using the True Skill Statistic (TSS), Area Under the Receiver Operating Characteristic Curve (ROC-AUC), and the Boyce Index. Spatial autocorrelation was addressed through spatial block cross-validation implemented with the *blockCV* package (Valavi et al., 2019). Block sizes were determined using variogram-based estimates of spatial autocorrelation, resulting in block dimensions of approximately 168 × 215 km for national-scale models and 27 × 51 km for regional-scale models.

Furthermore, to assess model extrapolation under novel environmental conditions, Multivariate Environmental Similarity Surface (MESS: see Appendixes 23–25 and 29–31) analysis was performed for both WorldClim-2.1 and CHELSA-2.1 projections. This method identifies areas where one or more predictor variables fall outside the range of the training data, thereby flagging locations where model predictions may be less reliable (Elith et al., 2010).

Variable selection for national-scale SDMs was conducted separately for each climate dataset using a two-step procedure. First, multicollinearity among predictors was reduced using Variance Inflation Factor (VIF) analysis with a threshold of VIF < 4. Second, repeated model runs and variable importance analyses were used to identify the four most influential predictors for each dataset. For WorldClim-based models, these were bio4 (Temperature Seasonality), bio10 (Mean Temperature of Warmest Quarter), bio12 (Annual Precipitation), and bio15 (Precipitation Seasonality). For CHELSA-based models, the selected predictors were chlsa_bio3 (Isothermality), chlsa_bio4 (Temperature Seasonality), chlsa_bio11 (Mean Temperature of Coldest Quarter), and chlsa_bio19 (Precipitation of Coldest Quarter).

Regional-scale SDMs followed a two-step nested approach adapted from Mateo et al. (2019; Fig. 2). First, models were calibrated using downscaled climatic variables only to generate a climate-driven habitat suitability surface. In the second step, this suitability raster was included as an additional predictor alongside four fine-resolution topographic and hydrological variables—Direct Insolation, Slope, Minimal Curvature, and Vertical Distance to Channel Network—selected through VIF filtering (VIF < 4). Predictor selection was further supported by inspection of response curves across algorithms and pseudo-absence sets to ensure stable and ecologically interpretable relationships.

2.4. Future Projections

Future projections were generated for the national-scale species distribution models using bioclimatic data from WorldClim 2.1 and CHELSA 2.1 in combination with three CMIP6 global circulation models (GFDL-ESM4, IPSL-CM6A-LR, and MPI-ESM1-2-HR). These models represent the subset of CMIP6 GCMs consistently available across both climate datasets, ensuring comparability between WorldClim- and CHELSA-based projections. Projections were conducted under two contrasting Shared Socioeconomic Pathways: SSP1–2.6 and SSP5–8.5.

For WorldClim 2.1, future projections covered the periods 2021–2040 and 2041–2060, while CHELSA 2.1 projections spanned 2011–2040 and 2041–2070. The bioclimatic variables retained in the baseline national-scale models were projected forward using ensemble outputs from GLM, GBM, and RF algorithms. Changes in habitat suitability were quantified using Species Range Change (SRC) analysis to assess potential patterns of habitat loss, gain, and persistence under alternative climate scenarios.

Future climate projections were applied exclusively to national-scale SDMs at 1 km resolution, focusing on macroclimatic controls on the distribution of *Castanea sativa* across Türkiye. Regional-scale SDMs developed for the Marmara region were restricted to current climate conditions and were not projected into the future. This reflects a deliberate modeling choice to prioritize spatial realism and ecological interpretability at fine resolution, rather than propagating uncertainty from coarse-resolution GCM outputs through aggressive downscaling. As a result, regional models provide a robust baseline representation of climate–topography interactions relevant to site-level forest management, while national-scale models capture broad-scale climate-driven range dynamics.

3. RESULTS

3.1. Model Performances and Distribution Patterns of National-scale SDMs

National-scale SDMs were developed using three algorithms (GBM, GLM, and RF) and two climate datasets (WorldClim 2.1 and CHELSA 2.1), resulting in 240 individual model runs. Model performance was evaluated using ROC-AUC, TSS, and the Boyce Index, and only models exceeding a minimum TSS threshold of 0.75 were retained for ensemble construction (Fig. 3).

Models calibrated with WorldClim 2.1 exhibited consistently high predictive performance across all algorithms (Fig. 4). Mean AUC values exceeded 0.95, TSS values ranged from approximately 0.65 to 0.85, and Boyce Index values ranged from 0.65 to 0.85. CHELSA 2.1–based models also showed high overall performance (Fig. 5), with mean AUC values above 0.95 and Boyce Index values generally between 0.70 and 0.85. Compared to WorldClim-based models, CHELSA-based models exhibited a moderate reduction in TSS scores (approximately 5–10%), particularly for GBM and RF; however, a substantial proportion of models still met the ensemble inclusion threshold (Fig. 3).

Ensemble predictions under current climate conditions revealed broadly similar national-scale suitability patterns across both climate datasets, with high suitability concentrated along the Black Sea coast and parts of the Marmara region (Fig. 3). However, CHELSA-based ensembles generally produced more spatially restricted suitability patterns compared to WorldClim-based ensembles, particularly in lowland and interior regions.

Response curves for the four selected WorldClim predictors (Appendix 3) showed declining suitability with increasing temperature seasonality (bio4), a unimodal response to mean temperature of the warmest quarter (bio10), increasing suitability with annual precipitation up to a plateau (bio12), and reduced suitability at high levels of precipitation seasonality (bio15). For CHELSA-based models (Appendix 4), suitability declined sharply with increasing isothermality (chlsa_bio3) and temperature seasonality (chlsa_bio4), exhibited a unimodal response to mean temperature of the coldest quarter (chlsa_bio11), and increased with winter precipitation up to a threshold (chlsa_bio19).

Current Projections for National-scale SDMs

Current (baseline) national-scale ensemble SDMs derived from CHELSA 2.1 and WorldClim 2.1 revealed broadly similar spatial patterns of habitat suitability for *Castanea sativa* across Türkiye, with notable differences in the extent and continuity of predicted suitable areas (Fig. 3).

CHELSA-based ensemble projections identified extensive and relatively continuous suitable habitats along major mountainous regions, particularly in the Black Sea, Marmara, and parts of the Mediterranean regions. In southern Türkiye, CHELSA-based models predicted a continuous belt of suitability extending from the Taurus Mountains toward the Amanos range. In western Anatolia, suitable areas were primarily associated with higher-elevation massifs, including the Aydın–Dilek Mountains, Bozdağ, Nif, Spil, Simav, and Kazdağı ranges. Within the Marmara region, suitability was concentrated on forested uplands such as the Yıldız (Strandzha) Mountains, the Northern Forests surrounding Istanbul, and the Samanlı–Katırlı mountain system.

In contrast, low-elevation coastal plains and major alluvial systems—such as the Bafra and Çarşamba plains in the Black Sea region and coastal lowlands in parts of Kocaeli and Sakarya—were consistently predicted as low-suitability areas in CHELSA-based models. Across the Black Sea region, suitability was largely confined to mid- and high-elevation zones, generally ranging from approximately 400 m to over 1,700 m in the Kaçkar Mountains.

WorldClim 2.1–based ensemble projections showed overall spatial agreement with CHELSA-based results at a national scale but differed in the magnitude and elevation range of predicted suitability. WorldClim-based models assigned moderate suitability values (approximately 0.6–0.7) to

several lowland areas identified as unsuitable in CHELSA-based projections, including portions of the Kandıra and Kaynarca lowlands in the Marmara region and the Bafra, Çarşamba, and Sinop plains in the Black Sea region. Additionally, WorldClim-based models predicted suitability at lower elevations (< 500 m) in some mountainous regions, including parts of the Kazdağı, Menteşe Mountains, and Uludağ massif.

Regional discrepancies were also observed in the Mediterranean region. While both datasets identified suitable habitats in known chestnut-bearing areas of southwestern Anatolia, WorldClim-based models generally predicted lower suitability values compared to CHELSA-based models, whereas CHELSA projections extended suitability into some eastern Mediterranean areas not highlighted by WorldClim-based ensembles.

Expected Species Range Changes for National-scale SDMs

Species Range Change (SRC) analyses revealed pronounced differences in projected habitat dynamics for *Castanea sativa* depending on the climate dataset, emission scenario, and time horizon (Fig. 6, Table 1). Across all global circulation models (GFDL, IPSL, and MPI) and both Shared Socioeconomic Pathways, WorldClim 2.1–based projections consistently indicated substantial reductions in suitable habitat area over time. Under the low-emission SSP1 scenario, projected habitat loss generally approached 40–60%, increasing to approximately 70–90% under the high-emission SSP5 scenario by mid-century.

In contrast, CHELSA 2.1–based projections suggested markedly lower levels of habitat contraction under comparable scenarios. Under SSP1, CHELSA-based models frequently projected the retention of approximately 80–90% of the current potential distribution, with substantially smaller net losses than those derived from WorldClim-based projections. Under SSP5, habitat loss increased across all models and GCMs, though projected contractions remained less severe than those estimated using WorldClim 2.1.

SRC patterns were broadly consistent across the three GCMs, although the magnitude of projected change varied among models and time periods (Fig. 6). WorldClim-based projections showed limited gains in newly suitable habitat and were dominated by extensive habitat loss, whereas CHELSA-based projections exhibited a greater proportion of stable and, in some cases, newly suitable areas, particularly in earlier projection periods. These differences are reflected in the wider interquartile ranges and higher median SRC values observed for WorldClim-based projections in Fig. 6.

At the individual model level, WorldClim–GFDL projections illustrated the extent of potential contraction, with total range change exceeding 70% by mid-century under SSP5 and showing minimal compensatory habitat gain. In contrast, CHELSA–GFDL projections indicated more moderate changes under SSP1, with larger proportions of stable habitat and, in some periods, net gains in suitable area. Comparable patterns were observed for IPSL and MPI models (Table 1), indicating that dataset-driven differences were consistent across GCMs rather than driven by a single climate model.

Overall, SRC analyses demonstrate that projected future distributions of *C. sativa* at the national scale are highly sensitive to the choice of climate dataset, with WorldClim 2.1 and CHELSA 2.1 producing divergent estimates of both the magnitude and spatial configuration of habitat change under future climate scenarios.

Table 1
Expected Species Range Changes (SRC) in the distribution of Sweet Chestnut according to given future projections (GCMs) and time periods.

Dataset	Global Circulation Models (GCM)	Period	Scenario	Percentage Loss	Percentage Gain	Area Change (%)	Habitat Loss (ha)	Not Suitable (ha)	Stable Habitat (ha)	Habitat Gain (ha)
chelsa2.1	GFDL	2011–2040	ssp1 (2.6)	32.952	3.843	-29.109	3035645.9	69026797	6177463.1	354404.9
chelsa2.1	GFDL	2011–2040	ssp5 (8.5)	31.118	11.418	-19.7	287001	68323608	6343091	1057594
chelsa2.1	GFDL	2041–2170	ssp1 (2.6)	34.495	2.326	-32.169	3181245.8	69164998.7	6031863.1	216203.2
chelsa2.1	GFDL	2041–2170	ssp5 (8.5)	75.835	7.057	-68.778	6993733.2	68729582.6	2219375.8	651619.3
wc2.1	GFDL	2021–2040	ssp1 (2.6)	47.216	3.95	-43.266	3212344.3	71188574.9	3539140.4	264892.1
wc2.1	GFDL	2021–2040	ssp5 (8.5)	77.437	0.916	-76.521	5246767.1	71392581.2	1504717.6	60885.6
wc2.1	GFDL	2041–2060	ssp1 (2.6)	61.221	1.584	-59.637	4159125.9	71347836.6	2592358.8	105630.3
wc2.1	GFDL	2041–2060	ssp5 (8.5)	90.616	1.31	-89.306	6125544.39	71366373.2	625940.3	87093.7
chelsa2.1	IPSL	2011–2040	ssp1 (2.6)	34.144	5.04	-29.064	3142346.3	68910689.4	6070762.6	470512.5
chelsa2.1	IPSL	2011–2040	ssp5 (8.5)	29.632	7.771	-21.861	2721015.8	68655185	6492093.1	726016.9
chelsa2.1	IPSL	2041–2170	ssp1 (2.6)	47.325	5.475	-41.85	4363094.6	68874464.8	4850014.4	506737.1
chelsa2.1	IPSL	2041–2170	ssp5 (8.5)	85.772	21.878	-63.895	7904598	67355004	1308511	2026198
wc2.1	IPSL	2021–2040	ssp1 (2.6)	68.008	2.719	-65.288	4613823.8	71272148.8	2137661	181318.1
wc2.1	IPSL	2021–2040	ssp5 (8.5)	93.113	2.73	-90.383	6292105.7	71271858.4	459379.1	181608.5
wc2.1	IPSL	2041–2060	ssp1 (2.6)	77.018	2.694	-74.324	5214401.2	71273900	1537083.5	179566.9
wc2.1	IPSL	2041–2060	ssp5 (8.5)	99.934	3.064	-96.87	6747100,1	71249908,6	4384,641	203558,3
chelsa2.1	MPI	2011–2040	ssp1 (2.6)	23.393	7.573	-15.82	2150056.1	68679715.1	7063052,8	701486.7
chelsa2.1	MPI	2011–2040	ssp5 (8.5)	23.051	15.799	-7.252	2125001	67925638	7088108	1455564
chelsa2.1	MPI	2041–2170	ssp1 (2.6)	28.95	10.859	-18.091	2657519	68377045	6555590	1004157
chelsa2.1	MPI	2041–2170	ssp5 (8.5)	92.165	0.021	-92.144	8502618,8	69379320,5	710490,8	1881,8
wc2.1	MPI	2021–2040	ssp1 (2.6)	59.258	1.276	-57.981	4022614.3	71368250.8	2728870.4	85216.1
wc2.1	MPI	2021–2040	ssp5 (8.5)	75.634	1.729	-73.905	5123111.2	71338252.1	1628373.5	115214.9
wc2.1	MPI	2041–2060	ssp1 (2.6)	42.581	3.561	-39.02	2904637	71215564	3846848	237903
wc2.1	MPI	2041–2060	ssp5 (8.5)	87.057	2.122	-84.935	5888155.1	71312259.5	863329.7	141207.5

3.2. Model Performances and Distribution Patterns of Regional-scale SDMs

Regional-scale species distribution models were developed at 120 m resolution using downscaled climatic predictors and fine-scale topographic and hydrological variables within a hierarchical framework. Model performance across algorithms was consistently high (Fig. 7a), with Area Under the Curve (AUC) values ranging from 0.95 to 0.99, True Skill Statistic (TSS) values between 0.80 and 0.85, and Boyce Index values generally spanning 0.60–0.85 across GBM, GLM, and RF models. Ensemble models were constructed using only those individual models exceeding a TSS threshold of 0.75, ensuring robust predictive performance.

Spatial patterns of habitat suitability differed markedly among regional-scale model variants (Fig. 8). Models based solely on topographic predictors emphasized terrain-driven gradients, while climate-only models highlighted broader suitability patterns inherited from downscaled macroclimatic conditions. The hierarchical model, which integrated the climate-driven suitability surface with fine-scale topographic and hydrological predictors, produced the most spatially heterogeneous and ecologically structured predictions. Compared to the national-scale model based on WorldClim 2.1 at 1 km resolution (Fig. 8e), regional-scale outputs revealed finer differentiation of suitable and unsuitable habitats, particularly along elevation gradients, slope transitions, and riparian corridors.

Variable importance analyses (Fig. 7b) and response curves (Fig. 9) indicated consistent and ecologically interpretable relationships across algorithms and ensemble models. Direct Insolation showed a strong negative association with habitat suitability, while Minimal Curvature exhibited a weakly negative response. Slope displayed a unimodal relationship, with peak suitability occurring at moderate to steep inclinations. Vertical Distance to Channel Network (VDCN) demonstrated a threshold-like response, with suitability increasing at intermediate distances from stream channels before stabilizing.

The hierarchical ensemble model (Fig. 7c) showed the closest spatial correspondence with known *Castanea sativa* distribution patterns derived from forest management and inventory records, particularly in mountainous and hydrologically structured landscapes. Compared to topographic-only or climate-only regional models, the hierarchical approach reduced overprediction in low-elevation areas and improved spatial coherence in complex terrain.

4. DISCUSSION

SDMs function as a conceptual laboratory for testing macroecological theory, particularly in relation to how scale, environmental predictors, and data resolution influence inferred species–environment relationships (Guisan & Thuiller, 2005; Franklin, 2010; Peterson et al., 2011). Continental- and national-scale SDMs, which are often favored due to broader data availability, primarily capture ecological processes operating at macroclimatic scales and are valuable for identifying large-scale distributional patterns and formulating general hypotheses (Guisan & Thuiller, 2005; Franklin, 2010; Rathore & Sharma, 2023). However, their utility for ecological interpretation and management planning in heterogeneous forest landscapes depends critically on how uncertainty and scale-dependence are addressed.

The use of bioclimatic variables in large-scale SDMs offers clear advantages over raw climate predictors, as variables such as temperature and precipitation seasonality summarize ecologically meaningful aspects of climate while reducing dimensionality and collinearity. This facilitates the representation of climatic niches with fewer predictors and enhances model interpretability (Elith et al., 2010). Nevertheless, SDMs based exclusively on bioclimatic variables remain limited in their ability to capture fine-scale ecological constraints, including topography-mediated microclimates, hydrological processes, and edaphic conditions. These omissions constitute a major source of uncertainty in SDMs and have been widely discussed in the literature (Beaumont et al., 2005, 2008; Araújo & Guisan, 2006; Elith & Leathwick, 2009).

Addressing uncertainty through a hierarchical, multi-scale framework

This study explicitly addressed several key sources of SDM uncertainty by combining national- and regional-scale modeling within a hierarchical framework. Environmental uncertainty was reduced by integrating high-resolution topographic and hydrological predictors into regional-scale models, allowing macroclimatic suitability to be constrained by local habitat conditions. Occurrence-related uncertainty was minimized by restricting presence data to pure, mature *Castanea sativa* stands derived from forest inventories and herbarium records, and by applying spatial thinning to reduce autocorrelation. Algorithmic uncertainty was addressed by ensemble modeling across three complementary algorithms (RF, GBM, GLM), retaining only well-performing models (TSS > 0.75). Spatial and temporal transferability were further improved through spatial block cross-validation, which limits optimistic bias arising from spatially clustered samples (Roberts et al., 2017; Valavi et al., 2019).

Uncertainty associated with future climate projections remains one of the most challenging aspects of SDM-based forecasting. Structural differences among Global Circulation Models (GCMs) and emission scenarios can lead to substantial divergence in projected outcomes (Beaumont et al., 2008). Following recommendations by Buisson et al. (2010), this study incorporated multiple GCMs (GFDL, IPSL, MPI) and scenarios within an ensemble framework, allowing dataset- and model-driven variability to be quantified rather than ignored. The pronounced divergence observed between WorldClim 2.1– and CHELSA 2.1–based projections highlights the importance of climate dataset choice as a critical, and often underappreciated, source of uncertainty.

Ecological interpretation of current distributions and predictor responses

Differences between WorldClim 2.1 and CHELSA 2.1 projections reflect not only spatial resolution but also methodological differences in climate interpolation. CHELSA explicitly incorporates elevation, wind fields, and orographic effects, which can improve precipitation estimates in complex terrain (Karger et al., 2017; Brun et al., 2022). In national-scale models, WorldClim-based predictions were primarily structured by temperature seasonality (bio4), mean temperature of the warmest quarter (bio10), annual precipitation (bio12), and precipitation seasonality (bio15). CHELSA-based models emphasized isothermality (bio3), temperature seasonality (bio4), mean temperature of the coldest quarter (bio11), and precipitation of the coldest quarter (bio19).

Response curves for these variables were ecologically coherent and consistent with existing literature. Temperature seasonality emerged as a dominant constraint, with *C. sativa* exhibiting optimal suitability under mild thermal variability, characteristic of maritime-influenced climates. Precipitation-related variables highlighted the species' reliance on sufficient moisture availability, particularly during winter and early spring. These patterns broadly align with findings from the Caucasus region (Metreveli et al., 2023) and reflect well-documented physiological sensitivities to summer drought stress, winter cold extremes, and shifts in seasonal water availability under climate change (Freitas et al., 2021).

Differences in variable importance between studies likely reflect geographic scope and underlying genetic differentiation among populations, as documented by Chiocchini et al. (2016). This interpretation is further supported by Castellana et al. (2021), who demonstrated clear signatures of local adaptation to climate in natural populations of *Castanea sativa* across southern Europe. Their results indicate that chestnut populations exhibit genetically structured responses to temperature and precipitation gradients, suggesting that regional climatic sensitivity detected by SDMs may partly reflect adaptive differentiation rather than purely plastic responses. Collectively, these findings highlight that macroclimatic suitability inferred from national-scale models represents a necessary but incomplete condition for species persistence, capturing broad physiological limits while remaining insensitive to fine-scale environmental filtering.

Against this backdrop, regional-scale modeling becomes essential for resolving how local terrain and hydrological context constrain the expression of climatically suitable niches. At the regional scale, topographic variables provided critical ecological refinement. Slope emerged as the most influential predictor, with optimal suitability occurring on moderately steep terrain (30–70%), consistent with field-based habitat syntheses and known preferences for well-drained, shaded slopes (Çepel, 1995; Gallardo L.J., 2001). Minimal Curvature and Vertical Distance to Channel Network jointly captured landscape position, indicating preference for concave mid-slope positions at intermediate distances from streams—locations that balance moisture availability with reduced risk of waterlogging and disease. Direct Insolation, while of lower statistical importance, contributed to explaining regional variation in microhabitat use, particularly along climatic gradients from humid Black Sea forests to drier Marmara landscapes.

Scale dependence and comparison of national- and regional-scale SDMs

Direct comparison of national- and regional-scale SDMs illustrates the consequences of scale choice for habitat inference. National-scale models provide a valuable overview of climatically permissive regions across Türkiye but inevitably smooth local environmental gradients, resulting in broader suitability patterns. This scale-induced smoothing can be interpreted as a form of niche distortion driven by spatial averaging, conceptually related to recent discussions of scale effects and niche truncation in SDMs (Goicolea et al., 2024), though expressed differently here. In this study, coarse-resolution models primarily inflated suitability by overlooking microhabitat constraints rather than underestimating environmental tolerances in a strict truncation sense.

Regional-scale models, by contrast, constrained macroclimatic suitability using fine-resolution topographic and hydrological predictors, producing more spatially explicit and ecologically realistic patterns. When a suitability threshold of 0.4 was applied (based on ensemble optimization), national-scale models predicted approximately 3,757 km² of suitable habitat within the Marmara region, whereas regional-scale models identified only 1,508 km²—a reduction of nearly 60%. This contraction reflects the exclusion of lowland and environmentally marginal areas and the concentration of suitability in topographically mediated refugia such as north-facing slopes and mid-elevation zones of the Uludağ Massif and Aydın Mountains. The closer correspondence of hierarchical regional models with forest inventory records supports the added value of nested modeling approaches in complex forest landscapes.

Implications for SDM practice and forest management

Beyond their biogeographical relevance, SDM outputs gain applied value when interpreted in the context of forest management and socio-ecological systems. In Türkiye, *Castanea sativa* is not only a forest tree species but also a cornerstone of multifunctional agroforestry systems and rural livelihoods (Wall et al., 2019). Recent work by Wall et al. (2025) emphasizes the need for unsiloed chestnut management strategies that integrate ecological suitability with livelihood priorities and policy frameworks. In this context, the hierarchical SDMs presented here provide spatially explicit guidance for identifying areas where ecological conditions are most compatible with sustainable chestnut management, thereby supporting place-based decision-making under climate change.

Taken together, these results demonstrate that both climate dataset choice and modeling scale strongly influence SDM outcomes for climate-sensitive forest species. While national-scale SDMs remain indispensable for broad-scale assessments and future scenario analysis, their outputs should be interpreted cautiously in heterogeneous terrain. Hierarchical, multi-scale frameworks provide a practical pathway for reconciling macroclimatic projections with local ecological constraints, improving both ecological realism and decision relevance for adaptive forest management and conservation planning.

5. CONCLUSIONS

This study demonstrates that algorithm choice, spatial scale, and climate dataset selection jointly shape species distribution model (SDM) outcomes for *Castanea sativa*. Among individual algorithms, Random Forest produced the most conservative predictions, while Generalized Linear Models tended to overestimate suitable habitat; Gradient Boosting Models yielded intermediate results. Because no single algorithm can be assumed to best represent ecological reality, the ensemble approach adopted here provides a more robust and ecologically defensible estimate of potential distribution by minimizing algorithm-specific bias. These findings reinforce ensemble modeling as a best-practice strategy for SDMs in heterogeneous forest landscapes.

Comparisons between national- and regional-scale models further highlight the importance of spatial scale. National-scale SDMs effectively capture broad climatic suitability patterns but tend to overgeneralize distributions by smoothing local environmental gradients. In contrast, regionally calibrated models that incorporate high-resolution topographic and hydrological predictors yield more spatially precise and ecologically constrained suitability patterns. This scale-dependent refinement underscores the value of hierarchical modeling frameworks for translating macroclimatic suitability into realistic habitat representations relevant for forest management and conservation planning.

Climate dataset choice emerged as a critical determinant of SDM outcomes, with terrain-aware datasets like CHELSA potentially offering advantages in orographically complex regions for precipitation-related variables. Future regional SDMs would benefit from systematic comparison of both WorldClim- and CHELSA-based predictors, or from hybrid approaches leveraging the strengths of each dataset.

Looking forward, extending regional-scale SDMs to future climate projections represents a logical next step. Downscaling GCM outputs using methods such as Geographically Weighted Regression or machine learning–based interpolation, explicitly linked to local topographic covariates, may substantially improve the spatial realism of future projections. Incorporating dynamic environmental and biotic variables—such as soil moisture, pathogen pressure, or pest dispersal—would further enhance the capacity of SDMs to assess compounded climate and biological threats, supporting adaptive management strategies for *C. sativa*.

Beyond single-species applications, SDMs offer substantial potential for multispecies and community-level analyses. Modeling the co-occurrence of ecologically associated species could help identify shared environmental constraints, biodiversity hotspots, and functionally important habitats. For *C. sativa*, integrating distributions of co-occurring tree species or ecologically linked taxa, such as disperser birds or keystone conifers, could yield more holistic insights into forest ecosystem dynamics under changing environmental conditions.

Finally, the central objective of this study was to ensure methodological transparency and reproducibility. By documenting all modeling steps in accordance with the ODMAP framework (Zurell et al., 2020), this work facilitates reproducibility, comparison, and future synthesis across SDM studies. Such standardized reporting is essential for advancing SDMs from predictive tools toward cumulative, theory-informed ecological science.

Declarations

Conflicts of interest/Competing interests

The authors declare no conflicts of interest.

Funding

This study originates from the PhD dissertation of Hüseyin Doğan and was supported by the Scientific Research Projects Coordination Unit of Istanbul University–Cerrahpaşa (Project No: 35792).

Author Contribution

O.Y.Y. conceived and designed the study, provided supervision and guidance, contributed to manuscript revision and editing, and was responsible for funding acquisition and project administration. H.D. contributed to data collection and curation, performed the modeling and analyses, and wrote the first draft of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgement

We gratefully acknowledge Prof. Dr. Orhan Sevgi and Prof. Dr. Gülriz Bayçu Kahyaoğlu for serving as co-advisors during the PhD research that formed the basis of this study. We thank Prof. Dr. Hatice Yılmaz for providing critical guidance at key stages of the research, and Prof. Dr. Ünal Akkemik for mentorship in forest ecology and for facilitating research opportunities throughout Dr. Hüseyin Doğan's doctoral program.

Data Availability

All data used in this study are publicly available. Processed data and model outputs generated during the research are available on Zenodo (https://doi.org/10.5281/zenodo.16882976) and on GitHub (https://github.com/ozgurhsyndgn/Phd).

References

1. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38(5):541–545
2. Araujo MB, Guisan A (2006) Five testing strategies for circumscribed predictions of species distributions. Ecography 29(5):699–708
3. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3(2):327–338
4. Beaumont LJ, Hughes L, Poulsen M (2005) Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecol Model 186(2):251–270
5. Beaumont LJ, Hughes L, Pitman AJ (2008) Why is the choice of future climate scenarios for species distribution modelling important? Ecol Lett 11:1135–1146
6. Brun P, Zimmermann N, Hari C, Pellissier L, Karger DN (2022) Global climate-related predictors at kilometer resolution for the past and future. Earth Syst Sci Data 14:5573–5603. 10.5194/essd-14-5573-2022
7. Brunsdon C, Fotheringham AS, Charlton M (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Anal 28(4):281–298
8. Buisson L, Thuiller W, Casajus N, Lek S, Grenoulliet G (2010) Uncertainty in ensemble forecasting of species distribution. Glob Change Biol 16:1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x
9. Castellana S, Martin MÁ, Solla A et al (2021) Signatures of local adaptation to climate in natural populations of sweet chestnut (Castanea sativa Mill.) from southern Europe. Ann For Sci 78:27. https://doi.org/10.1007/s13595-021-01027-6
10. Çepel N (1995) Forest Ecology. Department of Soil Science and Ecology, Faculty of Forestry. Istanbul University, Istanbul. (in Turkish)
11. Chiocchini F, Mattioni C, Pollegioni P, Lusini I, Martín MA, Cherubini M, Lauteri M, Villani F (2016) Mapping the Genetic Diversity of Castanea sativa: Exploiting Spatial Analysis for Biogeography and Conservation Studies. J Geographic Inform Syst 08:248–259. 10.4236/jgis.2016.82022
12. Conedera M, Krebs P, Tinner W (2004) The cultivation of Castanea sativa (Mill.) in Europe, from its origin to its diffusion on a continental scale. Veget Hist Archaeobot 13:161–179. https://doi.org/10.1007/s00334-004-0038-7
13. Conedera M, Manetti MC, Tinner W, Hubalek Z, Giacometti M (2016) Chestnut (*Castanea sativa* Mill.): origin, history, and culture. Genet Resour Crop Evol 63(5):873–899
14. Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci Model Dev 8:1991–2007. 10.5194/gmd-8-1991-2015
15. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling D, W. and, Kühn I (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628
16. Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40(1):677–697
17. Elith J, Kearney M, Phillips Sfg (2010) The art of modeling range-shifted species. Methods Ecol Evol 1:330–342. 10.1111/j.2041-210X.2010.00036.x
18. FAO (2020) *Global Forest Resources Assessment 2020: Key findings*. Rome
19. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315
20. Franklin J (2010) Mapping species distributions: spatial inference for ecological evaluation. Cambridge University Press
21. Franklin J (2013) Species distribution models in conservation biogeography: developments and challenges. Divers Distrib 19:1217–1223. https://doi.org/10.1111/ddi.12125
22. Freitas TR, Santos JA, Silva AP, Fraga H (2021) Influence of climate change on chestnut trees: A review. Plants 10(7):1463
23. Gallardo LJ (2001) Distribution of chestnut (Castenea sativa Mill.) forests in Spain: Possible ecological criteria for quality and management (focusing on timber coppices). Snow Landsc Res 76:477–481
24. Goicolea T, Adde A, Broennimann O, García-Viñas JI, Gastón A, José, Aroca-Fernández M, Guisan A, Mateo G (2024) R. Spatially-nested hierarchical species distribution models to overcome niche truncation in national-scale studies. Ecography, 2025: e07328. https://doi.org/10.1111/ecog.07328
25. Gueguen M, Blancheteau H, Thuiller W (2025) biomod2: Ensemble Platform for Species Distribution Modeling. R package version 4.2-6-2. https://biomodhub.github.io/biomod2/
26. Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2–3):147–186

27. Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x
28. Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI, Buckley YM (2013) Predicting species distributions for conservation decisions. Ecol Lett 16(12):1424–1435
29. Hanewinkel M, Cullmann DA, Schelhaas MJ, Nabuurs GJ, Zimmermann NE (2013) Climate change may cause severe loss in the economic value of European forest land. Nat Clim change 3(3):203–207
30. Hirzel HA, Le lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Model 199(2):142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017
31. IPCC (2021) Climate Change 2021: The Physical Science Basis. Cambridge University Press
32. Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., …Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Scientific Data, 4(1), 1–20.
33. Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo,J., … Marchetti, M. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest ecology and management, 259(4), 698–709.
34. Mateo RG, Gaston A, Aroca-Fernandez MJ, Broennimann O, Guisan A, Saura S, Garcia-Vinas JI (2019) Hierarchical species distribution models in support of vegetation conservation at the landscape scale. J Veg Sci 30:386–396
35. Metreveli V, Kreft H, Akobia I, Janiashvili Z, Nonashvili Z, Dzadzamia L, Javakhishvili Z, Gavashelishvili A (2023) Potential Distribution and Suitable Habitat for Chestnut (Castanea sativa). Forests 2023, 14, 2076. https://doi.org/10.3390/f14102076
36. Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling. Ecography 37:191–203. 10.1111/j.1600-0587.2013.00205.x
37. NASA/METI (2009) ASTER Global Digital Elevation Model (GDEM) Version 2. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/ASTER/ASTGTM.002
38. O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41(5):673–690. https://doi.org/10.1007/s11135-006-9018-6
39. OGM (2012) Chestnut Action Plan (2013–2017). Ministry of Forestry and Water Affairs, General Directorate of Forestry, Department of Silviculture. (in Turkish) (https://github.com/ozgurhsyndgn/Phd/blob/main/KESTANE%20EYLEM%20PLANI.pdf)
40. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB (2011) Ecological niches and geographic distributions (MPB-49). Princeton University Press
41. Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K.M., … McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). Methods in Ecology and Evolution, 5(5), 397–406.
42. R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
43. Rathore MK, Sharma LK (2023) Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. Biodivers Conserv 32:3053–3087. https://doi.org/10.1007/s10531-023-02648-1
44. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G.,… Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial,hierarchical, or phylogenetic structure. Ecography, 40(8), 913–929. https://doi.org/10.1111/ecog.02881
45. Sequeira AMM, Bouchet PJ, Yates KL, Mengersen K, Caley MJ (2018) Transferring biodiversity models for conservation: Opportunities and challenges. Methods Ecol Evol 9(5):1250–1264. https://doi.org/10.1111/2041-210X.12998
46. Thuiller W, Lavorel S, Araújo MB, Sykes MT, Prentice IC (2005) Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences, 102(23), 8245–8250. https://doi.org/10.1073/pnas.0409902102
47. Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD–a platform for ensemble forecasting of species distributions. Ecography 32(3):369–373
48. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol Evol 10(2):225–232
49. Villani F, Pigliucci M, Cherubini M (1994) Evolution of Castanea sativa mill, in Turkey and Europe. Genet Res 63(2):109–116
50. Villani F, Sansotta A, Cherubini M, Cesaroni D, Sbordoni V (2004) Genetic structure of natural populations of Castanea sativa in Turkey: evidence of a hybrid zone. J Evol Biol 17(6):1380–1394
51. Wall JR, Okan T, Köse C, Köse N, Aksoy EB (2019) Folk biological value and chestnut conservation in Turkey. Econ Bot 73(4):461–476
52. Wall J, Okan T, Köse C, Köse N, Aksoy EB (2025) Unsiloed agroforestry research and policy: Livelihood and multifunction as chestnut (Castanea sativa) management priorities for Türkiye. For Policy Econ 173:103474
53. Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell,S., … Elith, J. (2018). Outstanding challenges in the transferability of ecological models. Trends in Ecology & Evolution, 33(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

54. Yilmaz H, Yilmaz OY, Akyüz YF (2017) Determining the factors affecting the distribution of Muscari latifolium, an endemic plant of Turkey, and a mapping species distribution model. Ecol Evol 7(4):1112–1124
55. Yilmaz OY, Akkemik Ü, Dogan ÖH, Yilmaz H, Sevgi O, Sevgi E (2024) The missing part of the past, current, and future distribution model of Quercus ilex L.: the eastern edge. iForest 17:90–99. 10.3832/ifor4350-016
56. Zevenbergen LW, Thorne CR (1987) Quantitative analysis of land surface topography. Earth Surf Proc Land 12:47–56. 10.1002/esp.3290120107
57. Zurell, D., Franklin, J., König, C., Bouchet, P. J., Serra-Diaz, J. M., Dormann, C.F., … Merow, C. (2020). A standard protocol for reporting species distribution models.Ecography, 43(8), 1261–1277. https://doi.org/10.1111/ecog.04960

Figures

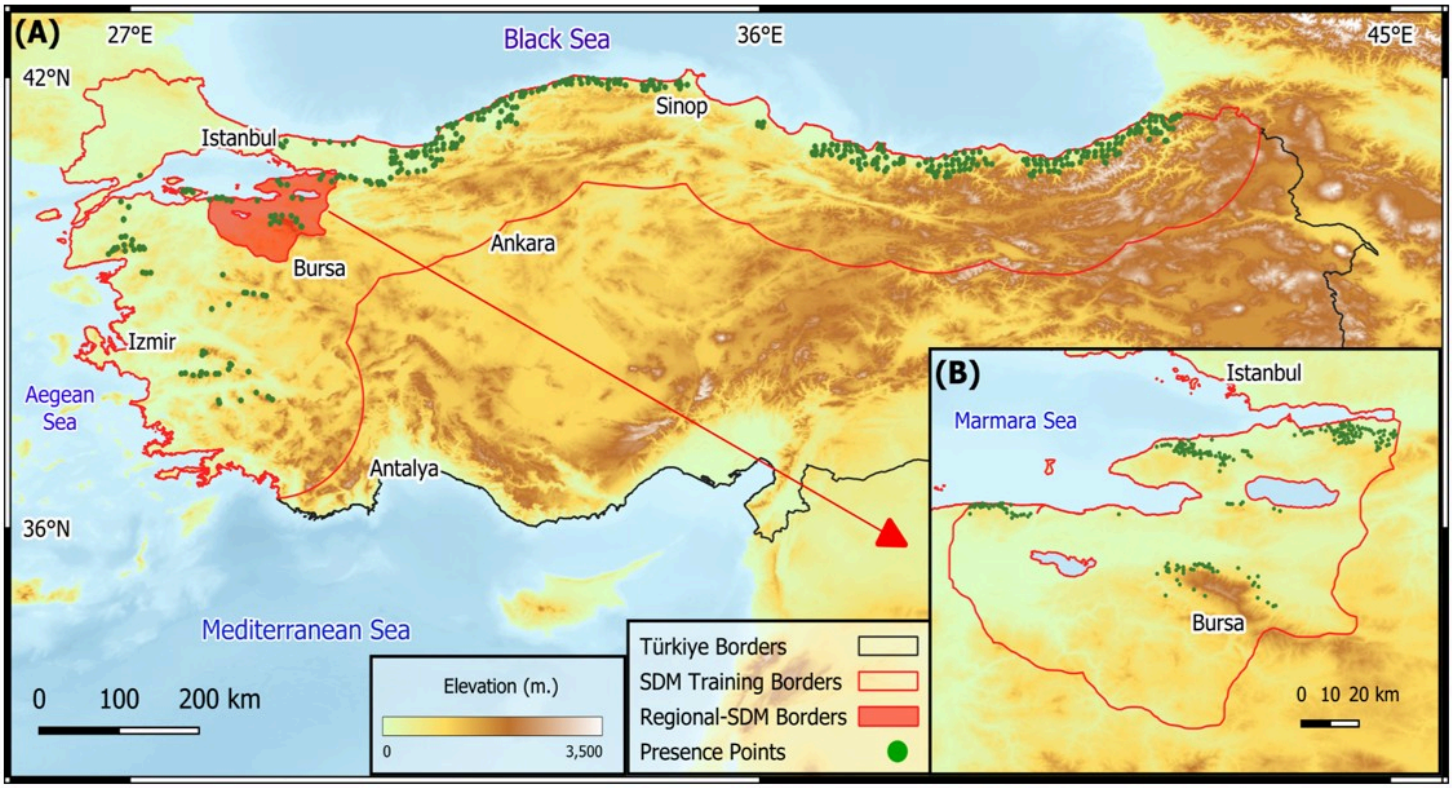


Figure 1

(A) Dark-green dots represent presence points used in both national-scale and regional-scale models. The red line indicates the area used to train the national-scale SDMs, which were later extrapolated to the entire country borders. (B) A zoomed-in view of the regional-scale study area in the Marmara region, where dark-green points represent regional-scale presence data and the solid red line outlines the regional SDM training extent.

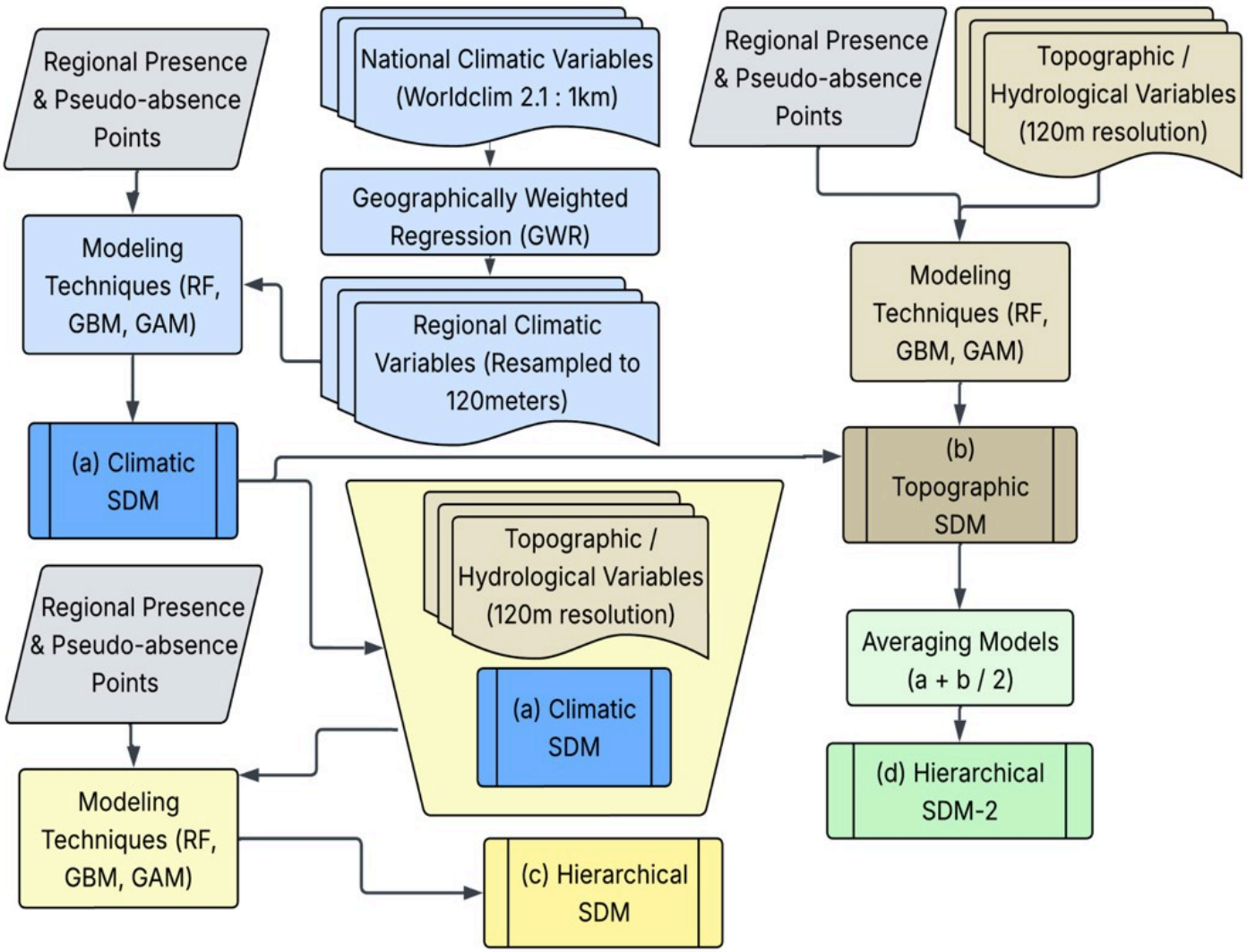


Figure 2

Workflow diagram of the hierarchical species distribution modeling (regional-scale SDM) approach, adapted and modified from Mateo et al. (2019).

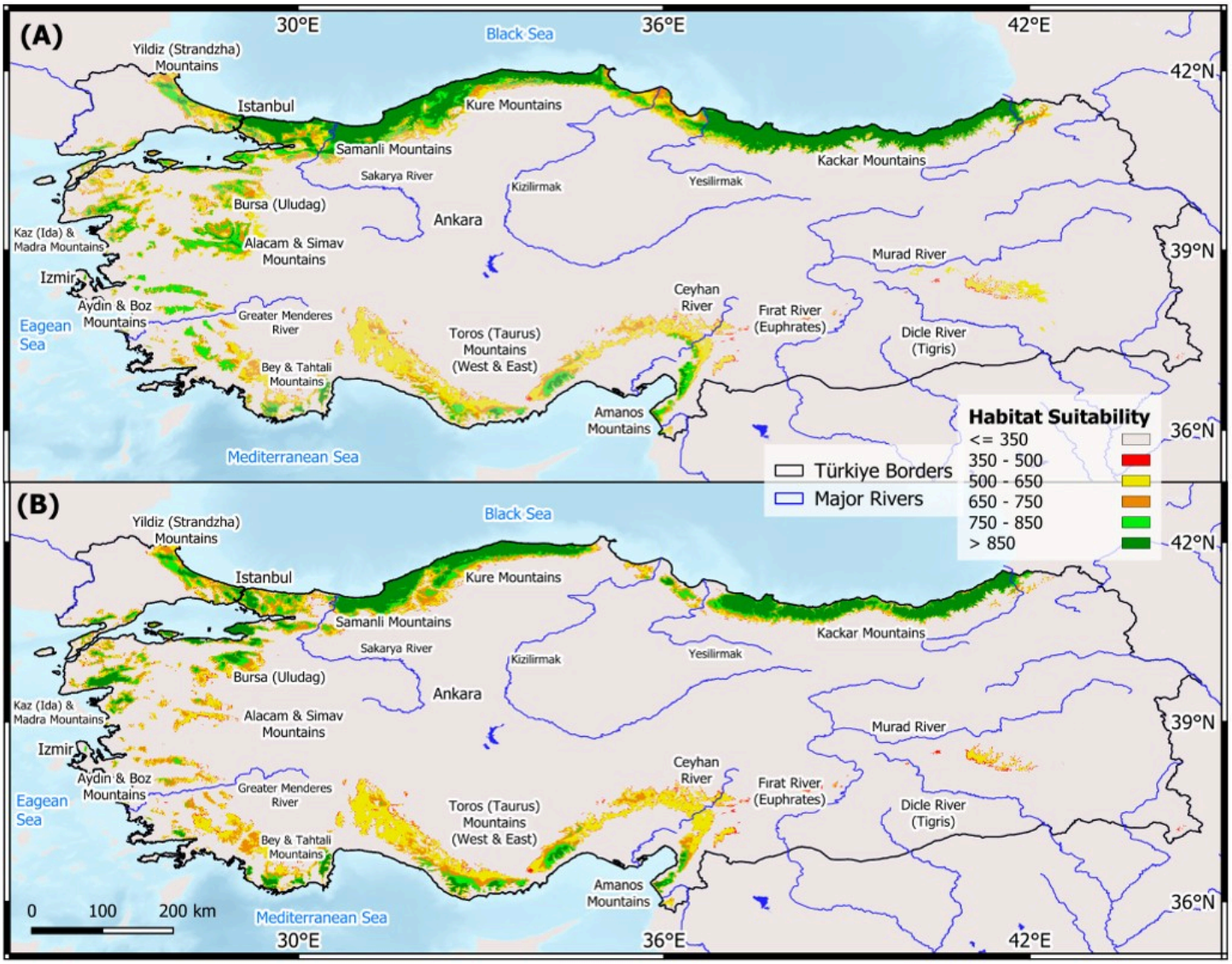


Figure 3

Predicted habitat suitability for Castanea sativa under current climate conditions using two different climate datasets in national-scale ensemble SDMs using 3 different algorithms (Ensemble-SDMs): (A) WorldClim 2.1 and (B) CHELSA 2.1.

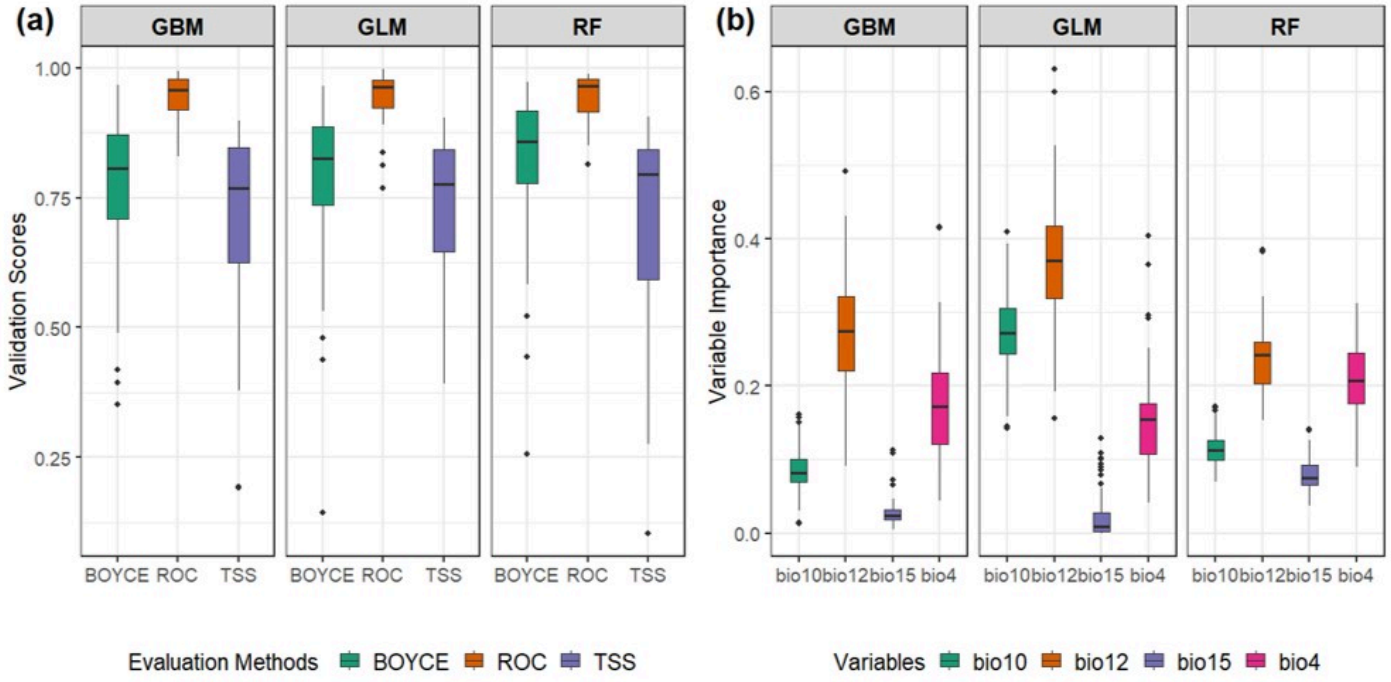


Figure 4

(a) Validation scores (Boyce Index, AUC, and TSS) for Castanea sativa distribution models built using the Worldclim 2.1 climate dataset across three algorithms. (b) Variable importance values for the four selected bioclimatic predictors from WorldClim 2.1: bio10 (mean temperature of warmest quarter), bio12 (annual precipitation), bio15 (precipitation seasonality), and bio4 (temperature seasonality).

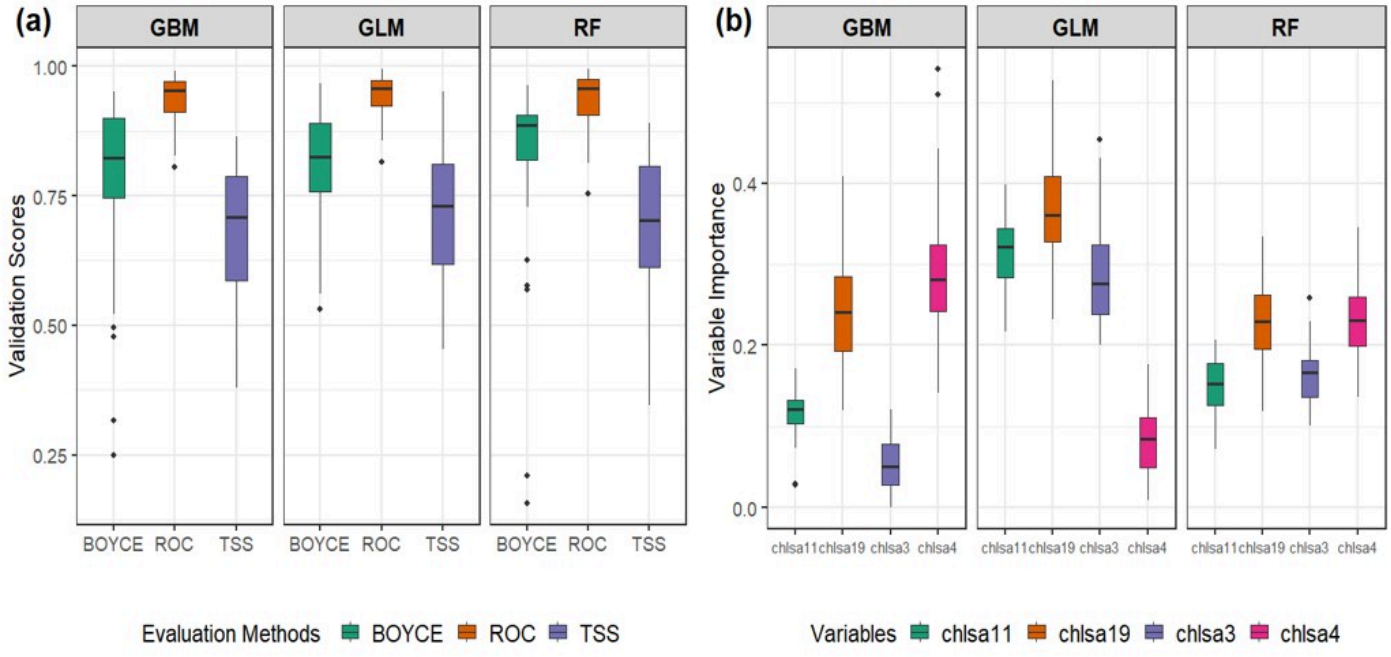


Figure 5

(a) Validation scores (Boyce Index, AUC, and TSS) for Castanea sativa distribution models built using the CHELSA 2.1 climate dataset across three algorithms. (b) Variable importance values for the four CHELSA-derived bioclimatic predictors (chlsa11: mean temperature of coldest quarter, chlsa19: precipitation of coldest quarter, chlsa3: isothermality, and chlsa4: temperature seasonality), as assessed by each modeling algorithm.

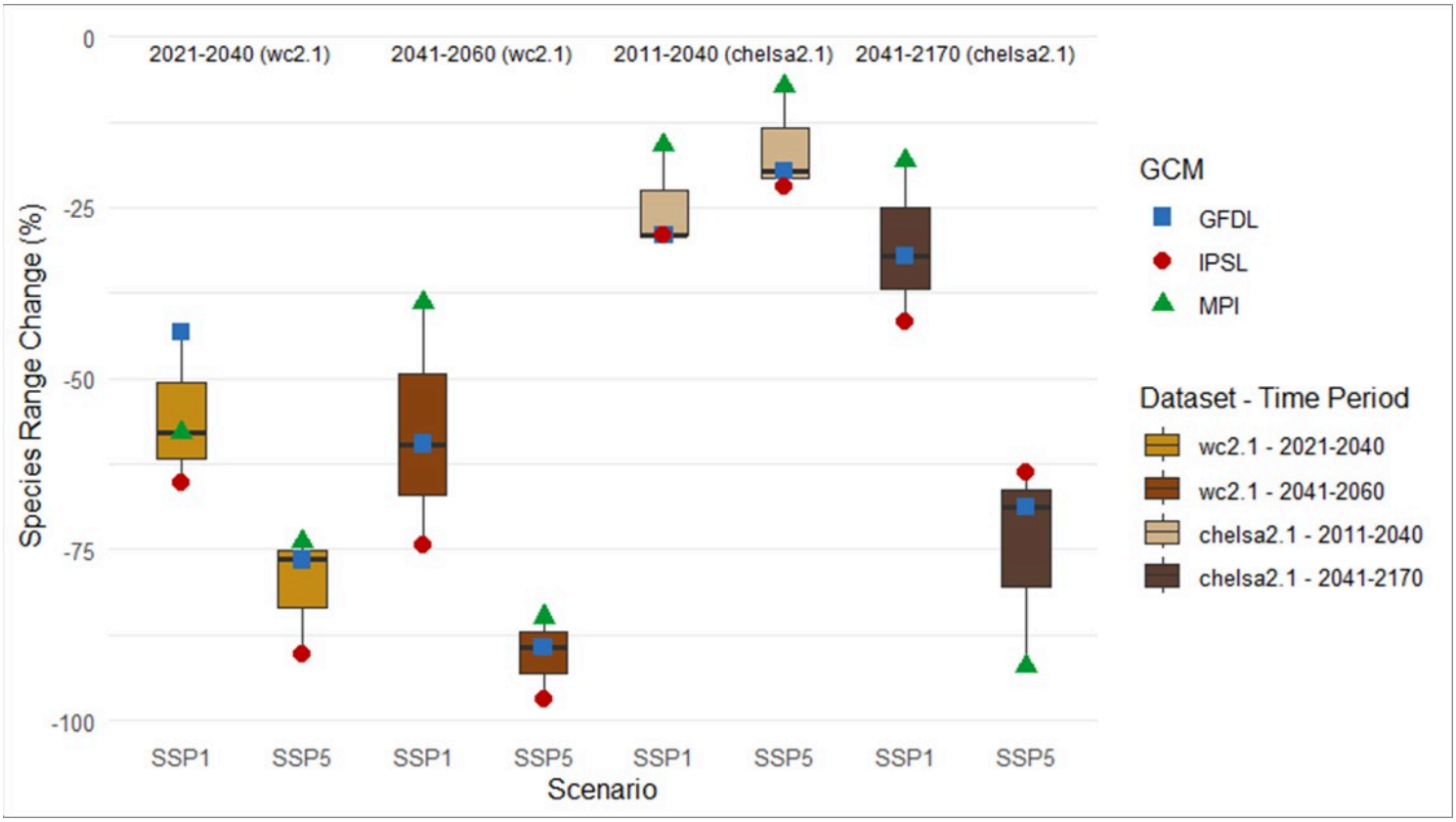


Figure 6

Boxplots depict the projected Species Range Change (SRC %) for two shared socioeconomic pathways (SSP1 and SSP5) across four time periods (2021-2040 and 2041-2060 for wc2.1 dataset in dark brown; 2011-2040 and 2041-2170 for chelsa2.1 dataset in light brown) with 3 different GCMs. The left group represents wc2.1 periods, while the right group represents chelsa2.1 periods. Shapes indicate mean values from Global Circulation Models (GCMs: GFDL in blue, IPSL in red, MPI in green), with boxes showing the interquartile range and whiskers showing the results range. Data derived from Table 1, using the Area Change (%) column to generate the boxplots, based on calculations of percentage loss and gain in habitat area across datasets, GCMs, periods, and scenarios (comparison concept adapted from Yilmaz et. al., 2024).

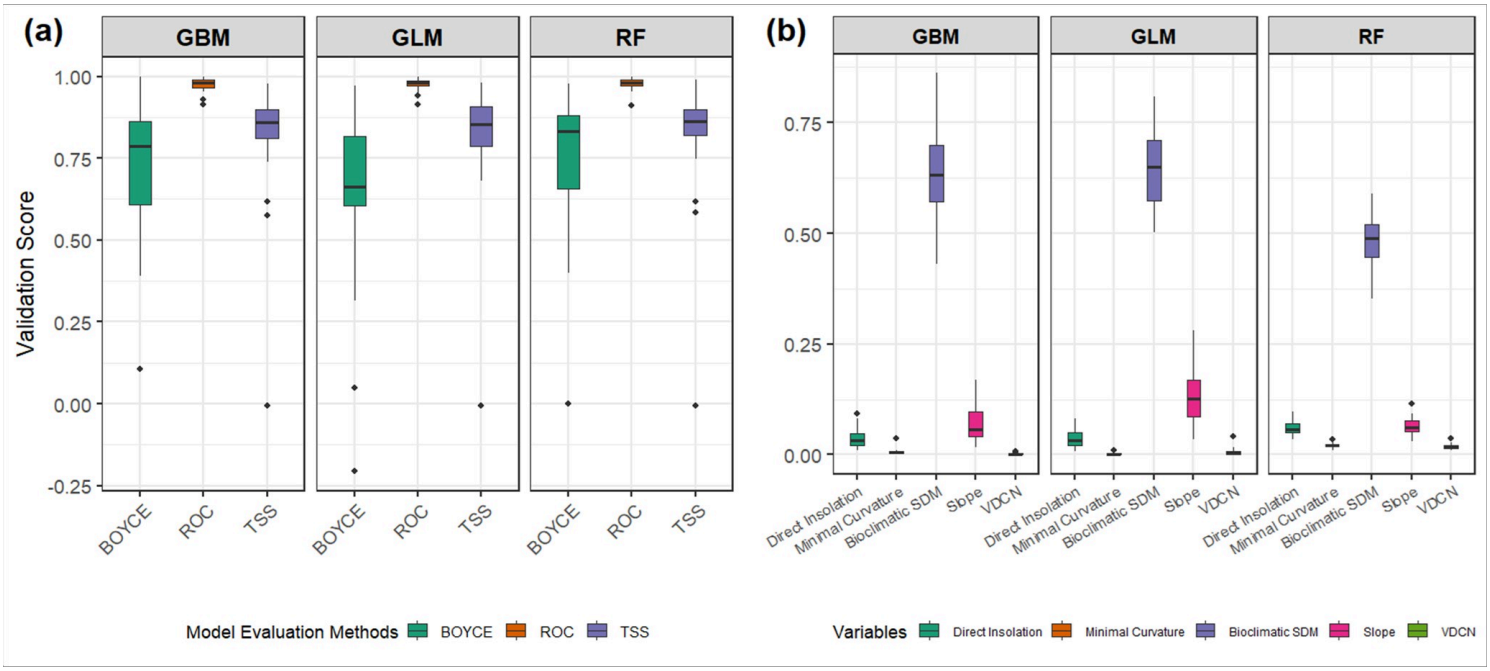


Figure 7

(a) Evaluation scores of the models created using the hierarchical method based on each algorithm and three different evaluation metrics (BOYCE, ROC, and TSS). (b) Importance scores of the variables used in the generated models.

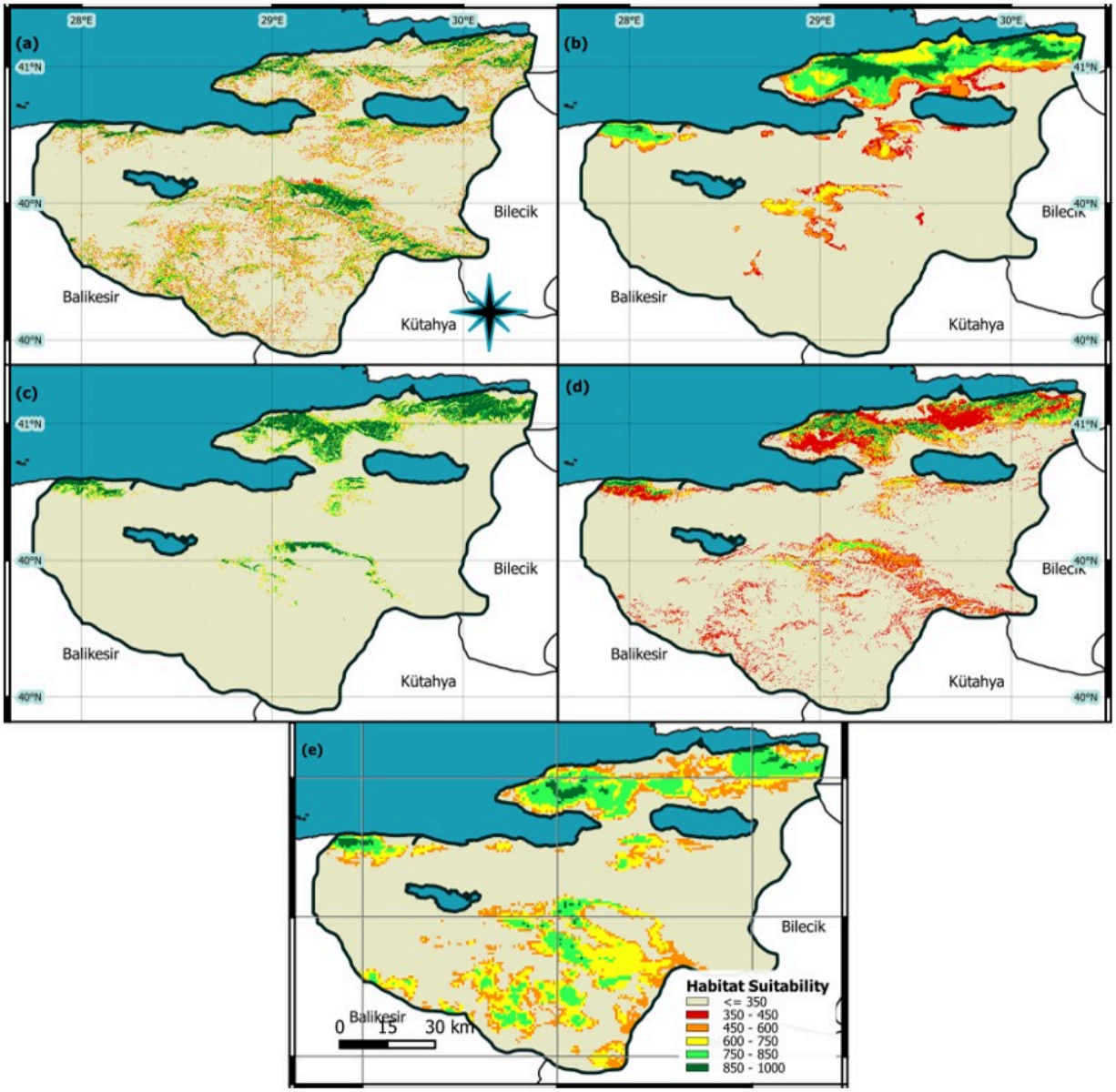


Figure 8

Current habitat suitability results produced at regional (a-b-c-d) and national (e) scales. The plots display regional-scale results generated for (a) Topographic, (b) Climatic, (c) Hierarchical, and (d) Topographic/Climatic Ensemble models, all at 120m resolution. Plot (e) shows the current habitat suitability results generated for comparison at the national scale using the WC2.1 dataset with 1km resolution.

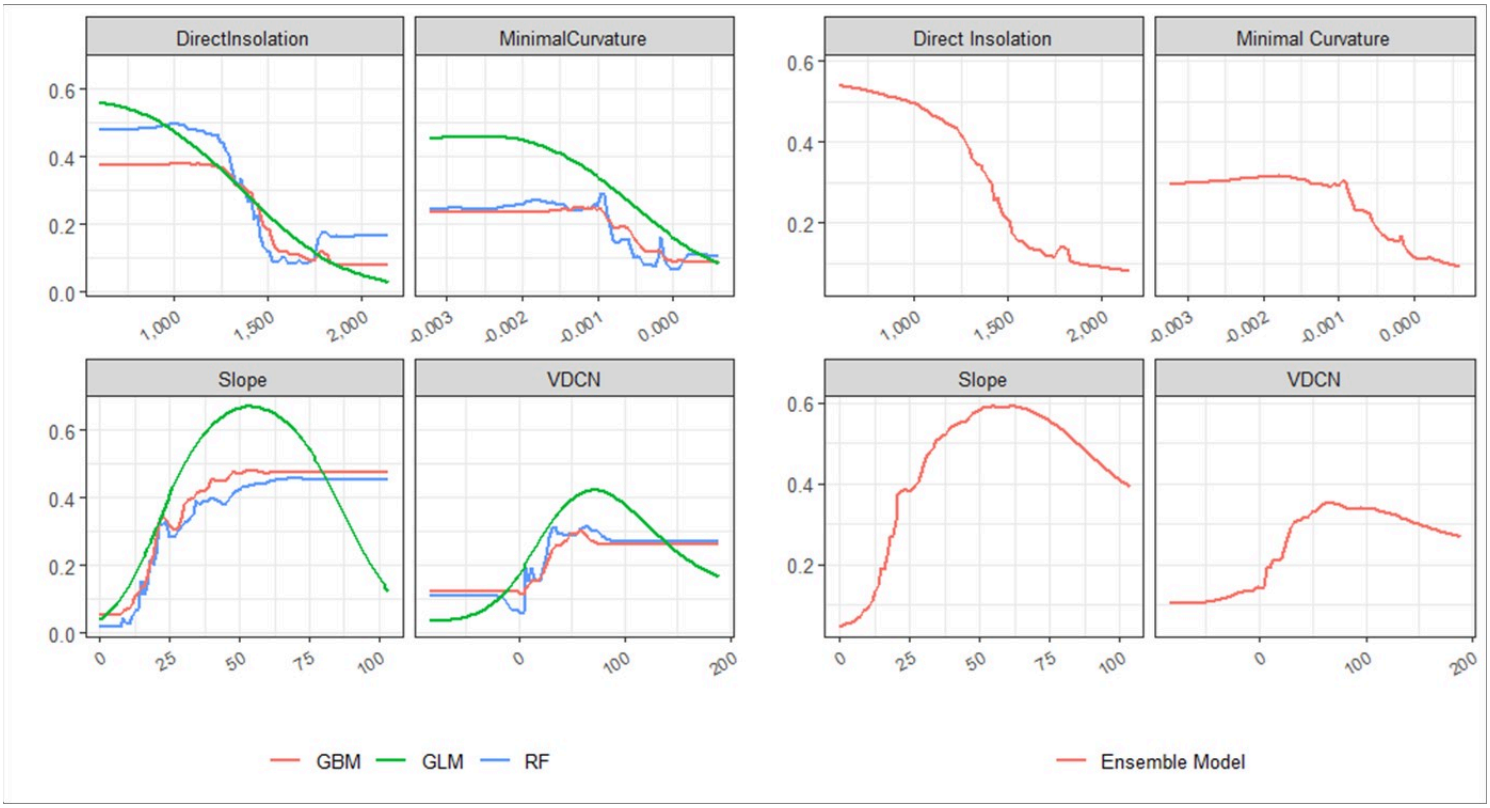


Figure 9

Response curves of variables used in regional-scale SDMs. The curves illustrate the relationship between environmental variables (Direct Insolation, Minimal Curvature, Slope, and VDCN) and the probability of species presence, showing the results for RF, GLM, and GBM algorithms (left panels) and the combined (ensemble) model (right panels).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- EPSupportingInformation.docx