International Institute for Applied Systems Analysis
IIASA
Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
Email: repository@iiasa.ac.at Telephone: +43 (0)2236 807 342

YSSP Report

Young Scientist Summer Program

Meeting 2030 biodiversity targets under land use constraints in the EU

Author: Melissa Chapman
Email: mchapman@berkeley.edu

Approved by

Supervisor: Piero Visconti
Co-Supervisor: Martin Jung
Program: Biodiversity, Ecology, and Conservation (BEC)
September 30, 2022

This report represents the work completed by the author during the IIASA Young Scientists Summer Program (YSSP) with approval from the YSSP supervisor.

It was finished by 30 September 2022 and has not been altered or revised since.

Supervisor signature:

ABSTRACT

In alignment with the global initiative to slow biodiversity loss, the European Union has committed to ambitious area-based conservation and restoration targets for the upcoming decade. Ensuring that these area-based targets are meaningfully achieved, while land requirements for commodity production and climate mitigation targets met, will require strategic planning across land uses, management measures, and jurisdictional boundaries. In this study we spatially optimize the allocation of restoration and conservation measures across different land-cover types to maximize species conservation and carbon sequestration objectives, subject to projected crop, pasture, and forestry production constraints by 2030. We show how simultaneously prioritizing landscapes for restoration and conservation can significantly improve the efficacy and coherence of area-based conservation initiatives compared to currently available disparate optimization of these measures. Additionally, we explore the impact of multilateral burden sharing and multi-objective weightings on optimal land allocation for production and the expected biodiversity and climate mitigation benefits of proposed solutions. Our results provide insight for the implementation of timely EU policy measures and a baseline for EU member states to evaluate the potential of their conservation pledges. Moreover, the methodology can easily be adopted to different contexts around the world and showcase how conservation planning can meaningfully optimize land allocation needs for both nature and people.

ACKNOWLEDGMENTS

- Dr. Piero Visconti (IIASA) and Dr. Martin Jung (IIASA) for supervision on the project formulation, methodology development and implementation.
- 'United States National Academy of Sciences' for the National Membership Organization funding.
- Dr. Carl Boettiger (University of California Berkeley) and Dr. Justin Brashares (University of California Berkeley) for helpful input on analysis.
- Matt Lewis for developing biodiversity data harmonization.
- Dr. David Leclere for providing data and thoughtful feedback on land use constraints.
- The IIASA Biodiversity, Ecology, and Conservation Group for useful feedback on methodology and data visualization.

Contents

List of Figures

List of Tables

1 Introduction

In an effort to halt biodiversity loss [1] and safeguard nature's contributions to people (NCP) [2], the Convention on Biological Diversity (CBD) is nearing the adoption a post-2020 global biodiversity framework [3]. Largely coalesced around area-based conservation targets (e.g., protecting 30% of lands and waters by 2030), this framework is influencing hundreds of national and sub-national conservation policies around the world as governments pledge to improve the protection of their biodiversity assets [4, 5]. However, area-based conservation measures are only as effective as the ecological objectives and NCPs (e.g., climate mitigation) they successfully target and address [6, 7]. The strategic implementation of these area targets, rather than meeting the area targets themselves, will shape the next decade of global biodiversity outcomes and decide whether we are able to bend the curve on biodiversity loss.

In alignment with the global post-2020 biodiversity goals, the European Union (EU) has committed to ambitious land conservation targets (EU Biodiversity Strategy for 2030). Complemented by the Nature Restoration Law (European commission, 2022), these decadal policies seek to not only expand the EU protected area network to cover at least 30% of land and waters but also enact a binding commitment to restore 20% of degraded ecosystems by 2030. With most EU land being under some form of human use and given anticipated geopolitical and climate impacts on the bio-economy, it is necessary that these area-based targets are achieved in a way so that land-use production needs are met, which will necessitate strategic planning across land use and management measures. Moreover, because biodiversity is not uniformly distributed across the EU, planning the implementation of these policies will require thoughtful consideration and intergovernmental cooperation to avoid equal burden sharing between member states of the EU.

Systematic conservation planning has long been applied to inform the spatial allocation of conservation actions [8, 9], being able to identify optimal priorities and actions so that the most complementary portfolio of species and NCPs is contained in a particular solution and managed optimally. Recent advances in optimization methods (e.g., linear programming

[10]) have allowed for exploring the optimal allocation of conservation measures at multiple scales [11], under competing objectives [12], or by subjecting the allocation to land use constraints [13]. Yet, much of these approaches have been applied in isolation and separately, and concurrently optimizing conservation and restoration allocation has to our knowledge never been attempted [12, 14]. It is furthermore necessitated by the EU policy agenda that synergies and co-benefits are optimally identified to avoid any land management portfolios to be left up to chance. For example, forest restoration in a given location might seemingly provide the best benefits to biodiversity and carbon sequestration when considered alone, however in a regional perspective it may be more preferable to allocate restoration to a another land-cover type (e.g., grassland) if this benefits species who are less ubiquitous and constrained to this region. Without planning conservation and restoration in concert, it is not easily feasible to assess whether priorities are optimally contributing to the preservation of species habitats and NCPs.

In this work, we aim to identify the areas that, if restored or conserved, would maximize the achievement of a series of targets for species conservation and carbon sequestration. Opposed to previous global conservation [12] and restoration [14] prioritizations, we estimate for the first time how and where conservation and restoration targets could be met simultaneously, while also considering existing and future constraints on land area (e.g. production needs of the EU bio-economy by 2030). We estimate how landscape-scale habitat requirements for European species of conservation concern (e.g. those in the Habitats and Birds directives) could be met by "maintaining" natural habitat (i.e., natural land cover not converted to production) rather than legal protection. This definition of conservation recognizes that some of the areas of conservation importance identified in our analyses will require specific national or international designation (e.g. protected area) or governance.

Leveraging best available data on the current and potential distributions of EU species, carbon sequestration, and land cover, we identify priority areas for how an optimal allocation of conservation, restoration, and production across land cover types could be met in the

EU. Through a series of different planning scenarios, we explore three main questions: (1) What are the implications of simultaneously optimizing the allocation of conservation and restoration measures across a landscape rather than considering the allocation of these measures separately? (2) What are the trade-offs and synergies between optimizing the allocation of conservation and restoration measures for biodiversity conservation and carbon sequestration objectives? Finally, (3) What are the implications of multilateral burden-sharing on the coherence, efficiency, and resilience of possible conservation and restoration allocation solutions?

2 Methods

2.1 Problem formulation

We formulate the spatial optimization of land-use decisions as a mixed linear programming problem, where the objective is to maximize the extent to which targets for species conservation and carbon sequestration are met through optimal zoning of conservation and restoration measures across land-use types in the EU (eq. 1), subject to spatial constraints in the applications of these measures.

$$min[\sum_{p}^{P}\sum_{z}^{Z}\sum_{s}^{S} w_s((t_s - \sum_{p}^{P}\sum_{z}^{Z} r_{s,p}k_{z,s}x_{p,z})/t_s)] \tag{1}$$

In the minimal shortfall objective (eq. 1), $p \in P$ indicates a given 100 km^2 planning unit, $z \in Z$ indicates a management zone (figure 1A), and $a_{p,z}$ is the area of the planning unit p allocated to zone z in the solution. $r_{s,p}$ is the amount of feature $s \in S$ in planning unit p and has the same unit as the respective target, t_s. Features include species of concern and carbon (see below for additional details). $k_{z,s}$ is a zone-specific parameter that defines the proportional contribution of zone z to achieving the target for feature s and is determined by the habitat preferences and land use specific threats for the species (see SI and tables **??**, and 3 for additional details), or the estimated carbon sequestration of that zone. w_s is a feature-specific weighting that gives higher or lower priority to achieving a given target when not all targets can be met (see 2.8 for additional details).

2.2 Management zones

We consider 25 management zones (decision possibilities) that can be allocated proportionally within each of the 43,498 planning units across the EU (figure 1A). Restoration zones consist of both natural land cover restored from production landscapes and low-intensity production landscapes restored from high-intensity production landscapes (table 1). Production zones consist of varying intensity crop, pasture, and forestry areas that can be maintained or created from other land-cover types. Finally, conservation zones capture the maintenance of natural land-cover types (i.e., land not allocated to production or restoration). We bound the allocation of each of these zones in a given planning unit based on the current and potential land cover, improving the realism of our proposed solution space and minimizing unnecessary and unrealistic ”reshuffling” (e.g. forests that should be restored to grasslands).

For the set of conservation zones in each planning unit, we consider the lower bound to be the current area of a given land-cover type within the planning unit that is presently within protected areas (Natura 2000 site [15]) or classified as primary forest [16]. The upper bound of conservation zones is the total extent of a given natural land-cover type in the planning unit (see below for additional information on land cover data). Note again that conservation here is distinct from legal protection. Instead, ”conservation” is natural land not allocated to production, urban area, or restoration. Importantly, this conceptualization of conservation allows us to capture the habitat contributions of all natural land in a given solution to optimize the biodiversity benefits of restoration allocation.

For the set of restoration zones, the lower bound is 0 (no restoration is required in any given planning unit). The upper bound is constrained by the potential distribution of the restored land-cover type and the current land cover available to be restored to that future

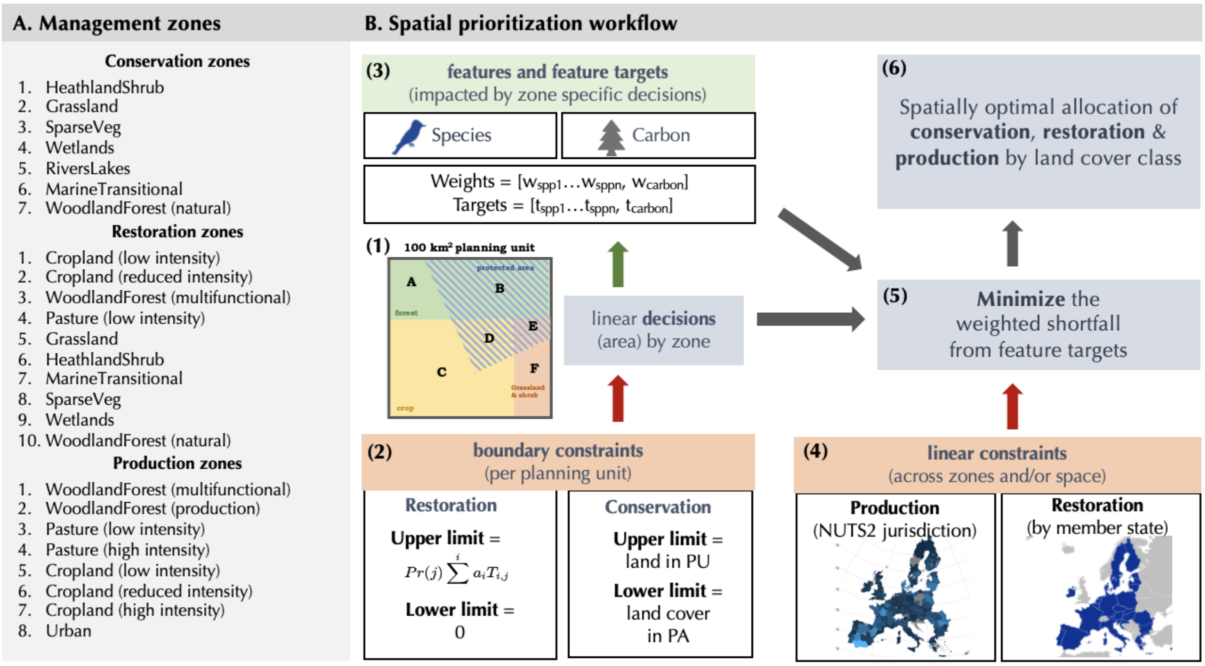

Figure 1: Schematic diagram of the prioritization analyses. (A) List of management zones, whose spatial allocation is optimized under a suite of different scenarios 2.8 in the analyses. (B) For each priority species, we set a target to meet by conserving or restoring habitat types for the species within their present range. For carbon, the aim was to maximize the amount of carbon stored in conserved or restored areas. Depending on the scenario variants, species were weighted in importance differently relative to carbon to explore the implications of putting more emphasis on different objectives. The area allocation of a planning unit to a given management zone was bounded with the bounds depending on the planning unit and zone (B2). In addition, the optimization included constraints on the area under restoration and how much area needed to be under production (grazing, farming, timber harvesting, B4) in 2030. The result of this constrained optimization is a series of maps identifying priorities for conservation, restoration, and production of food and timber products (B6)

land cover (according to logical transitions; see SI 6.1, eq. 4, and table 1 for additional details). For the set of production zones, the lower bound is 0 (no production is required in any given planning unit), and the upper bound is the land area of the planning unit minus the distribution of the natural land cover in currently protected areas.

2.3 Meeting 2030 production demands

To avoid setting more land aside for conservation and restoration than what would be needed given projected demands for timber and agricultural products, we included production area requirements for each NUTS2 jurisdiction for each of the seven production zones (high-intensity cropland, reduced intensity cropland, low-intensity cropland, high-intensity pasture, low-intensity pasture, production forestry, and multi-functional forestry) (Figure 1B(4)). For example, to ensure the allocation of sufficient high-intensity cropland to meet projected 2030 production needs, we require that the sum of the area allocated to the relevant cropland zones across planning units within each NUTS2 jurisdiction is greater than or equal to the specified 2030 target of that jurisdiction. These targets are derived from GLOBIOM [17], a global partial equilibrium model that explicitly integrates different land-use sectors (IIASA GLOBIOM, 2022), and which has been adapted specifically for EU contexts and policy problems. In this work we made use of the "Fit for 55" scenarios constructed as part of the EUCLIMIT project and that following a "Middle-of-road" design for the European bio-economy in 2030 and is fully consistent with the EU Corine land cover product (see 6.4).

We formally implement these constraints using eq. 2:

$$\sum_{p}^{P}\sum_{z}^{Z}(D_{i,k,p,z} \times X_{p,z}) \quad \theta \quad t_{i,k} \qquad \forall i \in I \quad \text{and} \quad \forall k \in K \tag{2}$$

where $i \in I$ denotes a set of planning units p (e.g., all planning units within a NUTS2 jurisdiction), and $k \in K$ a set of zones relevant to a given target $t_{i,k}$. $D_{i,k,p,z}$ denotes the constraint data associated with planning units $i \in I$ for zones $k \in K$.

2.4 Biodiversity data

We estimated the distribution of 1277 species included the EU habitats directive using an integrated species distribution modeling (iSDM) approach where best-available data sources (occurrence, preference, expert information - see SI section 6.5 for a full description of included data) are integrated into one joint prediction (see SI section 6.6 for prediction variables) using different types of linear and non-linear modeling approaches (full methodology for SDM in SI section 6.6.1). We also leverage species distribution models to depict the potential distribution of the species (*sensu* [18]), understanding potential in this context as the contemporary climatic, soil and natural vegetation conditions that would allow a species to persist in an area (full methodology for potential SDMs available in SI 6.6.2). Critically, and opposed to mapping current suitable habitat, the potential modelling approach considers only contemporary differences in climate and soil, and not any land-cover or land-use, aligning with the concept of the potential natural vegetation of Europe [19]. The predictions from the species distribution models used here thus aim to depict where a species might exist under current conditions while also allowing modest inter- and extrapolation from its current distribution.

2.5 Land cover data

We used data on the distribution of land-cover types as distinguished by the Corine 2018 dataset. The thematic legend of the Corine land-cover data (level 2) was re-categorized into different MAES habitat categories to allow for a crosswalk with species habitat preferences. Our input land cover differed from the MAES categories slightly for production lands, which we first split from natural lands using more disaggregate layers of the Corine dataset where necessary (e.g., grasslands where split into natural grasslands and pasture lands). We distinguished the current area of different management intensity classes of cropland, forests, and pasture using the current distribution of those production lands multiplied by the estimated percentage of intensity for a given production intensity at the 10 km^2 scale using outputs from G4M and EPIC models, constructed as part of the GLOBIOM Fit for 55 scenarios (see above).

To calculate the protected and unprotected land cover distributions within each PU, we rasterized protected areas in the Natura 2000 network to align with reclassified 100-meter

Corine land cover data. Then, we masked the 100-meter land cover data by the protected areas and aggregated this from the 100m resolution to the 10km grain size by calculating the proportion of 100m grid cells in each coarser grain for each land cover inside and outside of current protected areas. This was done to separate areas inside and outside protected areas, as in particular restoration inside such sites is specific objective in the EU Restoration law.

To identify potentially restorable land, we followed the concept of potential natural vegetation (*sensu* [18]). We define potential in this context as the contemporary climatic, topography, soil, and natural vegetation conditions that allow for a specific type of natural habitat (e.g., Forest, Wetland) to occur in an area. The full methodology and input data used for estimating potential land cover are available in the SI section 6.4.

2.6 Carbon data

For current carbon stocks, we used data on above-ground, below-ground, and soil organic carbon at risk from land-use change from [12]. These data were created by selecting and integrating the best available carbon maps for different vegetation types. All data are in units of tC/ha. For the analysis, we combined the current carbon layers by calculating the combined sum of above- & below-ground and soil organic carbon for Europe. We included the total amount as an additional feature in the prioritization (see SI section 6.7.1 for additional information). Similarly, for allocating restoration priorities in regard to carbon contributions, we needed to identify areas with high carbon sequestration potential. Here followed an approach that combined the different techniques from [14] and [20] for potential carbon estimation, allowing for a estimation of carbon potential per planning unit and natural land-cover types (see SI section 6.7.2 for full details on methodology).

2.7 Feature targets

Species targets are defined relative to the minimum habitat area (km^2) necessary to qualify the species for the Least Concern conservation status following IUCN criteria (see [12]). We equate this to the current area of habitat (AOH) within the range of the species, or 2200 km^2 in the case that the AOH is less than 2200 km^2 (eq. 3). To calculate the current AOH, we assess the proportion of each species range that is within corresponding MAES land cover preferences (”preferred” or ”suitable” habitat) for the given species. To differentiate the habitat provided by low and high-intensity production landscapes, we leverage frequently reported pressures on habitats and species associated with agriculture (EEA State of Nature Dataset; see SI section table 3, 4, and 2 for more information on how we classified threats by zone). Moreover, we set a maximum area target for any given species to $10\hat{6}$ km^2 to avoid infeasibly large AOH targets [12].

$$t_s = \min(\max(2200, AOH_s), 10^6) \tag{3}$$

As mentioned above (eq. 1), we define a zone-specific parameter $k_{z,s}$ that identifies the proportional contribution of zone z to achieving the target for feature s and is determined by the habitat preferences for the species. $k_{z,s}$ is 1 if zone z (e.g. mountain coniferous forest) fully contribute to achieving the conservation target for species s (e.g. the three-toed woodpecker). $k_{z,s}$ is 0 if the zone z does not contribute to the species' habitat. $k_{z,s}$ can be between 0 and 1 for species that persist in agricultural habitat but are sensitive to threats associated with the intensity and type of agriculture in zone z (see SI table 3, 4, and 2 for more details).

We set the carbon sequestration target to the maximum potential value of carbon in each planning unit and similar defined zone-specific parameters to understand the different contributions of different land-cover types. We jointly optimize for both biodiversity and carbon sequestration, and since we include only one target for carbon compared to multiple targets for biodiversity outcomes, the weighting of target shortfalls is set accordingly (see information on feature weighting scenarios below 2.8).

2.8 Scenarios

We explore 45 different scenarios of optimal land allocation. First, we vary the weighting of carbon shortfalls relative to biodiversity shortfalls (15 different weighting scenarios) to understand the synergies and trade-offs between these two objectives. Second, we compare solutions that only optimally allocate restoration to solutions that prioritize restoration allocation in the context of conservation and broader landscape constraints. Third, we set constraints (eq. 2) for meeting 20% restoration targets at (1) the EU scale and (2) the member state scale, allowing us to assess the impact of burden sharing between countries across different carbon/biodiversity objective weighting scenarios.

We use the *prioritzr* package [10] and the Gurobi solver (v9.5) [21] to build and solve each of our spatial optimization problem scenarios.

3 Results

The resulting conservation plans provide insight into the optimal allocation of restoration, conservation, and production measures under a variety of scenarios across the EU (a single solution scenario is shown in figure 2). We first compare plans for meeting 20% restoration commitments between scenarios that jointly optimize restoration and conservation (figure 3B) and scenarios that only consider restoration (figure 3A). We show that the two scenarios result in significant spatial differences (figure 3C) and land cover differences (figure 3D) in restoration priority. Because the coordinated prioritization framework integrates information about the broader landscape within which restoration decisions are made, it presents a more holistic perspective for both restoration and conservation, especially when compared to the less informed optimization scheme (figure 3B).

Using the coordinated prioritization framework (jointly optimizing the allocation of restoration, conservation, and production), we explore a suite of optimal land management allocation scenarios by varying the weight of biodiversity and carbon objectives (figure 4A).

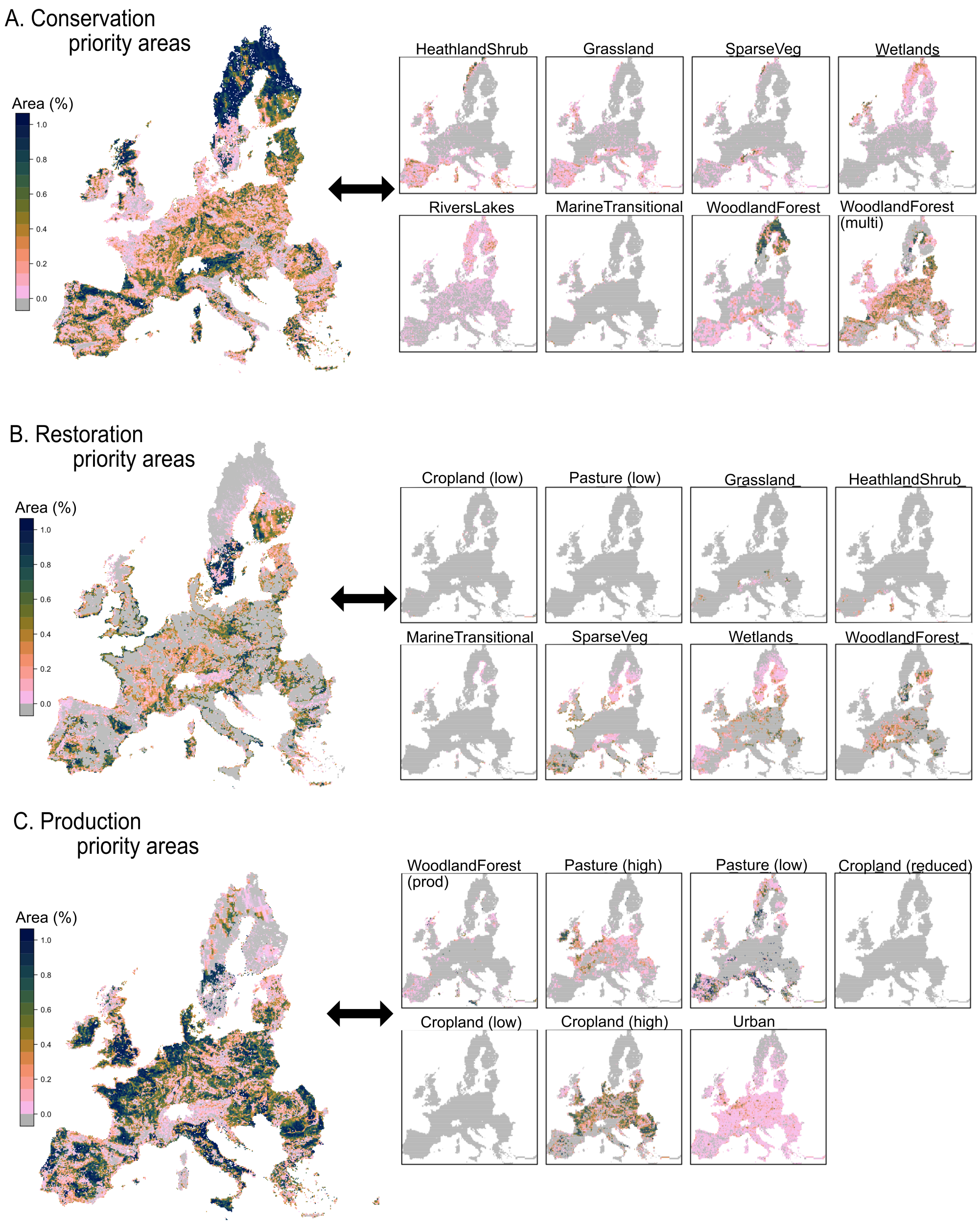

Figure 2: Optimal allocation of conservation, restoration, and production across land-cover types for one solution scenario (carbon target weight set to 20% of the sum of all species targets and restoration budgets defined at the member state scale). (A) The optimal allocation of conservation (maintained natural land) across the EU and the breakdown of conservation by land cover type. Notably, natural and multi-functional woodlands and forests make up over 50% of the conserved land priorities (B) The optimal allocation of restoration across the EU and the breakdown of restoration by land cover type. Notably, natural and multi-functional woodlands and forests make up over 50% of the conserved land priorities (C) Production area targets, while constant throughout scenarios at the sub-national (NUTS2) level, do vary in their spatial distribution as the result of the conservation and restoration priorities of that given scenario. Urban areas remain constant throughout all scenario solutions and are set to match 2018 urban area distributions in each planning unit.

Additionally, we explore how these trade-off curves vary depending on policy implementation (namely, burden sharing between countries) (figure 4A). On the efficiency frontier with higher shortfalls, we require that 20% restoration targets are not exceeded in any given member state (figure 4A), meaning the burden to meet the EU target is evenly distributed across countries. Notably, solutions are still optimized in the context of continental scale portfolios of species and carbon features, just with additional constraints. Because optimization of restoration targets at the EU scale is less constrained, we by definition see lower biodiversity and carbon shortfalls overall (figure 4A).

We show how shifting the distribution of responsibility to reach area-based targets changes the spatial priorities of restoration (figure 5A) and the distribution of natural land cover prioritized for restoration (figure 4B). Notably, burden sharing has a less significant impact on the distribution of restoration land cover types (figure 4B; figure 5D) and overall target shortfalls than the relative weighting of biodiversity and carbon sequestration objectives (figure 4A).

While the aggregate shortfall from biodiversity targets provides some insight into the overall efficacy of different scenarios, we also aimed at understanding the "winners" (species which benefit most) and "losers" (species which benefit least) of these different planning scenarios. We show that some species groups (e.g., birds) have significantly shifted their shortfall distributions due to objective weighting, while others (e.g., mammals) are less influenced by exact objective weightings (figure 4). This is unsurprising given that birds benefit disproportionately from the increased wetland restoration priorities that emerge in scenarios that more heavily weight biodiversity. Changing the burden distribution between countries does not appear to significantly impact any given species group more than others (figure 4E), but may be impacting individual species within those groups differently.

Despite the benefits, EU scale restoration targets result in some countries bearing a more substantial burden for meeting EU restoration targets (up to 49% of land prioritized for restoration in a single country, as opposed to a maximum of 20% area per country in the

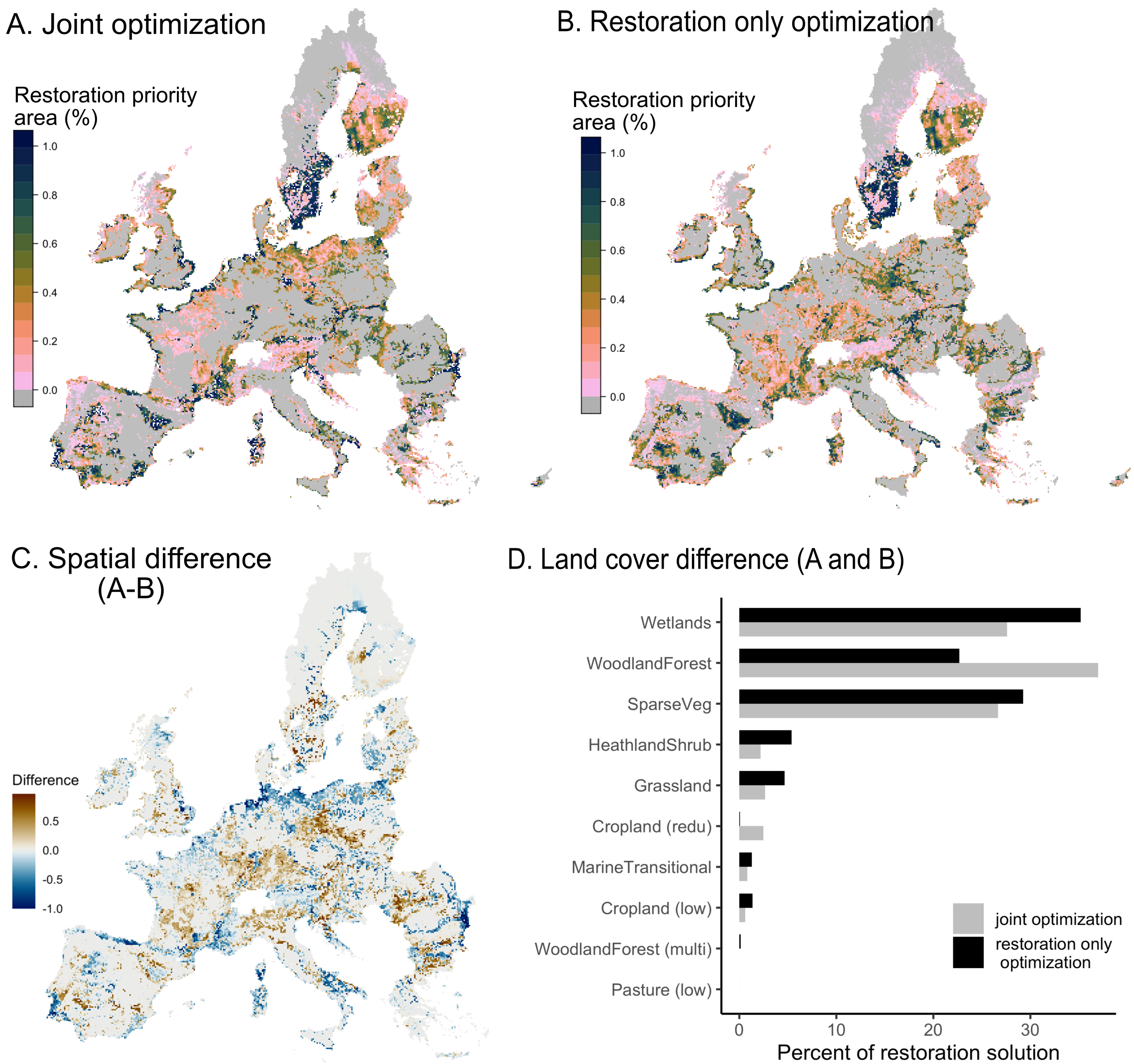

Figure 3: Using the same planning unit constraints and restoration transition matrix (fig 1) from figure 2, we compare solutions of restoration prioritization when (A) jointly optimized to meet restoration targets in the context of the optimal allocation of conservation and production lands (same as figure 2B) and (B) only considering restoration priorities. (C) Significant spatial differences and (D) land type differences in restoration priority emerge between the two solutions.

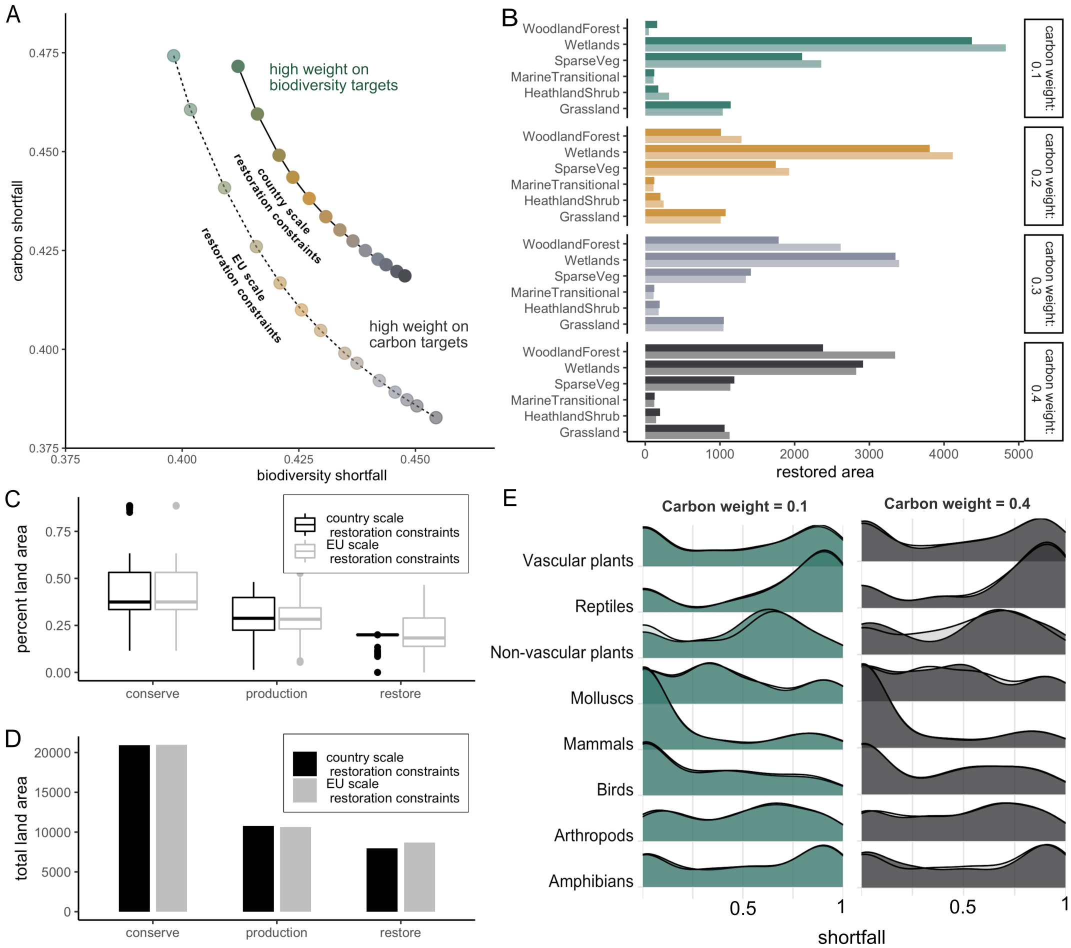

Figure 4: (A) We explore a suite of scenarios varying the weight of biodiversity and carbon objectives. Additionally, we explore how these trade-off curves vary depending on policy implementation. On the efficiency frontier with higher shortfalls, we require that 20% restoration targets are not exceeded in any given member state. In the lower frontier, we allow restoration to occur anywhere in the EU to contribute to an overall 20% area target. (B) Restoration per natural land cover also varies depending on the scenario. Here the bar colors align with the solutions from (A) and the shade with the policy implementation scenario (lighter colors align with EU scale priorities), showing that value weighting has a more significant impact on priorities than formalizing equity of burden between countries in this context. (C) While less constrained optimization allows for the allocation of conservation and restoration measures that provide larger biodiversity benefits (C) these solutions can result in up to 50% of individual member states being allocated to restoration as opposed to that limit being set at 20% in the constrained optimization. (D) Despite this difference in distribution, the total area restored in each scenario is only marginally different, driven by countries where reaching 20% restoration is not feasible in the solution, given production and restoration transition constraints. (E) We can see that some species groups (e.g., birds) have significantly shifted shortfall distributions while other species groups (e.g., mammals) are less influenced by exact objective weightings.

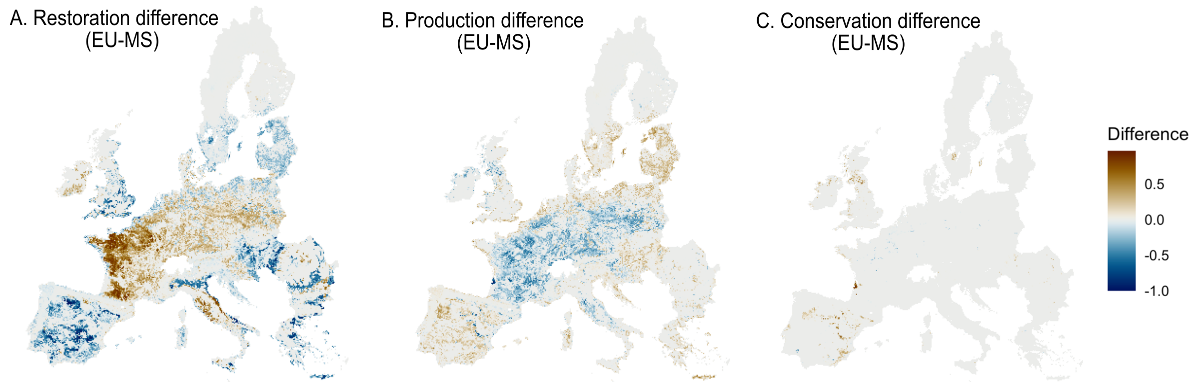

Figure 5: distribution of the responsibility to reach area-based targets significantly changes the spatial distribution of (A) restoration and (B) production priorities across the EU. Negative values (blue) indicate that planning units with more area are prioritized for (A) restoration, (B) production, and (C) conservation in the member-state target scenario. Positive values (brown) indicate planning units with more area prioritized in EU scale target scenarios.

constrained scenarios) (figure 4D). We sought to understand if concentrating solutions in some countries increased sociopolitical risks associated with the prioritized conservation and restoration allocation. Our results suggest the contrary (figure 6A-C). Optimization scenarios that allow for uneven distribution of restoration burden between countries generally resulted in a more significant portion of the restoration solution space allocated in countries with lower corruption indices, higher GDP, and higher political engagement (figure 6A-C) [22]. Our results show less significant difference in the distribution of conservation or production across these sociopolitical metrics (figure 6A-C).

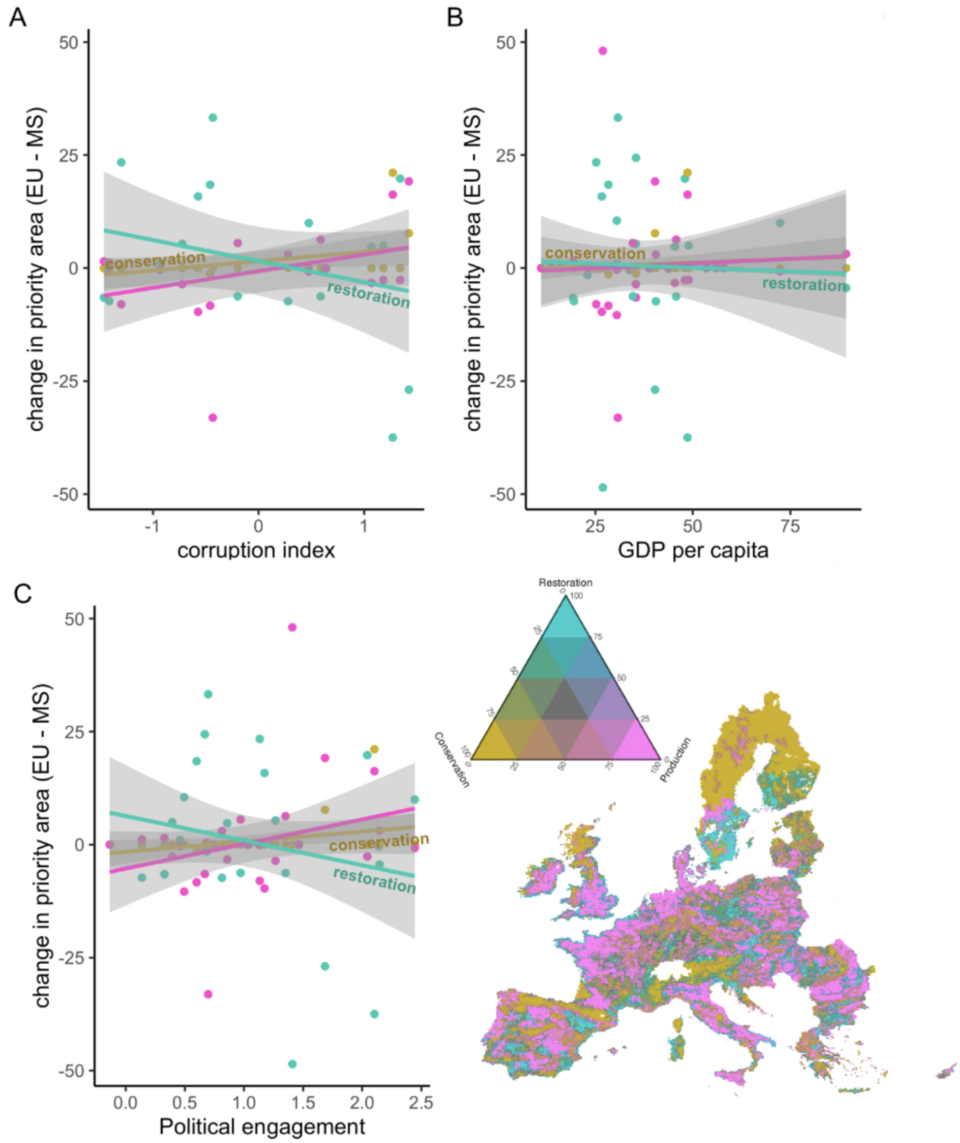

Figure 6: While less constrained optimization results in lower biodiversity and carbon shortfalls (figure 4A) understanding the risks on the resilience of conservation and restoration that the resultant solutions pose as the result of uneven burden might place is critical, particularly given the long time horizons of benefit accrual. Optimizations run with EU scale targets have a larger portion of the restoration solution space in countries with lower corruption indices (A), higher GDP (B), and high political engagement (C) in the EU context. The map shows the distribution of these three categories given member state restoration targets. Colors in the map align with land covers included in regression lines in A-C.

4 Discussion

Leveraging a suite of scenarios (figure 4A), our analysis provides insight into the optimal allocation of conservation, restoration, and production across land-cover types in the context of current EU policies (figure 2). We show that joint prioritization of multiple area-based conservation measures (i.e., conservation and restoration) produces significantly improved ecological and NCP outcomes from these measures compared to uncoupled planning (figure 3). Additionally, we explore how burden-sharing and cooperative planning between countries in multilateral agreements can improve the efficacy of area-based conservation and restoration measures (figure 4A), while not increasing the sociopolitical dimensions of risk associated with priorities (figure 6) [23].

Unlike existing approaches to spatial planning of conservation and restoration measures (which have to-date have been implemented separately [12, 14]), our analysis concurrently optimizes the allocation of both measures. This advance in coordinated spatial planning provides important insight for the EU and individual Member States on devising strategies to meet commitments under both the Nature Restoration Law and the EU 2030 biodiversity strategy. Moreover, because our analysis optimizes biodiversity and NCP portfolios across the entire EU, even in scenarios where restoration targets are set at the Member State scales (figure 4A; top line), our resulting plans provide information on how Member State implementation can be strategized to maximize benefits in the context of continental biodiversity portfolios and climate mitigation opportunities. Additional constraints at the Member State (or subnational; e.g., NUTS2) scale could be integrated into this framework (e.g., to reflect different values or priorities) and resulting plans would still provide more optimal EU-wide benefit to biodiversity and carbon sequestration than if Member States planned their strategies entirely independent of one another.

Recognizing that conservation priorities are normative at their core, we do not suggest that there is a single optimal allocation of conservation and restoration measures across the EU. Rather we showcase how the weightings of different objectives and policy targets might shift biodiversity and carbon sequestration outcomes (figure 4A). Providing a suite of spatial plans and their synergies can help decision-makers understand the feasibility different strategies, as well as identify places and measures that will likely be prioritized regardless of precise problem formulation or objective weightings.

We stress that while our analysis provides restoration priorities that align with area-based policy targets (figure 3) outlined by the EU restoration law, we do not claim or propose any areas for legal or strict protection as outlined by the EU 2030 Biodiversity Strategy (30% of land area in protected areas and 10% under strict protection). Such decisions need to be taken following in-depth and bottom-up investigations, also considering factors that were not available for inclusion in this work and at this scale (for example land tenure). The definition of conservation used in this study more closely aligns with the contemporary emphasis of biodiversity conservation on the importance of areas outside of traditional protected areas (”Other Effective Conservation Measures” (OECMs) [6, 3]). Our solutions reflect that legal designations will only be a subset of the areas identified as meaningfully contributing to biodiversity and carbon sequestration targets. Regardless, some portion of the conservation and restoration priority areas identified in our solutions would likely benefit from national or international protected area designation. Legal protection can be valuable where narrow-ranging and threatened species occur or where there is a need for active site management. Post-hoc analysis of solutions (e.g., identifying areas with high synergies across scenarios, high risk for development, or high importance for species of particular concern) might help provide priorities for expanding legally protected area networks across the EU.

While conservation planning at multilateral scales (e.g., global, continental) is more efficient along the dimensions specified in an optimization (as shown in figure 4A), and allows for consideration of large scale ecological processes otherwise overlooked, there are several potential shortcomings. For example, large-scale planning can be detrimental across important problem dimensions not explicitly included in the optimization. These excluded dimensions often reflect areas or values that do not have consistent data at aggregate scales. This can lead to disparities in the distributional equity of plans due to the lack of quantification of local natural resource use, land tenure, or the inability to consider cultural components of conservation benefits in standardized data formats [24]. Conservation planning that integrates knowledge and values across multiple scales may provide the most promising path forward [25]. Our methods provide insight into how we can, at once, set conservation targets acknowledging jurisdictional boundaries and differences, while concurrently considering the broader species distributions and habitat needs when prioritizing areas to meet those targets.

4.1 Next steps

Improving the realism of 2030 biodiversity policy implementation in our optimization problem will be a critical next step in this study. For example, our current solutions (across all scenarios) allocate only a small amount of restoration budgets to reducing the intensity of production lands (figure 3D). This solution pattern emerges because natural landscapes benefit species and carbon sequestration more than production landscapes in most contexts, even when production is done at low intensities. Furthermore, because we do not consider the opportunity costs of restoration, if plans can meet 2030 production requirements in a NUTS2 jurisdiction without low-intensity restoration allocation, the plans will favor natural land restoration. Therefore, exploring scenarios that require a specified distribution of restoration between natural land cover restoration and reduced intensity production restoration might provide more realistic insight into how incentive allocation would change biodiversity and carbon sequestration outcomes of the EU restoration law.

There are several other limitations to our current analysis, which will motivate subsequent work on this project. First, we do not consider the sensitivity of our results to uncertainties in underlying spatial data. Comparing solutions across different probability thresholds for species distribution models and land cover potential maps will be critical for understanding the sensitivity of our results to these underlying data. Second, in these analyses, we set species habitat targets at the EU scale when, in reality, setting habitat targets at the biogeographic or country scale will better align with policy goals set out by the EU commission. Moreover, disaggregating habitat targets might help improve the climate resilience of scenario solutions. Notably, we do not consider the implications of climate change risk and uncertainty on ecosystem threats or expected species range shifts [26]. Including projected species ranges under climate change scenarios (similar to [27]) as a component of biodiversity targets might help propose conservation and restoration measures that provide resilient biodiversity benefits beyond 2030. Finally, we do not yet consider connectivity of proposed conservation and restoration measures [28].

5 Conclusions

In this study, we explore the impact of multilateral burden sharing and multi-objective weighting on optimal land conservation and restoration allocation. We highlight the implications of these land management measures on expected biodiversity and climate mitigation outcomes across the EU. Our methodology allows for simultaneous planning of conservation and restoration measures under land use constraints, showing that strategic planning across land management measures can significantly improve the efficacy and coherence of area-based conservation implementation when compared to disparate planning. Our results provide insight for strategic implementation of the EU restoration law and 2030 biodiversity policy commitments, while creating a baseline for EU Member States to evaluate the potential of their conservation pledges. Our methods are broadly relevant to contexts around the world and showcase how spatial planning methods can meaningfully optimize land allocation in a way that benefits both nature and people.

6 Supporting Information

6.1 Restoration potential

Area of restoration potential within each 10 km^2 planning unit was further refined from the 10km grid cell potential using the existing land cover distributions within each PU. Restoring current land cover to a potential land cover is defined as possible only up to the probability potential of a given planning unit (from potential distribution model) that is currently within managed land cover types defined as feasible for restoration to a given potential land cover (table 1). Meaning the upper bound of potential area of restoration for a given land cover j in a given planning unit is given by eq 4

$$Pr(j) \sum_{i}^{I} a_i T_{i,j} \tag{4}$$

Where $Pr(j)$ is the probability of a given potential land cover in the PU, a_i is the percent of the PU area that is a given current land cover i and $T_{i,j}$is the logical transition between the current land cover and potential land cover (see table 1)

6.2 Management intensity threats

We map EEA land use threats onto our seven production zones. Species are then mapped to threats that they are sensitive to. If a species is sensitive to 3 or more threats but is considered suitable to the land use, $k_{s,z}$ is set to 0.1. If a species is sensitive to 1-3 threats but considered suitable to the land use, $k_{s,z}$ is set to 0.5. If a species is sensitive to 1-3 threats but considered suitable to the land use, $k_{s,z}$ is set to 1.

6.3 Management zones and constraints

As described in the main text, we set planning unit constraints on each zone as well as linear constraints that consider targets or budgets across zones and multiple planning units. These

	WoodlandForest_natural	WoodlandForest_multi	WoodlandForest_prod	HeathlandShrub_natural	Grassland_natural	SparseVeg_natural	Cropland_low	Cropland_med	Cropland_high	Urban_urban	Wetlands_natural	RiversLakes_natural	MarineTransitional_natural	Pasture_low	Pasture_high
WoodlandForest_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WoodlandForest_multi	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WoodlandForest_prod	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
HeathlandShrub_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Grassland_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SparseVeg_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cropland_low	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0
Cropland_med	0	0	0	1	1	1	1	0	0	0	1	0	1	0	0
Cropland_high	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0
Urban_urban	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wetlands_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RiversLakes_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MarineTransitional_natural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pasture_low	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0
Pasture_high	0	0	0	1	1	1	1	1	0	0	1	0	1	1	0

Table 1: Restoration transition matrix: we restrict restoration of a current land cover type (row) to a potential land cover type (column) based on logical and desirable transitions. For example, native vegetation is not proposed for restoration to another land cover type even if that vegetation is within the potential zone of another land cover type. Human modified land cover (cropland, pastureland) can be restored to a set of "native" land cover types only within the potential range of those land covers.

constraints are outline in table 5.

6.4 Land cover data

6.4.1 Potential distribution of natural land cover

Similar as for the distribution of species, for the identification of potentially restorable land we followed the concept of potentially natural vegetation (sensu [29]). We understand potential in this context as the contemporary climatic, topography, soil and natural vegetation

Code	Pressure/threat	Cropland (high)	Cropland (reduced)	Cropland (low)
A01	Conversion into agricultural land (excluding drainage and burning)	3	2	1
A25	Agricultural activities generating point source pollution to surface or ground waters	3	2	1
A26	Agricultural activities generating diffuse pollution to surface or ground waters	3	2	1
A29	Agricultural activities generating soil pollution	3	2	1
A30	Active abstractions from groundwater, surface water or mixed water for agriculture	3	2	1
A31	Drainage for use as agricultural land	3	2	1
A35	Agricultural crops for renewable energy production	3	2	1
A02	Conversion from one type of agricultural land use to another (excluding drainage and burning)	2	1	0
A03	Conversion from mixed farming and agroforestry systems to specialized (e.g., single crop) production	2	1	0
A04	Changes in terrain and surface of agricultural areas	2	1	0
A15	Tillage practices (e.g., ploughing) in agriculture	2	1	0
A16	Other soil management practices in agriculture	2	1	0
A17	Harvesting of crops and cutting of croplands	2	1	0
A18	Irrigation of agricultural land	2	1	1
A19	Application of natural fertilizers on agricultural land	2	1	1
A20	Application of synthetic (mineral) fertilizers on agricultural land	2	1	1
A21	Use of plant protection chemicals in agriculture	2	1	1
A22	Use of physical plant protection in agriculture	2	1	1
A34	Introduction and spread of new crops (including GMOs)	2	1	0
A05	Removal of small landscape features for agricultural land parcel consolidation (hedges, stone walls, rushes, open ditches, springs, solitary trees, etc.)	1	1	0
A07	Abandonment of management/use of other agricultural and agroforestry systems (all except grassland)	1	1	0
A08	Mowing or cutting of grasslands	1	1	0
A23	Use of other pest control methods in agriculture (excluding tillage)	1	1	2
A24	Waste management practices in agriculture	1	1	1

Table 2: Mapping EEA threats onto cropland intensity zones. Threat codes align with species, allowing for the differentiation of biodiversity contributions of different intensities of production

conditions that allow for a specific type of natural habitat (e.g. Forest, Wetland) to occur in an area. Critically this approach considers only contemporary and not future conditions and aligns with the concept of potential natural vegetation of Europe [19]. We first assembled a Europe-wide database on the distribution of habitats in Europe where we followed the thematic legend of the MAES habitat classification system at level 2, while ignoring any

Code	Pressure/threat	Pasture (Low)	Pasture (High)
A01	Conversion into agricultural land (excluding drainage and burning)	1	2
A04	Changes in terrain and surface of agricultural areas	1	2
A08	Mowing or cutting of grasslands	1	2
A09	Intensive grazing or overgrazing by livestock	1	2
A13	Reseeding of grasslands and other semi-natural habitats	1	1
A14	Livestock farming (without grazing)	1	2
A02	Conversion from one type of agricultural land use to another (excluding drainage and burning)	0	1
A03	Conversion from mixed farming and agroforestry systems to specialized (e.g., single crop) production	0	1
A05	Removal of small landscape features for agricultural land parcel consolidation (hedges, stone walls, rushes, open ditches, springs, solitary trees, etc.)	0	1
A06	Abandonment of grassland management (e.g., cessation of grazing or mowing)	0	1
A07	Abandonment of management/use of other agricultural and agroforestry systems (all except grassland)	0	1
A10	Extensive grazing or under grazing by livestock	0	0

Table 3: Mapping EEA threats onto pasture intensity zones. Threat codes align with species, allowing for the differentiation of biodiversity contributions of different intensities of production

strictly anthropogenic habitats (e.g. Urban, Cropland, Pasture) as well as Rivers and Lakes. Different data sources on the distribution of habitats differ in terms of their geographic spread and biases, thus in order to not rely on any single data source of European ecosystems we integrated habitat data from three different sources collated for Europe:

We took habitat information from the European Habitat Directive which gives the occurrence of all EUNIS habitats listed in the Article 17 of the habitats directive at a 10km resolution. We used a crosswalk developed by the European Environment Agency and Biodiversity Topic Centre to translate the EUNIS types to Corine CLC and subsequently to

Code	Pressure/threat	Forest (natural)	Forest (multi)	Forestry (prod)
B01	Conversion to forest from other land uses, or afforestation (excluding drainage)	0	0	0
B02	Conversion to other types of forests including monocultures	0	1	1
B03	Replanting with or introducing non-native or non-typical species (including new species and GMOs)	0	1	1
B04	Abandonment of traditional forest management	0	1	2
B05	Logging without replanting or natural regrowth	0	1	1
B06	Logging (excluding clear cutting) of individual trees	0	1	0
B07	Removal of dead and dying trees, including debris	0	1	2
B08	Removal of old trees (excluding dead or dying trees)	0	1	1
B10	Illegal logging	0	1	1
B11	Cork extraction and forest exploitation excluding logging	0	0	0
B12	Thinning of tree layer	0	1	1
B13	Burning for forestry	0	0	0
B14	Suppression of fire for forestry	0	0	0
B15	Forest management reducing old growth forests	0	0	1
B16	Wood transport	0	1	1
B17	Tillage practices in forestry and other soil management practices in forestry	0	0	1
B18	Application of natural fertilizers	0	1	1
B19	Application of synthetic fertilizers in forestry, including liming of forest soils	0	1	2
B20	Use of plant protection chemicals in forestry	0	1	2
B21	Use of physical plant protection in forestry, excluding tree layer thinning	0	1	1
B22	Use of other pest control methods in forestry	0	1	1
B23	Forestry activities generating pollution to surface or ground waters	0	1	2
B24	Forestry activities generating air pollution	0	1	1
B25	Forestry activities generating marine pollution	0	1	2
B26	Forestry activities generating soil pollution	0	1	2
B27	Modification of hydrological conditions, or physical alteration of water bodies and drainage for forestry (including dams)	0	1	1
B28	Forests for renewable energy production	0	1	1
B29	Other forestry activities, excluding those relating to agroforestry	0	1	2
B09	Clear-cutting, removal of all trees	0	0	1

Table 4: Mapping EEA threats onto forestry intensity zones. Threat codes align with species, allowing for the differentiation of biodiversity contributions of different intensities of production

MAES level 2 categories. We further made use of point occurrence datasets for key habitats in the new EEA suitability predictions. Here the habitat categories were reclassified into the respective MAES types, following the CLC to MAES crosswalk. Finally, we prepared

					Linear constraints							
			Planning unit constraint		Production				Restoration			
Action	Zone	Land cover	Lower bound	Upper bound	*k*	Logic	Amount	Scale *(i)*	*k*	Logic	Amount	Scale *(i)*
Restoration	1	Cropland (low intensity)	0	*see restoration potential methods*	1	>=	2030 GLOBIOM	NUTS2	7	>=	20% area	(EU or country depending on scenario)
	2	Cropland (reduced intensity)	0		2	>=	2030 GLOBIOM	NUTS2	7			
	3	WoodlandForest (multifunctional)	0		3	>=	2030 GLOBIOM	NUTS2	7			
	4	Pasture (low intensity)	0		4	>=	2030 GLOBIOM	NUTS2	7			
	5	Grassland	0						7			
	6	HeathlandShrub	0						7			
	7	MarineTransitional	0						7			
	8	SparseVeg	0						7			
	9	Wetlands	0						7			
	10	WoodlandForest	0						7			
Conservation	11	HeathlandShrub	area in PA	area in PU								
	12	Grassland	area in PA	area in PU								
	13	SparseVeg	area in PA	area in PU								
	14	Wetlands	area in PA	area in PU								
	15	RiversLakes	area in PA	area in PU								
	16	MarineTransitional	area in PA	area in PU								
	17	WoodlandForest	area in PA	area in PU								
Production	18	WoodlandForest (multifunctional)	90% current area	1	3	>=	2030 GLOBIOM	NUTS2				
	19	WoodlandForest (production)	90% current area	1	5	>=	2030 GLOBIOM	NUTS2				
	20	Pasture (low intensity)	90% current area	1	4	>=	2030 GLOBIOM	NUTS2				
	21	Pasture (high intensity)	90% current area	1	5	>=	2030 GLOBIOM	NUTS2				
	22	Cropland (low intensity)	90% current area	1	1	>=	2030 GLOBIOM	NUTS2				
	23	Cropland (reduced intensity)	90% current area	1	2	>=	2030 GLOBIOM	NUTS2				
	24	Cropland (high intensity)	90% current area	1	6	>=	2030 GLOBIOM	NUTS2				
Urban	25	Urban	100% current area	100% current area								

Table 5: Summary of management zones and constraints

point occurrence data from the openly available land-cover and land-use database LUCAS [30]. The LUCAS database contains stratified and repeated survey records of local land-cover and land-use types for Europe [30] with the latest one being available for the year 2018. We took the lucas survey records and reclassified the land-cover type (”$lc1_{l}able$”) to the natural MAES ecosystem categories, discarding all anthropogenic created habitats (Cropland, Urban, ...) in the process.

As predictors for the potential habitat modelling, we considered data on the potential distribution of land cover ([31], [19]) as well as long-term average climatic conditions in Europe where we used downscaled bioclimatic ERA5 indicators over the last 40 years (1979 to 2018) from the European Copernicus program. These climatic indicators represent Essential Climate Variables (ECV) such as the surface energy, drought or moisture all of which are known to be important factors in delineating the range and environmental niche. Specifically, we made use of the BIOCLIM data BIO01 to BIO19, average aridity and cloud cover, the annual sum of frost days, potential evaporation and volumetric soil water as well as different characterizations of the number of growing degree days and the start, end and length of the growing season. We also included a predictor that quantified the Euclidean distance to the ocean from each terrestrial grid cell, given the importance of some wetland habitats to brackish water and coastal conditions. We prepared data groundwater and soil conditions, specifically data on the Ph value and Calcium Carbonate content of groundwater resources as well as estimates of the depth to groundwater in meters. We also included data from a thematic layer of a European soil lithology classification system [32] owing to the importance of difference soil types. The individual lithology classes were included as factorial combinations in the modelling.

We estimated the potential distribution of the habitat by relating and contrasting presence-only and pseudo-absence points with another [33]; [34]. This is a widely applied approach usually for species distribution modelling, which although being scale-dependent, is in this context desirable since our aim is to map the potential natural distribution of a habitat. Although it is possible to create predictions of potential natural habitat as multi-nominal problem, e.g. where each class has a different and exclusive probability to potentially occur [29], we decided instead to estimate the distribution of the habitat non-exclusively, since in many areas of Europe there is a potential for more than one habitat to potentially occur under natural conditions especially when succession trajectories are unknown. Thus, for each habitat dataset we created pseudo-absence points randomly distributed within the modelling domain. We rasterized these 10-km Article 17 data and applied a distance transform to them, e.g. there is a monotonically decreasing probability of a habitat type to occur outside the Article 17 reporting data. The resulting layer was then included as additional predictor.

We used non-linear and tree-based Bayesian Regression Trees (BART) for the estimation of the potential habitat, which have the benefit of being able to quantify complex non-linear interactions between variables and the consideration of prior information in a Bayesian framework [35], [36]. For the regression trees we used a logistic model formulation of the response by assuming the habitat presence and pseudo-absences to be Bernoulli distributed, e.g. $y_{habitat}$ $Pr(y = 1|x|)$.

We fitted the BART models with 500 tree and 50 burn-in iterations across four MCMC chains through the use of the ‘dbarts’ R-package [35]. From the resulting posterior of the fitted model and for each grid cell, we summarized the median and lower (5%) and upper (95%) percentile of the posterior, thus allowing us to spatially represent the uncertainty of each individual habitat type prediction. The resulting predictions thus contain an estimate of the probability of a potential occurrence of a given natural habitat for each 10km grid cell.

6.5 Biodiversity data

6.5.1 Biodiversity data collation

Openly available biodiversity data sources in Europe are heterogeneous in type, format and purpose; and to be able to use them in an integrated SDM type of approach, a considerable amount of data harmonization and format control is necessary. Throughout we followed the taxonomic “backbone” of GBIF and codified functions to harmonize and match taxonomic names from different data sources to the GBIF taxonomy backbone of 2021, (GBIF Secretariat (2021)). We focused throughout on terrestrial species listed in the EU Article 12 (Birds directive) and Article 17 (Habitats directive).

Firstly, we obtained presence-only records from GBIF for all animal and plant species in the database. We excluded fossil specimens, and those with invalid spatial coordinates, and applied standard data pre-preprocessing steps for unstructured citizen science data using the ‘CoordinateCleaner‘ R package [37]. We removed duplicate points (those within a 2-km distance within the same year) and highly uncertain records. Additionally, we collated taxonomic group specific data for bird and plant species. For birds we made use of eBird data, which we processed similarly as GBIF records above. Potential absence data were inferred from sites where full communities were recorded and for which the focal species had never been recorded. To further limit the number of total absence sites, we first took eBird sites where the species had been recorded, and spatially buffered these by 200 km, excluding any sites within this buffer zone. From the remaining potential absence sites, we randomly selected an equal number of absences as presences sites in which the species was recorded, up to a maximum of 500 per species. For plant species, we obtained presence-absence data from comprehensively inventoried vegetation plot data collated on the SPlotOpen database [38]. We filtered these data to Europe, and the representative subset species, as well as excluding any observations prior the year 2000 as above, and to records with a positional uncertainty of less than 2 km. We inferred absence data similarly to eBird but using a lower maximum of 100 absences at maximum, because of the smaller size of this dataset.

We furthermore obtained from the (Natura2000 database), a source commonly used for biodiversity accounting in Europe. For the course of this work, we treated the species information reported at various sites (polygons) in the Natura2000 network as presence-absence information, recognizing that surveys in some sites might be incomplete or outdated and not all species are necessarily recorded during. A R-package was created to format these data for the modelling (https://github.com/iiasa/rN2000). We supplemented these data with species checklists for Important Bird Areas (IBAs, [39]) across Europe, adding presence-absence records per species for the polygon area of the IBAs where the species was recorded or not.

We further obtained polygonal global, European, and Mediterranean species ranges from the IUCN Red List version 2021-2 (IUCN 2021) and from BirdLife International (BirdLife International and Handbook of the Birds of the World 2020). These data were filtered to only include areas where species were recorded as extant, possibly extant, or possibly extinct, and included all seasonal occurrences. Where existing for a given species, we further compiled habitat preference (land-cover and elevation) and threat information using data from the IUCN Red List (IUCN 2021) and (EEA preferences). Those estimates were linked to species-specific priors (see below).

Finally, we also obtained the 2020 spatial distribution data for birds listed in the Article 12 Birds Directive and for animals and plants listed in the Article 17 Habitats Directive, excluding sensitive species. Although these data do include population estimates recorded at the (sometimes sub-) Member State level, for this work we used these data only as occurrence data recorded as presence-only atlas data on a 10-km grid across Europe. For each species with sufficient data we fitted integrated species distribution models (see below). For those listed in the directives and for which we could not find enough raw data to fit a distribution model (see below), we used the provided distribution data from the EEA.

6.6 Predictor variables

For the statistical modelling to develop data we prepared a series of environmental predictors related to topography, soil conditions, climate and land cover.

In both planning and species distribution modelling we make use of land cover and land-use data that is consistent with the European Ecosystem accounting framework such as the Mapping and Assessment of Ecosystems and their Services (MAES) system. For the current distribution of land cover, we used data from the (Corine 2018 dataset). For mapping the potential natural distribution of a species (see below) we additionally considered data on the potential distribution of land cover ([19]; [29]) matched to the same legend. The thematic legend of the Corine land-cover data was then recategorized into different MAES categories through a crosswalk (Source), however, differed from the MAES categories as we split the class Pasture (2.3.1) from other Grasslands as considered by the MAES Grassland class. We furthermore separated the Forest class in the MAES categorization as "Natural", "Production" and "Multifunctional" forest following the stratification used in the G4M-GLOBIOM link (see above). These steps were taken to align the thematic legend of the MAES level 2 categorization more with the legend returned from GLOBIOM and furthermore to allow for some level of management intensity in the distinguishment of the classes.

In addition to land cover and land use we also considered data on topography of European landscapes making use of the (EU DEM ver1.1). We considered the mean elevation (in m), the topographic position index (TPI, [40]) and the aspect of the topography, which we transformed into eastness and northness estimates through a sinusoidal and cosinus transformation respectively. This transformation is necessary to avoid circularities in units caused by the degrees-based characterization of an aspect layer.

We used long-term average climatic conditions in Europe. Specifically, we leveraged downscaled bioclimatic ERA5 indicators over the last 40 years (1979 to 2018) from the European Copernicus program. These climatic indicators represent Essential Climate Variables (ECV) such as the surface energy, drought or moisture all of which are known to be important factors in delineating the range and environmental niche. Specifically, we made use of the BIOCLIM data BIO01 to BIO19, average aridity and cloud cover, the annual sum of frost days, potential evaporation and volumetric soil water as well as different characterizations of the number for growing degree days and the start, end and length of the growing season [41]. For those parts of European member states which are missing in the European downscaled Copernicus Climate products (Such as the Spanish Canary islands) we used the average values of the global rather than the downscaled climate product instead.

For current and future species projections of species distributions we considered only variables related to land cover as well as temporally static variables which are unlikely to change in future scenarios such as for instance altitude. On the other hand, for predictions of the potential natural distribution of a species we made use of all predictors excluding those related to land cover and land use. Further, not all predictors were used for each species-specific prediction. For bird species we additionally included a layer depicting the Euclidean distance to the ocean from each land grid cell, given the importance of marine waters to many onshore nesting grassland and wetland species. For plant species we furthermore included spatial-explicit predictors related to groundwater and soil conditions, specifically data on the Ph value and Calcium Carbonate content of groundwater resources as well as estimates of the depth to groundwater in meters from [42]. In particular, groundwater Ph and Calcium carbonate do not cover small islands (Madeira, Canary Islands) as well as Cyprus, which is why we filled any remaining missing values of the predictors with a spatial prediction from a random forest model, using spatial proximity and climatic variables as covariates [29]. We also included data from a thematic layer of a European soil lithology classification system owing to the importance of different soil types to growth and niche space of plant species. The individual lithology classes were included as factorial combinations in the modelling.

Most observational species occurrence points in Europe are known the be biased towards areas with higher accessibility and wealth, with critically Easter European member states having a comparably lower density of records compared to western and northern European member states. Besides the integration of multiple datasets, we attempted to control for such sampling biases through a model-based control following [43]. To do so, we first took the presence-only and presence-absence localities of all biodiversity sources considered in this work (see above) and rasterizing them for the target background, resulting in a counted number of points for any coarser grained grid cell, which we then aggregated overall as the total sum of all occurrences. Furthermore, we prepared data on the accessibility of land ([44]) and the human population density per grid cell using data from the GHSL product for Europe ([45]). The biased background grid was then created by first calculating an adjusted log transformation of each individual layer through the following equation ($log(1 + x - min(x) * w)$), where w stands for a numerical weight reflective of the direction of bias (-1 for accessibility, 1 for all others), and afterwards the individual layers normalized to a range of 0 to 1 and averaged.

Finally, all layers were aggregated from their original resolution to a 10-km (Lamberts Equal Area projection) grain size determined by the background modelling domain layer by either calculating the proportion of grid cells in each coarser grain if these binary data, by calculating the bilinear resampled average of all values within the coarser grain, or – in the case of multinominal categorical data – calculating the mode of all finer grained classes. For the modelling (see below) all continuous variables were standardized to the mean and divided by their standard deviation, thus ensuring compatibility in terms of units. All covariates were matched against the modelling extent and made consistent with NUTS2 representation of European countries. Any missing data at the pixel level not filled or extrapolated at this stage was filled with missing data across all covariates. All calculations and variable preparations were done in the Lambertz equal-area projection for Europe and aggregated using GDAL and R packages such as “Raster” or “sf”.

6.6.1 Current distribution of a species

We estimated the distribution of species using an integrated species distribution modelling (iSDM) approach where different best-available data sources (occurrence, preference, expert information) are integrated into one joint prediction using different types of linear and non-linear modelling approaches.

We collated for each species the available suitable data (see above), separating between different types, namely presence-only, presence-absence and presence-only data to be used in the form of spatial-explicit offsets (e.g. expert range). For species occurring in Natura 2000 or Important Bird area (IBA) sites, we assumed that the species occur in all Natura 2000 sites in which a presence was indicated and that suitable habitats are homogeneous within Natura 2000 sites (which, given the small size of many sites and the SDM modelling grain of 10km is a reasonable assumption). We sampled at random across all sites presence point estimates up to a number of two-times as many as there are other occurrence observations (see above), to broadly characterize the environmental conditions prevalent in those sites. We furthermore created an equal number of absence point data which we sampled at random across all sites excluding the ones where the species has been recorded as present. Because of computational reasons and to further reduce sampling biases, we applied thinning on point data for all species with more than 200 records. The process of spatial thinning removes occurrence points at random from areas that are oversampled, for example because of sampling or spatial biases in the database [46]. Notably thinning only removes points from grid cells where there are multiple and never removes all points from any grid cell completely. For presence-only records from GBIF – usually the largest data source by size – we first applied a bias thinning, where we preferentially removed observations from 10km grid cells considered as biased (based on occurrence information across all species). In addition, and across all point occurrence datasets, we also removed at random observations until a minimum number of 10 points at maximum has been reached. This approach ensures that presence and absence information (where existing) are relatively homogeneously distributed in density across the European land area, thus representing average conditions for suitable species habitats across the modelling period.

Whenever presence-only atlas or expert-delineated information on the occurrence of species existed, such as for example from the global, Mediterranean or European IUCN assessments, the SER Amphibian and Reptile atlas, the Atlas Hymenoptera data or for polygon information from the EU Habitats directive or Bird directive data, we included this information as spatial offsets. For IUCN we only used those parts of the range where the species is permanently resident or which are part of its breeding distribution. We then followed an approach by Domisch et al. and first binarized the range estimates and then calculated the Euclidean distance from the boundary of the range to all other grid cells in the modelling background [47]. The resulting distance layer was then rescaled to a range from 0 (furthest away) to 1 (within the range). Notably we used different approaches depending on whether bird or non-bird species were estimates, using infinite and 20-km distance transformations respectively. All offsets created in this way were log-transformed before adding them to any model using presence-only information and in the case multiple offsets were supplied, these were combined first via simple multiplication.

To avoid overprediction into novel areas, the predictions were spatially constrained by a broad environmental zoning layer [12]. This was done by removing broad zones in which there are no contemporary occurrence points to avoid, for instance, extrapolations from a Mediterranean into boreal climatic zones, while also allowing modest extrapolations within similar environmental conditions. It should be noted that this zoning was only included for current projections and not for any future scenarios. We tested for co-linearity between included variables, removing those that were highly co-linear (Pearson's r ¿ 0.7) unless they were known to be of particular importance to a species.

We estimated the potential distribution of the species through an ensemble modelling approach (stacked SDM, [48] [49])) using state-of-the-art machine learning and Bayesian algorithms that complement each other's strengths. Model structure and response were determined based on data type, with Poisson Process models being fitted for presence-only datasets and logistic regressions for presence-absence data. We fitted tree-based regressions using the XGBoost modelling approach [50]. XGBoost makes use of gradient descent boosting, supports variable regularization and also non-linear tree-based regressions. XGBoost Models were fitted using a total of 10000 boosting iterations, a learning rate of 0.001 and Gamma parameter of 4 (larger is more conservative) for regularization. We also used another gradient descent boosting algorithm (GDB) available from the 'mboost' R-package [51]. GDB models makes use of non-linear baselearners (splines) for additive inference similar to the popular Generalized Additive Models (GAMs), however in contrast to GAMs it also supports variable regularization directly through boosting and additional baselearners (see below on priors). Here models were fitted using a total of 2500 boosting iterations and a learning rate of 0.001 per iteration. Bayesian regularized regressions were fitted using the 'BoomSpikeSlab' R-package (Scott 2022) and 10 000 MCMC iterations and four MCMC chains. Lastly we used non-linear and tree-based Bayesian Regression Trees (BART) for estimation which, compared to the other two models, have the benefit of being able to quantify complex non-linear interactions between variables and the consideration of prior information in a Bayesian framework. We fitted the BART models with 200 tree and 50 burn-in iterations across four MCMC chains through the use of the 'dbarts' R-package [35].

Models were only fitted for those species for which at least – after thinning - 20 data points were available, assuming that species with fewer records have not been sampled comprehensively enough to make inferences about their potential distribution. Further, we made use of simple rules to avoid fitting overly complex models for limited number of observations. Only linear models (boosted and non-boosted) were fitted for species with fewer than 100 observational points and for species with point observations fewer than 1.5 times the minimum data size of 20, we did only fit linear Bayesian Poisson Process models and not use any non-linear or boosted approaches to avoid overfitting [52]. Linear regressions, compared to non-linear ones, usually fare better when the goal is extrapolation and are less prone to model overfitting. In case only presence-only information from GBIF was available for a species, we furthermore included spatial effects as covariate using polynomial-transformed coordinates [53].

Integrating prior information on species habitat and elevational preferences and distances to known occurrences can improve range estimates. We obtained information on the susceptibility of species to certain habitat and elevational preferences from the EEA habitat preference database and IUCN. Preferences to certain land-cover types in the IUCN habitat preferences or respectably Corine land-cover categories were remapped to MAES categories. Priors can help to stabilize and avoid mapping unrealistic response functions to certain covariates [54]. Elevational preferences were included as specific threshold transforms on the raw elevation data and were used instead of the raw elevation data instead. This approach thus creates two separate discrete elevational bounds in which a species might or might not exist. For GDB and xgboost we used monotonicity constrained priors [55] assuming either increasing, in case the habitat was preferred or suitable, or positive constraints. We specified monotonically constrained baselearners for both XGBoost and GDB [55], which are constrains placed on the linear and non-linear effects to follow certain directions. Previous studies have shown that the use of such monotonicity constrains in SDMs can results in more ecological plausible response functions [55]. For Bayesian Poisson-Process models we used Zellner-style spike and slab priors with two parameters, a coefficient for a Gaussian prior on the mean coefficient of the covariate and a inclusion probability which states the probability by which a certain variable is to be used or allowed to be regularized from the model [56]. For habitats preferred by a species we used mean coefficients of 3 and an inclusion probability of 1, for suitable habitats we used a coefficient of 1 and 0.5 respectively and for occasionally occupied habitats we used 0.1 in both cases. Similarly, for BART models we specified priors as transition probabilities for the variable so that the regression tree is forced - with a certain probability, here 0.75 –

to generate a split for a given variable.

On the full point occurrence dataset, we then applied a spatial block cross validation scheme using the blockCV R-package [57] [58]. Specifically, we created three spatial folds of training and testing data to evaluate each of the three algorithms on. However for species with very few occurrence records overall (¡ 50) and where the creation of spatial folds failed (owing to points being too close in distance), we instead implemented default randomized folds where 25% of data was removed at random. All predictors were scaled prior to model fitting by subtracting the mean from each value and dividing by their standard deviation to ensure comparable unit scales. We included among the final predictors also the bias variable (see above), which was controlled during the prediction [43], thus helping to reduce some of the spatial biases in available occurrence datasets. Ensemble of different datasets per species were integrated and thus separate models were estimated for each spatial block and for each data type [59].

Each separate model prediction was binarized using a 0.05 percentile threshold and then validated using the withheld data to obtain an estimate of the True-Skill-Statistic (TSS). We used the TSS values to create a weighted mean ensemble of all predictions. Individual predictions from different models were first normalized before this step owing to the different units (relative rate of occurrence compared to relative occurrence probability). We then thresholded all ensembles using a 5% minimum percentile threshold on all observed data points (across datasets), thus creating a conservative estimate of where suitable habitat for a species might or might not persist in Europe [52]. We used percentile-based thresholds [12] opposed to approaches maximizing any performance metric since they can be applied across different predictions and dataset types (presence-only and presence-absence). Furthermore it assumes that the least suitable habitat at which the species is known to occur is the minimum suitability value for the species, while allowing for some flexibility so that outliers do not bias the threshold. For each species the validation statistics, the predicted suitability and the thresholded map is then retained. All modelling was done by in the integrated modelling framework ibis.iSDM coded for R (Jung, 2022, https://iiasa.github.io/ibis.iSDM/).

6.6.2 Potential distribution of a species

The goal of the potential distribution modelling is to obtain a depiction of the potential distribution of the species (sensu [18]). We understand potential in this context as the contemporary climatic, soil and natural vegetation conditions that would allow a species to persist in an area. Critically, and opposed to the mapping of current suitable habitat, this approach considers only contemporary differences in climate and soil, and not any land-cover or land-use, aligning with the concept of potential natural vegetation of Europe [19]. The predictions from the species distribution models used here thus aim to depict where a species might exist under contemporary conditions, while also allowing modest inter- and extrapolation from its current distribution.

While for the current estimation of species distribution (see above) we considered each biodiversity data type separately, for the potential distribution of the species we merged presence-only and pseudo-absence points with each other [33] [34]. This is a widely applied approach for SDM mapping, which however is scale dependent and can result in an overestimation of the niche [60] [54], yet in this context this is acceptable given that our aim is to map the widest possible potential distribution of species (although we modestly constrain the prediction, see below). We first combined all cleaned and filtered point occurrence data into one joint dataset, removing duplicates per grid cell in the process. Although there can be benefits in modelling these datasets jointly for more constrained predictions [61], our aim is to identify and characterize the maximum potential extent of the environmental niche of a species given contemporary conditions. For each dataset we created pseudo-absence points randomly distributed within the modelling domain, but spatially weighted them so that pseudo-absences preferentially fall into areas with high bias (see above).

We used a similar ensemble modelling approach as for the current estimates of species ranges (see above), however used logistic regressions throughout, using pseudo-absence points for presence-only datasets. Further for validating the ensemble models we evaluated the predictions in terms of their accuracy through the F1 score, which is calculated as the ratio of the model precision (true positives) and the recall (also known as sensitivity). We specifically chose the F1 score for evaluation since it maximizes correct predictions and thus can help to ensure that most training occurrence points are retained. The final ensemble prediction was then created as a weighted mean of the nine different F1 scores (3 spatial blocked subsets per algorithm). All modelling was done by in the integrated modelling framework ibis.iSDM coded for R (Jung 2022, https://iiasa.github.io/ibis.iSDM/).

6.7 Carbon data

6.7.1 Current distribution of carbon

For current carbon stocks we used data on above-ground, below-ground and soil organic carbon at risk from land-use change from [12]. These data were created by selecting and integrating best available carbon maps for different vegetation classes. For more detail on the integration and handling of individual data layers see [12]. All data are in units of tC/ha and for the analysis we combined the current carbon layers by calculating the combined sum of above- & below-ground and soil organic carbon for Europe and included it as an additional feature in the prioritization.

6.7.2 Potential carbon

To spatially allocate specific restoration priorities, we would need to identify areas with high carbon sequestration potential. Here we used an approach that combined the different techniques from [14] and [20]. First, we created a regular sampling grid at 1km for each current MAES ecosystem type reclassified from Corine 2018 (see above) and extracted the fraction of the respective land cover type. We then extracted estimates of current reference carbon data (in tC/ha) from the above, below and soil organic carbon data layers from [12] as well as from the European specific JRC Biomass map [62] and the FAO Soil organic carbon map (ver 1.5.0). For each of the different types of carbon products (below, above and soil carbon) we then calculated a consensus estimate (arithmetic mean) per gridded 1km point. We further extracted estimates on whether a given point locality was situated in a peat land as considered by the European peatland map or land covered by primary forest [16]

From the resulting extracted estimates, we then selected for each natural land cover type (e.g. Grassland, Heathland, Marine inlets, Sparsely vegetated land, Wetland and Forest Woodland) a total for 10000 reference points for modelling training. We ensured that (a) the respective land cover type currently covers at least 50% of a given 1km grid cell, (b) average carbon density estimates are in the largest 95% percentile of values, (c) the reference points were preferentially sampled in remaining European peat and primary forest sites for the Wetland, Marine Inlets and Forest & Woodland classes, (d) points were geographically representative by covering each European biogeographical regions (adjusted for area) and (e) that extracted mean carbon density estimates were corrected for the fraction of non-natural land contained within them. Instead of using a single reference value for the carbon contained in non-natural systems [14], we calculated the average estimate of all non-natural land cover types in MAES (e.g. cropland and urban).

As above for potential species distributions and land cover, we then subjected these reference estimates to a spatial extrapolation approach. Here we followed an approach set out by [20] and estimated potential carbon density as $CDensPo = f(S, T, C)$, where potential carbon density is predicted as a function of soil, S, topographic, T, and climatic, C, factors. We used the same predictors as for potential land cover and species occurrences. Finally, we corrected each estimate by the amount of potentially occurring natural land cover.

7 References

References

[1] Georgina M Mace, Mike Barrett, Neil D Burgess, Sarah E Cornell, Robin Freeman, Monique Grooten, and Andy Purvis. Aiming higher to bend the curve of biodiversity loss. *Nature Sustainability*, 1(9):448–451, 2018.

[2] Sandra Díaz, Unai Pascual, Marie Stenseke, Berta Martín-López, Robert T Watson, Zsolt Molnár, Rosemary Hill, Kai MA Chan, Ivar A Baste, Kate A Brauman, et al. Assessing nature's contributions to people. *Science*, 359(6373):270–272, 2018.

[3] Convention on Biological Diversity. First draft of the post-2020 global biodiversity framework. 2021.

[4] Haigen Xu, Yun Cao, Dandan Yu, Mingchang Cao, Yuxiao He, Michael Gill, and Henrique M Pereira. Ensuring effective implementation of the post-2020 global biodiversity targets. *Nature Ecology & Evolution*, 5(4):411–418, 2021.

[5] Lindsay M Dreiss and Jacob W Malcom. Identifying key federal, state and private lands strategies for achieving 30x30 in the us. *bioRxiv*, 2021.

[6] Sean L Maxwell, Victor Cazalis, Nigel Dudley, Michael Hoffmann, Ana SL Rodrigues, Sue Stolton, Piero Visconti, Stephen Woodley, Naomi Kingston, Edward Lewis, et al. Area-based conservation in the twenty-first century. *Nature*, 586(7828):217–227, 2020.

[7] Robert L Pressey, Piero Visconti, Madeleine C McKinnon, Georgina G Gurney, Megan D Barnes, Louise Glew, and Martine Maron. The mismeasure of conservation. *Trends in Ecology & Evolution*, 36(9):808–821, 2021.

[8] Christopher Robert Margules and Robert L Pressey. Systematic conservation planning. *Nature*, 405(6783):243–253, 2000.

[9] Emma J McIntosh, Robert L Pressey, Samuel Lloyd, Robert J Smith, and Richard Grenyer. The impact of systematic conservation planning. *Annual Review of Environment and Resources*, 42:677–697, 2017.

[10] JO Hanson, R Schuster, N Morrell, M Strimas-Mackey, ME Watts, P Arcese, J Bennett, and HP Possingham. prioritizr: Systematic conservation prioritization in r. *R package version*, 3(1.1), 2017.

[11] Li Zhu, Alice C Hughes, Xiao-Qian Zhao, Li-Jing Zhou, Ke-Ping Ma, Xiao-Li Shen, Sheng Li, Ming-Zhang Liu, Wu-Bing Xu, and James EM Watson. Regional scalable priorities for national biodiversity and carbon conservation planning in asia. *Science Advances*, 7(35):eabe4261, 2021.

[12] Martin Jung, Andy Arnell, Xavier De Lamo, Shaenandhoa García-Rangel, Matthew Lewis, Jennifer Mark, Cory Merow, Lera Miles, Ian Ondo, Samuel Pironon, et al. Areas of global importance for conserving terrestrial biodiversity, carbon and water. *Nature Ecology & Evolution*, 5(11):1499–1509, 2021.

[13] Constance Fastré, Willem-Jan van Zeist, JEM Watson, and Piero Visconti. Integrated spatial planning for biodiversity conservation and food production. *One Earth*, 4(11):1635–1644, 2021.

[14] Bernardo BN Strassburg, Alvaro Iribarrem, Hawthorne L Beyer, Carlos Leandro Cordeiro, Renato Crouzeilles, Catarina C Jakovac, André Braga Junqueira, Eduardo Lacerda, Agnieszka E Latawiec, Andrew Balmford, et al. Global priority areas for ecosystem restoration. *Nature*, 586(7831):724–729, 2020.

[15] European Commission. Natura 2000 data and maps. 2020.

[16] Francesco Maria Sabatini, Hendrik Bluhm, Zoltan Kun, Dmitry Aksenov, José A Atauri, Erik Buchwald, Sabina Burrascano, Eugénie Cateau, Abdulla Diku, Inês Marques Duarte, et al. European primary forest database v2. 0. *Scientific data*, 8(1):1–14, 2021.

[17] Petr Havlík, Hugo Valin, Aline Mosnier, Stefan Frank, Pekka Lauri, David Leclère, Amanda Palazzo, Miroslav Batka, Esther Boere, Albert Brouwer, et al. Globiom documentation. *International Institute for Applied Systems Analysis, Laxenburg, Austria*, 2018.

[18] Tomislav Hengl, Madlene Nussbaum, Marvin N Wright, Gerard BM Heuvelink, and Benedikt Gräler. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6:e5518, 2018.

[19] U. Bohn, G. Gollub, C. Hettwer, Z. Neuhäuslová, T. Raus, H. Schlüter, and H. Weber. Karte der natürlichen vegetation europas/map of the natural vegetation of europe. 2004.

[20] Wayne S Walker, Seth R Gorelik, Susan C Cook-Patton, Alessandro Baccini, Mary K Farina, Kylen K Solvik, Peter W Ellis, Jon Sanderman, Richard A Houghton, Sara M Leavitt, et al. The global potential for increased storage of carbon on land. *Proceedings of the National Academy of Sciences*, 119(23):e2111312119, 2022.

[21] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2018.

[22] Michael Coppedge, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, Steven Fish, Adam Glynn, Allen Hicken, et al. Varieties of democracy. *Dataset v7*, 2017.

[23] Oskar Rydén, Alexander Zizka, Sverker C Jagers, Staffan I Lindberg, and Alexandre Antonelli. Linking democracy and biodiversity conservation: Empirical evidence and research gaps. *Ambio*, 49(2):419–433, 2020.

[24] Melissa S. Chapman, William K. Oestreich, Timothy H. Frawley, Carl Boettiger, Sibyl Diver, Bianca S. Santos, Caleb Scoville, Katrina Armstrong, Hannah Blondin, Kevin Chand, Danielle E. Haulsee, Christopher J. Knight, and Larry B. Crowder. Promoting equity in the use of algorithms for high-seas conservation. *One Earth*, 4(6):790–794, 2021.

[25] Conservation needs to integrate knowledge across scales., journal = Nature ecology evolution, volume = 6, number = 2, pages = 118-119, year = 2022, author = Rebecca Chaplin-Kramer and Kate A. Brauman, Jeannine Cavender-Bares and Sandra Díaz and Gabriela Teixeira Duarte and Brian J. Enquist and Lucas A. Garibaldi and Jonas Geldmann and Benjamin S. Halpern and Thomas W. Hertel and Colin K. Khoury and Joana Madeira Krieger and Sandra Lavorel and Thomas Mueller and Rachel A. Neugarten and Jesús Pinto-Ledezma and Stephen Polasky, Andy Purvis and Victoria Reyes-García and Patrick R. Roehrdanz and Lynne J. Shannon and M. Rebecca Shaw and Bernardo B. N. Strassburg and Jason M. Tylianakis and Peter H. Verburg and Piero Visconti and Noelia Zafra-Calvo.

[26] Valentin Popov, Payal Shah, Rebecca K Runting, and Jonathan R Rhodes. Managing risk and uncertainty in systematic conservation planning with insufficient information. *Methods in Ecology and Evolution*, 13(1):230–242, 2022.

[27] Joshua J Lawler, D Scott Rinnan, Julia L Michalak, John C Withey, Christopher R Randels, and Hugh P Possingham. Planning for climate change through additions to a national protected area network: implications for cost and configuration. *Philosophical Transactions of the Royal Society B*, 375(1794):20190117, 2020.

[28] Maria Beger, Anna Metaxa, Arieanna C. Balbar, Jennifer A. McGowan, Remi Daigle, Caitlin D. Kuempel, Eric A. Treml, and Hugh P. Possingham. Demystifying ecological connectivity for actionable spatial conservation planning. *Trends in Ecology and Evolution*, 2022.

[29] Tomislav Hengl, Markus G Walsh, Jonathan Sanderman, Ichsani Wheeler, Sandy P Harrison, and Iain C Prentice. Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. *PeerJ*, 6:e5457, 2018.

[30] Raphaël d'Andrimont, Momchil Yordanov, Laura Martinez-Sanchez, Beatrice Eiselt, Alessandra Palmieri, Paolo Dominici, Javier Gallego, Hannes Isaak Reuter, Christian Joebges, Guido Lemoine, et al. Harmonised lucas in-situ land cover and use database for field surveys from 2006 to 2018 in the european union. *Scientific data*, 7(1):1–15, 2020.

[31] Tomislav Hengl, Martin Jung, and Piero Visconti. Potential distribution of land cover classes (Potential Natural Vegetation) at 250 m spatial resolution, January 2020.

[32] Gergely Tóth, Luca Montanarella, Vladimir Stolbovoy, Ferenc Máté, Katalin Bódis, Arwyn Jones, Panos Panagos, and Marc Van Liedekerke. Soils of the european union. *JRC Scientific and Technical Reports, Office for Official Publications of the European Communities, Luxembourg*, 2008.

[33] Antoine Guisan and Wilfried Thuiller. Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9):993–1009, 2005.

[34] Morgane Barbet-Massin, Frédéric Jiguet, Cécile Hélène Albert, and Wilfried Thuiller. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in ecology and evolution*, 3(2):327–338, 2012.

[35] Vincent Dorie, Hugh Chipman, Robert McCulloch, Armon Dadgar, R Core Team, Guido U Draheim, Maarten Bosmans, Christophe Tournayre, Michael Petch, Rafael de Lucena Valle, et al. Package 'dbarts'. 2022.

[36] Colin J Carlson. embarcadero: Species distribution modelling with bayesian additive regression trees in r. *Methods in Ecology and Evolution*, 11(7):850–858, 2020.

[37] Alexander Zizka, Daniele Silvestro, Tobias Andermann, Josué Azevedo, Camila Duarte Ritter, Daniel Edler, Harith Farooq, Andrei Herdean, María Ariza, Ruud Scharn, et al. Coordinatecleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5):744–751, 2019.

[38] Francesco Maria Sabatini, Jonathan Lenoir, Tarek Hattab, Elise Aimee Arnst, Milan Chytrỳ, Jürgen Dengler, Patrice De Ruffray, Stephan M Hennekens, Ute Jandt, Florian Jansen, et al. splotopen–an environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography*, 30(9):1740–1764, 2021.

[39] Paul F Donald, Lincoln DC Fishpool, Ademola Ajagbe, Leon A Bennun, Gill Bunting, Ian J Burfield, Stuart HM Butchart, Sofia Capellan, Michael J Crosby, Maria P Dias, et al. Important bird and biodiversity areas (ibas): the development and characteristics of a global inventory of key sites for biodiversity. *Bird Conservation International*, 29(2):177–198, 2019.

[40] Jeroen De Reu, Jean Bourgeois, Machteld Bats, Ann Zwertvaegher, Vanessa Gelorini, Philippe De Smedt, Wei Chu, Marc Antrop, Philippe De Maeyer, Peter Finke, et al. Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, 186:39–49, 2013.

[41] H. Wouters, J. Berckmans, R. Maes, E. Vanuytrecht, and K. De Ridder. Downscaled bioclimatic indicators for selected regions from 1979 to 2018 derived from reanalysis, version 1.0, 2021.

[42] Michal Hájek, Borja Jiménez-Alfaro, Ondřej Hájek, Lisa Brancaleoni, Marco Cantonati, Michele Carbognani, Anita Dedić, Daniel Dítě, Renato Gerdol, Petra Hájková, et al. A european map of groundwater ph and calcium. *Earth system science data*, 13(3):1089–1105, 2021.

[43] David I Warton, Ian W Renner, and Daniel Ramp. Model-based control of observer bias for the analysis of presence-only data in ecology. *PloS one*, 8(11):e79168, 2013.

[44] D J Weiss, Andy Nelson, HS Gibson, W Temperley, Stephen Peedell, A Lieber, M Hancher, Eduardo Poyart, Simão Belchior, Nancy Fullman, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336, 2018.

[45] Martino Pesaresi, Guo Huadong, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen, Matina Halkia, Mayeul Kauffmann, Thomas Kemper, Linlin Lu, et al. A global human settlement layer from optical hr/vhr rs data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):2102–2131, 2013.

[46] Valerie A Steen, Morgan W Tingley, Peter WC Paton, and Chris S Elphick. Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12(2):216–226, 2021.

[47] Sami Domisch, Adam M Wilson, and Walter Jetz. Model-based integration of observed and expert-based information for assessing the geographic and environmental distribution of freshwater species. *Ecography*, 39(11):1078–1088, 2016.

[48] Matthias F Biber, Alke Voskamp, Aidin Niamir, Thomas Hickler, and Christian Hof. A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. *Journal of Biogeography*, 47(1):114–129, 2020.

[49] Roozbeh Valavi, Gurutzeta Guillera-Arroita, Jose J Lahoz-Monfort, and Jane Elith. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr*, 92:e01486, 2021.

[50] Yang Chen, Lixia Ma, Dongsheng Yu, Haidong Zhang, Kaiyue Feng, Xin Wang, and Jie Song. Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests. *Ecological Indicators*, 135:108545, 2022.

[51] Benjamin Hofner, Andreas Mayr, Nikolay Robinzonov, and Matthias Schmid. Model-based boosting in r: a hands-on tutorial using the r package mboost. *Computational statistics*, 29(1):3–35, 2014.

[52] Cory Merow, Matthew J Smith, and John A Silander Jr. A practical guide to maxent for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10):1058–1069, 2013.

[53] Sami Domisch, Martin Friedrichs, Thomas Hein, Florian Borgwardt, Annett Wetzig, Sonja C Jähnig, and Simone D Langhans. Spatially explicit species distribution models: A missed opportunity in conservation planning? *Diversity and Distributions*, 25(5):758–769, 2019.

[54] Henrik Hannemann, Katherine J Willis, and Marc Macias-Fauria. The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. *Global Ecology and Biogeography*, 25(1):26–35, 2016.

[55] Benjamin Hofner, Torsten Hothorn, Thomas Kneib, and Matthias Schmid. A framework for unbiased model selection based on boosting. *Journal of Computational and Graphical Statistics*, 20(4):956–971, 2011.

[56] Wen Cui and Edward I George. Empirical bayes vs. fully bayes variable selection. *Journal of Statistical Planning and Inference*, 138(4):888–900, 2008.

[57] Roozbeh Valavi, Jane Elith, José J Lahoz-Monfort, and Gurutzeta Guillera-Arroita. blockcv: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *bioRxiv*, page 357798, 2018.

[58] Hanna Meyer, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. Importance of spatial predictor variable selection in machine learning applications–moving from data reproduction to spatial prediction. *Ecological Modelling*, 411:108815, 2019.

[59] Robert J Fletcher Jr, Trevor J Hefley, Ellen P Robertson, Benjamin Zuckerberg, Robert A McCleery, and Robert M Dorazio. A practical guide for combining data to model species distributions. *Ecology*, 100(6):e02710, 2019.

[60] Changwan Seo, James H Thorne, Lee Hannah, and Wilfried Thuiller. Scale effects in species distribution models: implications for conservation planning under climate change. *Biology letters*, 5(1):39–43, 2009.

[61] Nick JB Isaac, Marta A Jarzyna, Petr Keil, Lea I Dambly, Philipp H Boersch-Supan, Ella Browning, Stephen N Freeman, Nick Golding, Gurutzeta Guillera-Arroita, Peter A Henrys, et al. Data integration for large-scale models of species distributions. *Trends in ecology & evolution*, 35(1):56–67, 2020.

[62] Valerio Avitabile and Andrea Camia. An assessment of forest biomass maps in europe using harmonized national statistics and inventory plots. *Forest ecology and management*, 409:489–498, 2018.