BIODIVERSITY RESEARCH

WILEY Diversity and Distributions

Spatially explicit species distribution models: A missed opportunity in conservation planning?

Sami Domisch[1] ⓘ | Martin Friedrichs[1,2] ⓘ | Thomas Hein[3] ⓘ | Florian Borgwardt[3] ⓘ |
Annett Wetzig[1] | Sonja C. Jähnig[1] ⓘ | Simone D. Langhans[1,4,5] ⓘ

[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

[2]Department of Biology, Freie Universität Berlin, Berlin, Germany

[3]Institute of Hydrobiology and Aquatic Ecosystem Management, University of Natural Resources and Life Sciences, Vienna, Austria

[4]Department of Zoology, University of Otago, Dunedin, New Zealand

[5]BC3 – Basque Centre for Climate Change, Leioa, Spain

Correspondence
Sami Domisch, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany.
Email: domisch@igb-berlin.de

Funding information
German Federal Ministry of Education and Research, Grant/Award Number: 01 LN1320A; H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 748625; European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 642317

Editor: Gwen Iacona

Abstract

Aim: Systematic conservation planning is vital for allocating protected areas given the spatial distribution of conservation features, such as species. Due to incomplete species inventories, species distribution models (SDMs) are often used for predicting species' habitat suitability and species' probability of occurrence. Currently, SDMs mostly ignore spatial dependencies in species and predictor data. Here, we provide a comparative evaluation of how accounting for spatial dependencies, that is, autocorrelation, affects the delineation of optimized protected areas.

Location: Southeast Australia, Southeast U.S. Continental Shelf, Danube River Basin.

Methods: We employ Bayesian spatially explicit and non-spatial SDMs for terrestrial, marine and freshwater species, using realm-specific planning unit shapes (grid, hexagon and subcatchment, respectively). We then apply the software GUROBI to optimize conservation plans based on species targets derived from spatial and non-spatial SDMs (10%–50% each to analyse sensitivity), and compare the delineation of the plans.

Results: Across realms and irrespective of the planning unit shape, spatially explicit SDMs (a) produce on average more accurate predictions in terms of AUC, TSS, sensitivity and specificity, along with a higher species detection probability. All spatial optimizations meet the species conservation targets. Spatial conservation plans that use predictions from spatially explicit SDMs (b) are spatially substantially different compared to those that use non-spatial SDM predictions, but (c) encompass a similar amount of planning units. The overlap in the selection of planning units is smallest for conservation plans based on the lowest targets and vice versa.

Main conclusions: Species distribution models are core tools in conservation planning. Not surprisingly, accounting for the spatial characteristics in SDMs has drastic impacts on the delineation of optimized conservation plans. We therefore encourage practitioners to consider spatial dependencies in conservation features to improve the spatial representation of future protected areas.

KEYWORDS
Bayesian hierarchical modelling, connectivity, GUROBI, integer linear programming, spatial autocorrelation, spatial unit

1 | INTRODUCTION

In the light of the ongoing decline in global biodiversity (Pimm et al., 2014), the implementation of protected areas in the terrestrial, marine and freshwater realms is yet the most widely used conservation approach to reduce the loss of biodiversity. Consequently, how to delineate protected areas so that they produce optimal outcomes for the targeted conservation features in a cost-effective way has been widely covered in the systematic conservation planning literature (Barnes, Glew, Wyborn, & Craigie, 2018; Margules & Pressey, 2000). In the past, conservation goals mainly reflected habitat or species representation. Only recently the focus has shifted towards considering environmental and ecological processes, which are essential for securing species persistence (e.g., Klein et al., 2009). Such processes shape the distribution and abundance of species (Pressey, Cabeza, Watts, Cowling, & Wilson, 2007) with connectivity playing a paramount and distinct role in terrestrial (Lockwood, 2010), marine (Carr et al., 2003), and freshwater ecosystems (Hermoso, Filipe, Segurado, & Beja, 2018). Incorporating spatial connectivity in the planning process has, therefore, important implications for designing protected areas (Daigle et al., 2018; van Teeffelen, Cabeza, & Moilanen, 2006; Weeks, 2017). This fact is also reflected in the software that is used in conservation planning, such as MARXAN (Ball, Possingham, & Watts, 2009) or ZONATION (Lehtomäki & Moilanen, 2013). All of them base, among other parameters, the selection of potential planning units on algorithms that account for their spatial connectivity.

When accounting for spatial connectivity, conservation plans inherently build the protected areas based on the spatial dependencies in the planning units as well. For example, it is vital to account for the spatial structure of the environment around a given planning unit for assessing its importance as part of a protected area (Daigle et al., 2018; Weeks, 2017). However, this key characteristic of spatial dependency is rarely applied in the underlying conservation features themselves, which provide the basis for the conservation planning.

The most widely used conservation feature in conservation planning is the geographical distribution of multiple species, as species distributions are often best known compared to other biodiversity facets (such as functional or phylogenetic characteristics of species; McGill, Dornelas, Gotelli, & Magurran, 2015). However, species point occurrence data alone are not useful in systematic conservation planning, which requires range-wide and seamless data of the targeted conservation features (Tulloch et al., 2016). This is likely one of the major challenges for conservation planners—to overcome the Wallacean shortfall, that is, to know the full geographical distribution of species (Bini, Diniz-Filho, Rangel, Bastos, & Pinto, 2006; Meyer, Weigelt, & Kreft, 2016).

Currently, species distribution models (SDMs) are the main tools used to produce such range-wide species distribution data (Guisan & Thuiller, 2005). SDMs assess habitat preferences in an environmental envelope given the species occurrences and the environmental predictors at the respective locations, and project a probabilistic habitat suitability index across the study area (Elith & Leathwick, 2009).

Species distribution models are considered useful in conservation planning when used within a "structured and transparent decision-making process" (Guisan et al., 2013). In reality, however, most conservation plans are still based on species' surrogates such as habitat maps, expert-derived species distributions, or connectivity surfaces (Tulloch et al., 2016). As reviewed by Tulloch et al. (2016), SDMs are not yet widely used in conservation planning because of several constraints: (a) poor availability of species occurrence data and (b) respective predictor data, (c) uncertainties associated with the data and (d) the wish to understand processes rather than patterns, requiring alternative and additional data. Here, (a–c) are clearly methodological issues that need to be assessed with care, while (e) goes beyond the expectations towards SDMs, since per definition, they do not deliver information on processes. Hence, it is key to distinguish between constraints that can be tackled (Dormann, 2007; Fourcade, Besnard, & Secondi, 2018; Pearson & Dawson, 2003), and those that are beyond the actual purpose of SDMs (Araújo & Peterson, 2012).

Species distribution models require rigorous testing for methodological issues and statistical shortcomings (Record, Fitzpatrick, Finley, Veloz, & Ellison, 2013; Tulloch et al., 2016). Given that spatial conservation plans incorporate spatial information among planning units, it is obvious that SDMs that yield the conservation feature data should make use of the same information. However, the few studies that modelled and used species distributions for conservation planning are typically not accounting for such spatial dependencies (e.g., in the terrestrial (Rondinini et al., 2011), freshwater (Esselman & Allan, 2011) and marine realms (McGowan et al., 2013), but see White, Schroeger, Drake, and Edwards (2014)). The disadvantage of such spatially implicit SDMs (hereafter referred to as "non-spatial SDMs") is that they do not account for spatial dependence or irregular sampling intensity (e.g., Latimer, Wu, Gelfand, & Silander, 2006), and frequently violate the assumption of independent samples and spatial units (Araujo, Pearson, Thuiller, & Erhard, 2005; Hampe, 2004). Assuming independence of spatial units means that the model is not aware of the predicted habitat suitability of a given species in neighbouring planning units. In contrast, the optimization of the spatial conservation plan, that is, the selection of planning units, does account for the spatial relation among them, favouring neighbouring ones over those that are far apart.

In contrast, spatially explicit SDMs ("spatial SDMs") account for the proximity and mobility (i.e., connectivity) in species populations. Hence, they provide more powerful inference about species distributions and niche relations (Latimer et al., 2006; De Marco, Diniz-Filho, & Bini, 2008). Incorporating the assumption of spatial dependencies in the data has, therefore, the potential to provide more robust SDM predictions (Guisan & Thuiller, 2005; Record et al., 2013). Most importantly from a conservation planning perspective, spatial SDMs make use of the spatial information and dependencies of the species and environment in the planning units. Habitat suitability predictions from spatial SDMs are generally more contiguous (Domisch, Wilson, & Jetz, 2016) and less patchy than those derived from non-spatial SDMs. The question remains, whether such

contiguous predictions have the potential to maximize the management efficiency of spatial plans.

Species distribution models can be made spatially explicit in numerous ways (see Dormann, 2007, for a review of various methods). Spatial aspects are included in the SDM given the spatial autocorrelation of species distributions (Besag, 1974; Record et al., 2013). Spatial autocorrelation is a common phenomenon in ecology, as nearby locations (given species occurrences and environmental predictors) tend to be more similar compared to those located farther apart (Legendre, 1993; Record et al., 2013). One possible method to add the spatial dimension in SDMs can be, for instance, through spatial random effects (Dormann, 2007). Spatial random effects can be computationally demanding. However, their use may be warranted—or even required—when the task is to limit the false positive and false negative predictions of probability of occurrence within a spatial unit (Record et al., 2013). Limiting false positive/negative predictions is crucial in conservation planning, since the over- or underestimation of suitable habitats (and therefore species occurrences) impacts

species protection itself and produces unnecessary costs, that is, for establishing protected areas in erroneous locations where the target species do not actually occur.

Here, we analyse whether accounting for spatial dependencies in both steps, the spatial prioritization process and in the underlying conservation features (here: species), influences the arrangement of potential protected areas. Using species survey data, we build nonspatial and spatial SDMs and compare the resulting mapped spatial conservation plans across a range of conservation targets, that is, proportions of potential species distribution areas. We hypothesize that (a) spatial SDMs would outperform non-spatial SDMs in terms of model evaluation scores, since spatial SDMs account for the influence of proximity in species populations; (b) this effect cascades through to the spatial conservation plans, and that those derived from spatial SDMs would differ significantly from the ones based on non-spatial SDMs. We test these hypotheses in three case studies covering terrestrial, marine and the freshwater realms. We do so, because protected areas in each realm are typically planned using
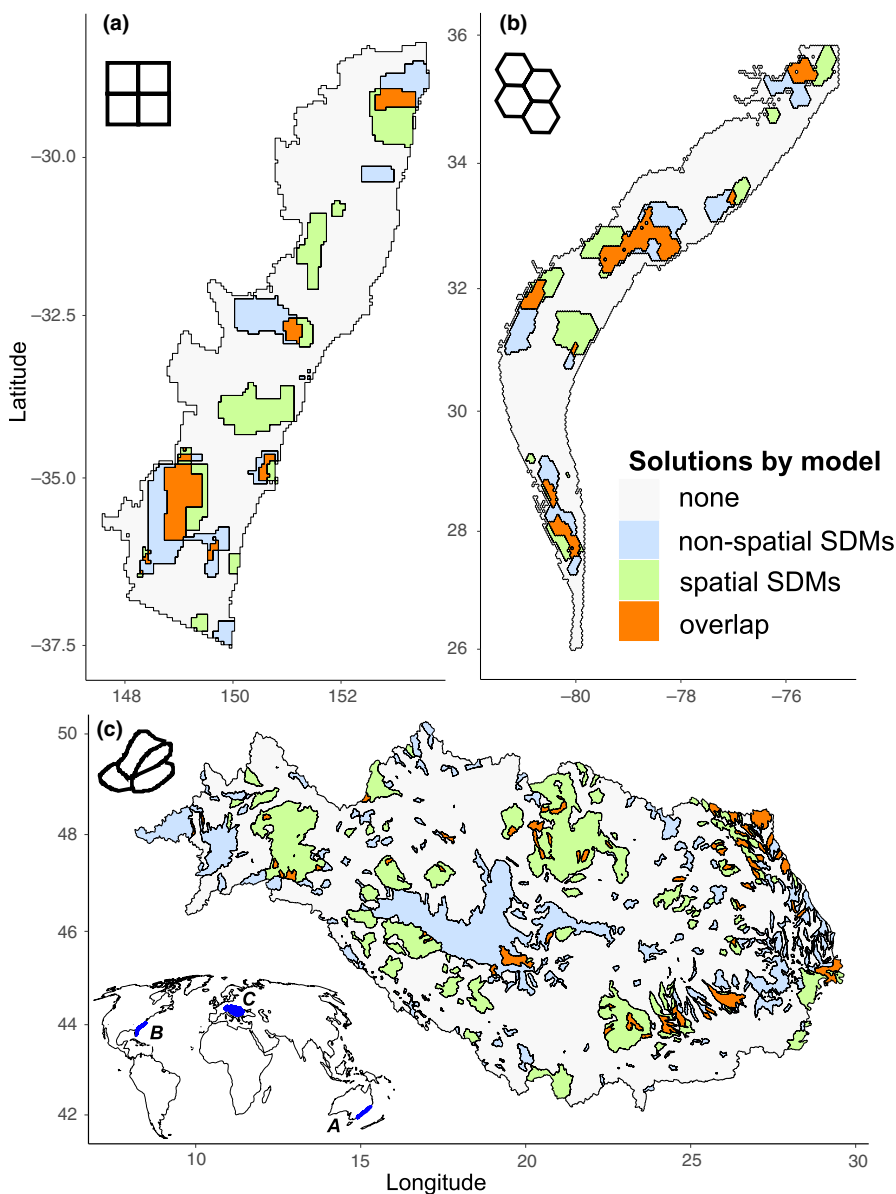

FIGURE 1 Spatial arrangement of potential protected areas derived from non-spatial and spatial SDMs, as well as their overlap in the terrestrial (a), marine (b) and freshwater realms (c), each using a specific planning unit shape (grids, hexagons, subcatchments). Spatial plans were defined for a 20% conservation target within a 10% gap to optimality. The inset shows the location of the study areas in Australia, US East Coast and the Danube river basin. See Supporting Information Figures S3–S5 for maps showing a wider range of conservation targets

different shapes of spatial units (grids, hexagons and subcatchments, respectively), and the planning unit shape has shown to impact the spatial pattern and effectiveness of protected areas (Nhancale & Smith, 2011).

2 | METHODS

2.1 | General workflow of the analysis

Our aim is to test whether, and if so, how the addition of spatial information in the SDMs leads to different species habitat suitability predictions, and hence to different spatial conservation plans. The analysis is divided into five stages: (a) preparing the data and aggregating it into spatial units; (b) testing the degree of spatial autocorrelation in the species data to see whether adding spatial random effects in the SDMs is warranted; (c) running two SDMs for each species, one that does not use the spatial information (non-spatial SDM) and another one that takes advantages of the spatial information via the spatial random effects (spatial SDM); (d) using the outputs from the spatial and the non-spatial SDMs as separate inputs to calculate spatially optimized conservation plans with varying conservation targets; and, finally, (e) comparing the spatial conservation plans derived from the non-spatial and the spatial SDMs. As the non-spatial and spatial SDMs are, except for the spatial random effects, identical, this allows us to single out the effect that the spatial information in SDMs has on the subsequent spatial conservation plans.

2.2 | Study areas and species data

The terrestrial case study (Figure 1a) comprised survey data of 33 *Eucalyptus* species in Eastern Australia derived from Fithian, Elith, Hastie, and Keith (2015). This dataset encompasses species detections/non-detections sampled annually from 1970 to 2013 as well as environmental data covering climate and topography and derivatives thereof (Fithian et al. 2015). Environmental data layers were re-sampled from the original 250-m spatial resolution by factor 20 to obtain a grid with the individual squares sizing 0.05° (approximately 25 km^2), using the RASTER package in R (Hijmans & van Etten, 2018; R-Core-Team, 2018). This yielded 7,763 grid cells that were used for all subsequent analyses.

The marine case study (Figure 1b) consisted of monitoring data of 53 reef fish species in the Southeast U.S. Continental Shelf region. Fish detections and non-detections were derived from the Southeast Reef Fish Survey (SERFS) that was conducted annually from 1990 to 2013 (Bacheler et al., 2014; SEAMAP-SA, 2017). Environmental data on ocean topography, currents, nutrients and light were obtained from the Bio-Oracle (approx. 9 km spatial grain at the Equator, Tyberghein et al., 2012; Assis et al., 2018) and MARSPEC datasets (approx. 1 km, Sbrocco & Barber, 2013). All environmental data were harmonized to 7,123 hexagonal grids of approx. 19 km^2, and the species data were summarized for each hexagonal grid cell (created using the R-package SP; Pebesma & Bivand, 2005; Bivand, Pebesma, & Gomez-Rubio, 2013).

The freshwater case study (Figure 1c) was based on detection/non-detection data of 85 fish species across the Danube river basin. Fish survey data were derived from the EFI+ and the BioFresh project databases, sampled during 1955–2007 (Schinegger, Pletterbauer, Melcher, & Schmutz, 2016; Zupancic, 2015). We used the HydroBASIN dataset (Lehner & Grill, 2013) and selected all level 12 subcatchments draining into the Black Sea, comprising of 7,376 subcatchments with an average size of 108 km^2. We then extracted climatic (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), topographic (Amatulli et al., 2018) and land cover variables (Tuanmu & Jetz, 2014, all at ~1 km spatial grain), as well as the number of dams (Lehner et al., 2011) for each subcatchment, the latter representing a proxy for human impact besides land use.

Species data, including species detections, non-detections and sampling dates, were aggregated to the planning units. Within each planning unit, environmental data were aggregated using various techniques (e.g., average, standard deviation; see Supporting Information Table S1). In the freshwater realm, we summed the precipitation across each subcatchment and routed the upstream accumulated precipitation along the hydrography to mimic run-off (Domisch, Amatulli, & Jetz, 2015).

From a large set of layers per case study, we selected those that are meaningful from an ecological perspective to describe the distribution of the species. The decision regarding the choice of predictors was based on a mixture of expert knowledge, data availability and on preliminary model runs: For each realm, we did multiple preliminary model runs where we tested which set of predictors yields the highest model evaluation scores. Hence, we aimed to get the best-possible model predictions for each realm. The set of predictors among realms is different as each realm needs to take realm-specific characteristics into account. Highly intercorrelated environmental predictors were omitted (i.e., where Pearson's correlation tests between different predictors yielded a correlation coefficient >0.7 or <−0.7) (Dormann et al., 2013, see Supporting Information Figure S1 for all predictors used). All (continuous) predictors were centred (so all predictors have a mean of 0) and scaled by dividing by their standard deviations. The final number of predictors was 13, 11 and 8 for SDMs in the terrestrial, marine and freshwater realms, respectively.

2.3 | Defining spatial connectivity

Regarding spatial SDMs, the connectivity among planning units for the terrestrial and marine realms was computed from a polygon shapefile using the R-package SPDEP (using the queen's move; Bivand, Hauke, & Kossowski, 2013; Bivand & Piras, 2015). This procedure identifies all first-order neighbours of each focal planning unit, that is, those planning units that are directly adjacent and connected. In the freshwater realm, it is crucial to account for the longitudinal connectivity among planning units (Abell, Allan, & Lehner, 2007; Hermoso, Linke, Prenda, & Possingham, 2011). Hence, the spatial connectivity was defined as all upstream subcatchments connected within a 100 km (as-the-fish-swims) distance of a given subcatchment. This distance was chosen as a trade-off between hydrological

connectivity (which can exceed 100 km) and the computational demand and time for calculating SDMs and spatial plans. Note that with this information, full connectivity matrices (range-wide and up/downstream) are built within SDMs and the spatial prioritization. In other words, each planning unit "knows" which planning units are connected to it.

2.4 | Detecting spatial autocorrelation

We first assessed the degree of spatial autocorrelation in species occurrences among planning units (where occurrences were aggregated within planning units). Here, we used the Monte Carlo simulation method and calculated Moran's I and the associated p-value for each species (at $\alpha = 0.05$), using the R-package SPDEP (with 999 permutations). Spatial weights were kept identical to the connectivity definition of planning unit above, specifying proportional weights to the connected planning unit (row standardization).

2.5 | Species distribution modelling

We ran Bayesian occupancy models for predicting the habitat suitability of species in each realm. Our aim was to minimize the methodological shortcomings in SDMs that could be ported to the spatial prioritization. Hence, we refrained from artificial pseudo-absences to draw species non-detections, but used survey data with repeated visits to account for the detection probability of species at given sites. Specifically, we ran zero-inflated binomial SDMs in a Bayesian framework using the HSDM R-package (Vieilledent et al., 2014) with "hSDM.ZIB" and "hSDM.ZIB.iCAR" functions for the non-spatial and spatial SDMs, respectively. These functions integrate two processes, a Bernoulli suitability and a Binomial observability process, into a hierarchical zero-inflated binomial model. The Bernoulli suitability process uses species point records and environmental predictors as the response and explanatory variables, respectively (Vieilledent et al., 2014):

$$z_i \sim \text{Bernoulli}(\theta_i) \tag{1}$$

$$\text{logit}(\theta_i) = X_i\beta \tag{2}$$

where z_i is a random variable describing the binary habitat suitability at planning unit$_i$, which follows a Bernoulli distribution of parameter θ_i; that is, the probability that the habitat is suitable in the planning unit$_i$. θ_i is expressed as a linear model combining the matrix of environmental predictors X_i and parameters β using a logit link function.

In the spatial models, Equation 2 is extended by an intrinsic conditional autoregressive model (iCAR) in the suitability process:

$$\text{logit}(\theta_i) = X_i\beta + \rho_i \tag{3}$$

where ρ_i is the spatial random effect of the planning unit$_i$. ρ_i accounts for the spatial autocorrelation of the presence probabilities

variability in suitability that is not explained by the environmental variables:

$$\rho_i \sim \text{Normal}(\mu_\rho V_\rho / n_i) \tag{4}$$

where μ_i is the mean of ρ_i in the neighbourhood of i, V_ρ is the variance of the spatial random effects, and n_i is the number of neighbours for the planning unit$_i$ (see also Latimer et al., 2006, for the formal description of the iCAR). Thus, we assume that in addition to the environmental characteristics within the planning units, the species' occurrence probabilities also depend on the occurrence probabilities of the neighbouring planning unit (Vieilledent et al., 2014). A planning unit with "poor" environmental conditions next to planning units with "good" environmental conditions will result in higher estimated suitability, than if its neighbours also show "poor" environmental conditions (Domisch et al., 2016).

Regarding the observability process, we counted the number of repeat surveys within each planning unit; that is, we aggregated the surveys over time, yielding the information on how often each planning unit was visited. The model then estimates the probability of observation given the species presence in a planning unit, where we assume that, if the species was observed at least once during multiple visits in a given planning unit, the habitat is suitable and the absence of the species during other visits in the planning unit is due to imperfect detection. For instance, a given species was observed once in planning units A and B. Planning unit A was visited once, yielding a detection probability of 1, whereas B was visited 10 times (yielding a detection probability of 1/10 = 0.1). In combination with the suitability process (i.e., environmental conditions) and the spatial autocorrelation (i.e., the spatial neighbourhood information), the observability information yields more robust estimates in unsampled locations, that is, whether a species actually occurs there but was not detected, or if it is actually not present (and hence not detected). The probability of observing the species (δ_ρ) was specified as:

$$y_i \sim \text{Binomial}(z_i \times \delta_i t_i) \tag{5}$$

$$\text{logit}(\delta_i) = \gamma_0 \tag{6}$$

where y_i is a vector of the total number of observed presences in planning unit$_i$. z_i is the binary habitat suitability in planning unit$_i$ from the suitability process (Equation 1), and t_i denotes the total number of visits within the planning unit$_i$, including the non-detections as described above. y_i follows a binomial distribution of the combination of the habitat suitability z_i at planning unit$_i$ and the parameters δ_i and t_i. In other words, δ_i is the probability of observing a species in a location, if it was actually present. We estimate only the intercept (γ_0) using a logit link function and assume a spatially constant detection probability (due to the relatively small occurrence datasets and limited spatial replication needed to understand the role of covariates in variable detection probabilities).

We ran three Markov chain Monte Carlo (MCMC) simulations with 200,000 iterations each, a burn-in phase of 50,000 iterations and a thinning interval of 10. Model convergence was assessed by the multivariate potential scale reduction factor (MPSRF; Brooks & Gelman, 1998). For the suitability process, we used the coefficients from an initial, non-spatial generalized linear model (GLM) as initial values, and both suitability and observability processes used uninformative priors centred at zero with a fixed large variance of 100 (Domisch et al., 2016). The prior distribution for the variance of the spatial random effects followed a uniform distribution, that is, a flat prior where the upper bound of the variance is set to 10.

Species data were split into training (70%) and validation (30%) sets, and model performance for each species was evaluated using the area under curve (AUC), true skill statistic (TSS), sensitivity and specificity (true positive and negative predictions, respectively) and the deviance information criterion (DIC). Note that the non-spatial SDMs did not have any pre-defined spatial configuration during the SDM calibration and prediction; that is, each planning unit was considered independent of each other in the SDMs, while all other data were kept constant (i.e., identical random seed in the SDMs, and identical predictors and species data including the subsets for validation). The spatial prioritization was undertaken using the realm-specific spatial connectivity as described above, for predictions derived from both non-spatial and spatial SDMs.

2.6 | Spatial prioritization

We transformed the predictive posterior mean probability maps from SDMs into a semi-binary scheme using TSS as a threshold (Allouche, Tsoar, & Kadmon, 2006), where all values below the threshold were converted to zero and values above the threshold retained their original values. This overcomes the problem of inflating the spatial prioritization with many planning units having low probabilities of occurrence (e.g., 10 planning units with probabilistic values of 0.1 would equal one planning unit having a value of 1). Simultaneously, this procedure retained the information of varying probabilities of occurrence (as recommended by Tulloch et al., 2016) above a certain level of confidence as given by TSS (i.e., within the range of the predicted "presence" of a species). We refrained from locking-in any current, already established protected areas as this would have decreased the flexibility in selecting protected areas given only the conservation features, which was the aim of this study.

To create the boundary files (describing the spatial connectivity) for the terrestrial and marine case studies, we used the QMarxan toolbox in QGIS (QGIS-Development-Team, 2017). For the freshwater realm, we applied an inverse-distance connectivity penalty, where subcatchments located closer to a given focal planning unit would get a higher penalty if not chosen as part of the protected area (opposed to those planning units located more distant, as proposed by Hermoso et al. 2011).

We first calibrated the boundary length modifier (BLM) for our analyses. The BLM is dimensionless and balances the spatial aggregation and patchiness of spatial plans. For the BLM calibration, we used the software MARXAN v.2.43v (Ball et al., 2009) within the R-package MARXAN (Hanson & Watts, 2015). We created all necessary input files in base R, except the "input.dat" file, where for convenience, we used the MARXAN R-package. We then calibrated the BLM following the recommendations by Ardron, Possingham, and Klein (2008). First, we ran MARXAN with a fixed species penalty factor of 10, and BLM values set to 0, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1,000, 10,000, 100,000 to approximate the range of the optimal BLM value. We plotted the connectivity against the BLM (Supporting Information Figure S6) and identified the elbow, that is, the point from where an increase in the compactness (higher BLM values) has no major effect on the connectivity in spatial plans anymore. Similarly, we also mapped the spatial plans to visually confirm the increasing compactness derived from increasing BLM values. In a second step, we repeated the previous analyses and maps with setting the BLM ranging within the approximate BLM value (e.g., between 10 and 100). Again, we ran sensitivity runs within increments of 10 and identified the BLM to 35, 20, 0.15 for the terrestrial, marine and freshwater realms, respectively (Supporting Information Figure S6).

We then used the GUROBI OPTIMIZER 7.5 software (Gurobi Optimization, 2017) to find optimal conservation planning solutions based on integer linear programming (ILP) within the R-package PRIORITIZR (Hanson et al., 2017). ILP has shown to out-compete traditional simulated annealing tools (e.g., Marxan, Ball et al., 2009) in terms of time and accuracy (Beyer, Dujardin, Watts, & Possingham, 2016), using the same input files as MARXAN. We varied the conservation target for all species in the spatial plans, ranging from 10% to 50% in 5% increments. For instance, a conservation target of 20% means that for each species 20% of its suitable habitat (as given by the SDM predictions) are required in this specific spatial plan.

For each run, we allowed a 10% gap to optimality in the spatial plans (as trade-off between optimality, and the time the optimizer takes to converge). To rule out the effect of the planning unit itself (cost, i.e., area of the planning unit), we set a constant value of 1 as the cost of each planning unit. Likewise, we set no weights for single species. This means that only the conservation features, that is, the probabilistic information on species habitat suitability and the connectivity penalty (set by the BLM), were decisive to the objective function that was set to minimize the total number of planning units to be part of the selected spatial plan (Game & Grantham, 2008). Note that opposed to, for example, MARXAN, GUROBI provides one optimal solution and hence requires that all species meet the conservation target.

2.7 | Statistical analyses

For each realm and model type, we compared model accuracy given the model evaluation scores, the estimated detection probability, the range size estimates of model outputs considering all planning units that had a probability value above the threshold (equal to binary predictions), and the summed probability of habitat suitability across all species within each planning unit as a

proxy for species richness (Mateo, Felicísimo, Pottier, Guisan, & Muñoz, 2012). For each conservation target run (10%–50%) derived from spatial and non-spatial SDMs, we (a) assessed whether species were covered by the spatial plans, (b) counted the relative number of planning units needed to fulfil the given target of the spatial plan and (c) compared the spatial overlap of spatial plans by assessing the percentage of overlapping spatial units.

3 | RESULTS

3.1 | Spatial autocorrelation

Across the terrestrial, marine and freshwater realms, all but three of the 171 modelled species showed a significantly positive spatial autocorrelation ($\alpha > 0.05$ for the freshwater fish species White-eye bream [*Ballerus sapa*], sunbleak [*Leucaspius delineates*] and brown trout [*Salmo trutta lacustris*], see Supporting Information Table S2). Species occurrences across planning units were hence non-randomly distributed, warranting the use of spatial random effects within SDMs to account for spatial autocorrelation.

3.2 | Model performance

The SDM performance indicators AUC, TSS, sensitivity and specificity were on average consistently higher for the spatial SDMs than for

the counter-parts (Figure 2). DIC was on average lower for the spatial SDMs in the terrestrial and freshwater realm, but higher in the marine realm (marine non-spatial vs. spatial SDMs; DIC: 515 ± 485 vs. 534 ± 596, mean ± SD). All chains of the SDMs converged successfully yielding a MPSRF value of 0.999. The detection probability across realms was on average consistently higher in the spatial than in the non-spatial SDMs (Figure 2f,l,r).

3.3 | Range size estimates

Spatial SDMs yielded generally more compact and less dispersed habitat suitability estimates (see exemplary maps in Supporting Information Figure S1), and this effect is also mirrored in the range size estimates. Range size estimates derived from spatial SDMs were significantly lower in the freshwater and terrestrial realms than in the non-spatial models. This means that the predictions derived from spatial SDMs were more restrictive and species' suitable habitats were predicted to occur in less planning units than in the non-spatial SDMs (see exemplary maps in Supporting Information Figure S1; non-spatial vs. spatial SDMs in the terrestrial realm: 1,278 ± 750 vs. 1,057 ± 690 planning units, paired *t*-test: *t* = 2.95, *df* = 32, *p* = 0.006; freshwater: 2,874 ± 1,491 and 2,532 ± 1,409 planning units, mean ± SD, paired *t*-test: *t* = 2.0493, *df* = 84, *p* = 0.044). In the marine realm, no significant difference in range size estimates was observed (non-spatial vs. spatial SDMs: 2,698 ± 1,564 and 2,628 ± 1,500 planning units, paired *t*-test: *t* = 0.783, *df* = 52, *p* = 0.436).
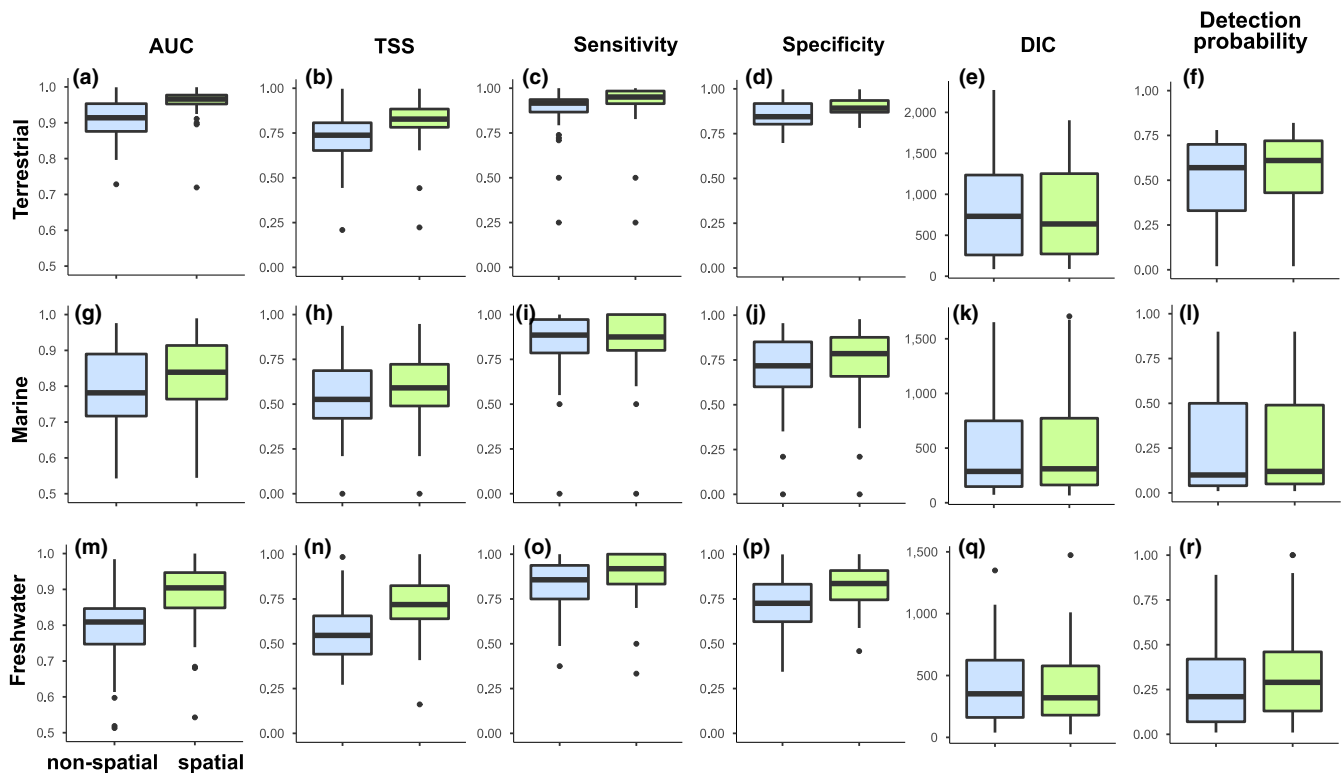

FIGURE 2 Model evaluation scores representing AUC, TSS, sensitivity, specificity and DIC, as well as the estimated detection probability, summarized across 33 terrestrial (a–f), 53 marine (g–l) and 85 freshwater species (m–r) derived from non-spatial (blue) and spatial SDMs (green). Bars represent median values and boxes the 1st and 3rd quartiles, respectively
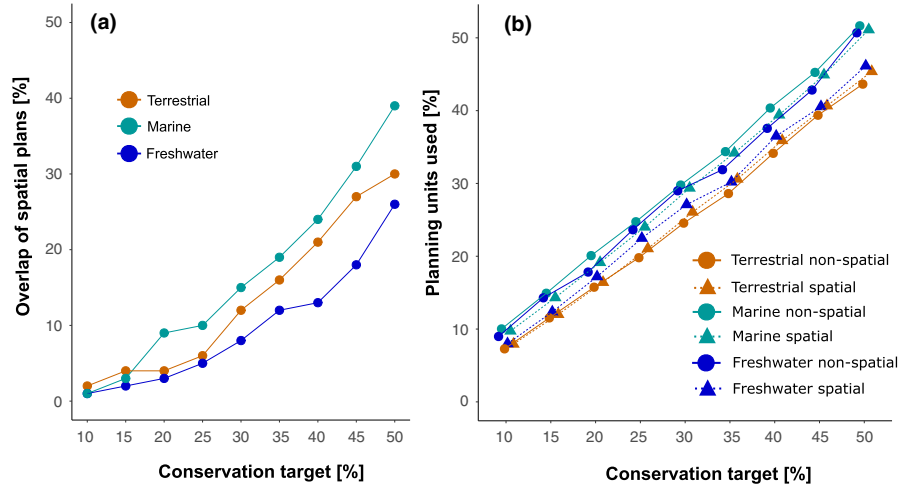
FIGURE 3 (a) Spatial overlap in potential protected areas based on spatial and non-spatial SDMs, along a conservation target of 10%–50% in the terrestrial, marine and freshwater realm. Please see Supporting Information Figures S3–S5 for the mapped spatial plans of each conservation target. (b) Relative number of planning units required for reaching the solution in spatial plans for the different realms with conservation targets ranging from 10%–50% in the terrestrial, marine and freshwater realms. Points and triangles along solid and dotted lines represent spatial plans derived from non-spatial and spatial SDMs, respectively

3.4 | Summed probabilities of habitat suitability

Across realms, spatial SDMs produced on average a lower summed habitat suitability across planning units (i.e., a proxy for species richness), than predicted by non-spatial SDMs (non-spatial vs. spatial SDMs in the terrestrial realm: 3.00 ± 1.28 vs. 2.37 ± 1.00 planning units, paired t-test: $t = 61.845$, $df = 7,762$, $p < 0.001$; marine: 12.27 ± 3.01 vs. 11.79 ± 2.97 planning units, mean ± SD, paired t-test: $t = 28.458$, $df = 7,122$, $p < 0.001$; freshwater: 21.36 ± 11.91 vs. 16.50 ± 9.48 planning units, paired t-test: $t = 46.057$, $df = 7,375$, $p < 0.001$). See Supporting Information Figure S2 for a spatial representation of the differences in summed probabilities among non-spatial and spatial SDMs.

3.5 | Spatial similarity of spatial plans

All species targets across SDM types and spatial conservation plans were met in all spatial prioritization runs (data shown in the Pangaea repository). The degree of spatial overlap of potential protected areas derived from spatial and non-spatial SDMs was target-dependent. The overlap was lowest for the smallest conservation target (10%) with the per cent overlap ranging from 1% to 2%, and increased up to a maximum of 30% to 39% for a conservation target of 50% (Figure 3a, and Supporting Information Figures S3–S5). The number of required planning units increased linearly with increasing conservation targets: the higher the conservation target, the higher the required amount of planning units to meet the given target (Figure 3a).

The relative difference in the number of planning units needed for a given solution between spatial and non-spatial SDMs was within a margin of 5% (Figure 3b). No significant differences in the number of planning units between spatial and non-spatial models across conservation targets could be observed.

4 | DISCUSSION

Incorporating connectivity has been successfully adopted in systematic conservation planning for building species migration and movement corridors (Margules & Pressey, 2000). Yet, this concept has been largely neglected in the underlying conservation features (e.g., species) that are used in the spatial planning. Our study provides a non-exhaustive comparison, shedding light on the effects of non-spatial versus spatially explicit conservation features in systematic conservation planning. We show that (a) spatially explicit predictions of species' probabilistic habitat suitability outperform those derived from non-spatial SDMs across the three realms, each with a specific landscape configuration and consequently distinct species dispersal and connectivity characteristics. Moreover, (b) spatially explicit predictions provide lower range size estimates, indicating less dispersed habitat suitability predictions. This effect cascades into the conservation planning as (c) the spatial arrangement of spatial conservation plans differs fundamentally between those derived from non-spatial and spatial model predictions. Simultaneously, (d) both spatial and non-spatial approaches require a similar number of planning units for delineating potential protected areas.

Spatial SDMs yielded habitat suitability predictions that were on average more accurate given their better model performance (Figure 2), than those derived from non-spatial SDMs. This pattern is in line with previous findings from Record et al. (2013) due to three well-known factors (reviewed by Dormann, 2007): (a) distance-related dispersal or species interactions, (b) non-linear species–environment relationships and (c) spatially structured environmental predictors causing spatial structuring in the response, that is, the modelled species' habitat suitability (Besag, 1974). In addition, Crase, Liedloff, Vesk, Fukuda, and Wintle (2014) showed that if such induced spatial autocorrelation in species distributions is not accounted for, model predictions are likely to remain inaccurate. Contrary to Thibaud, Petitpierre,

Broennimann, Davison, and Guisan (2014), who found that spatial autocorrelation itself has only a minor relative effect on model prediction accuracy (compared to the choice of predictors, sample sizes, sampling designs and modelling technique), our analyses reveal that accounting for spatial autocorrelation leads to a substantial improvement of model outputs (given that all other data were kept constant). This effect cascades into the resulting spatial plans: Those derived from spatial SDM predictions are likely to provide a better representation of areas needed to potentially protect, than spatial conservation plans derived from non-spatial SDM predictions. In addition, spatial SDMs yielded, on average, a slightly higher detection probability and thus reduced the potential false absences caused by the imperfect detection of species (MacKenzie et al., 2017). Reducing the risk of overlooking species is therefore crucial for an effective conservation planning (Petitot, Manceau, Geniez, & Besnard, 2014).

We found only a minor overlap of potential protected areas based on spatial and non-spatial SDMs for the commonly applied conservation targets of 10%–20% (e.g., the global target of 17% set by the Convention on Biological Diversity, CBD, 2010; Veach, Minin, Pouzols, & Moilanen, 2017). Interestingly, spatial plans, regardless of the modelling approach, required a similar amount of planning units. On average, spatial SDMs predicted a lower range size per species and, therefore, a lower number of species predicted to have a suitable habitat per planning unit. Hence, spatial plans using predictions from spatial SDMs require—relatively speaking—more planning units (e.g., assuming that non-spatial and spatial SDMs would predict a suitable habitat of a species into 100 and 90 planning units, respectively, the spatial plans should cover theoretically 20 and 18 planning units given a target of 20%). This is likely to derive due to the, on average, lower summed probability of suitable habitats per planning unit in spatial SDM predictions (i.e., species richness sensu Mateo et al., 2012). Consequently, the optimizer (GUROBI, in our case) had to choose from a smaller pool of available planning units. Hence, the optimizer is forced to seek additional planning units that lead to a remarkably different spatial arrangement of spatial conservation plans (Figures 1 and 3a). While we did not explicitly test for differences among realms, this pattern was similar across different landscape configurations and planning unit shapes. Broadening the perspective, non-spatial SDMs tend to create artificial species "hot spots" by predicting a higher species richness per planning unit, and the derived spatial plans would take advantage of the umbrella effect, that is, where the conservation of a species confers protection to a larger number of co-occurring species (Roberge & Angelstam, 2004). In our study, however, this pattern does not depict the best-possible outcome given the lower model evaluation scores. In turn, spatially explicit SDMs have the potential, though not reducing the quantity of protected areas, to increase their quality (Barnes et al., 2018).

We acknowledge that a 10% gap to optimality, as we used it here, can introduce noise in the spatial arrangement of potential protected areas (e.g., Figure 3a). Even with this noise, however, differences in spatial plans were distinct in our study. While a more stringent gap (e.g., 0.5%, Beyer et al., 2016) could yield more accurate spatial plans, we intended to balance uncertainties in the spatial plans with those most probably apparent in the underlying data (note that due to

computational reasons, we only extracted the mean posterior probability of SDMs). In other words, species and environmental data, as well as the aggregation of data into planning units, potentially lead to a certain level of uncertainty in the habitat suitability predictions in our study (Beale & Lennon, 2012). Hence, spatial plans should be given flexibility to accommodate such uncertainties in terms of the selection of spatial units during the spatial prioritization.

We singled out the effect of spatial autocorrelation in conservation features to explore the effect of the connectivity among planning units on the subsequent spatial conservation planning. We encourage future studies to further explore these effects using "real-world" exercises, where the current and already established protected areas could be locked-in during the spatial prioritization along with various costs for the planning units. This would decrease the flexibility in the selection of protected areas, and given the low overlap between non-spatial and spatial conservation features, the impact on the outcome remains to be seen. Similarly, the degree of fragmentation (Cabeza, 2003) and the minimum patch size depend on the available area and financial resources (as a large amount of small patches are more expensive to manage (Smith, Minin, Linke, Segan, & Possingham, 2010)). Likewise, the implementation of various cost measures, such as area, land price or human influence, could be added in the analyses—a topic that was however beyond the scope of our study. Hence, testing spatial plans and with these additions and interactions has the potential to further reveal differences stemming from non-spatial and spatially explicit modelled conservation features. In our study, the species data were contingent on publicly available survey data across the three realms, and hence, the species can be considered generalist species. Rare species are considered particularly important for spatial planning; however, such data were not publicly available for our study. It remains therefore to be seen whether SDMs of rare species would yield similar differences in spatial plans as shown in our study.

We note that by accounting for species' observability by employing a hierarchical SDM, we aimed to limit the shortcomings of SDMs that else do not account for the fact that species might be overlooked during the sampling. Regarding the addition of the spatial information in SDMs, we want to highlight that instead of running hierarchical Bayesian models, other SDM techniques could be equally used, such as Generalized Linear Mixed Models using spatial random effects, or other SDMs algorithms that use spatial eigenvectors as a covariate containing information about the spatial structure in the study area (see Dormann, 2007, for a review on various methods).

5 | CONCLUSIONS

Employing SDMs that account for spatial dependencies in conservation features provide a promising way forward to increase the quality of protected areas, however without increasing the area (and possibly costs) needed. They yield fundamentally different spatial conservation plans given more accurate species habitat suitability predictions, compared to SDMs that ignore such spatial dependencies. This highlights the importance of using best-possible modelling practices sensu

Dormann (2007) and Guisan et al. (2013), which is needed for a wider acceptance of SDMs in systematic conservation planning (Tulloch et al., 2016). Hence, we encourage modellers and practitioners to carefully assess the possibilities in adding spatial SDMs to their systematic conservation planning to derive sufficient options for choosing the optimal spatial arrangement of protected areas.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

DATA ACCESSIBILITY

All results supporting the study are available at https://doi.pangaea.de/10.1594/PANGAEA.889033 (all model predictions, model evaluation scores and spatial conservation plans) for each realm. All environmental data used for modelling are available in the following repositories for download: www.worldclim.org, www.earthenv.org, http://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev01, www.bio-oracle.org, www.marspec.org, and Supporting Information of Fithian et al. (2015). Species data are available for download in the following repositories: Supporting Information of Fithian et al. (2015), http://sero.nmfs.noaa.gov/sustainable_fisheries/caribbean/fish_indep_wkshp/surveys/sefis/, and www.freshwaterplatform.eu.

ORCID

Sami Domisch (iD) https://orcid.org/0000-0002-8127-9335

Martin Friedrichs (iD) https://orcid.org/0000-0003-0644-7869

Thomas Hein (iD) https://orcid.org/0000-0002-7767-4607

Florian Borgwardt (iD) https://orcid.org/0000-0002-8974-7834

Sonja C. Jähnig (iD) https://orcid.org/0000-0002-6349-9561

Simone D. Langhans (iD) https://orcid.org/0000-0001-9581-3183

REFERENCES

Abell, R., Allan, J. D., & Lehner, B. (2007). Unlocking the potential of protected areas for freshwaters. *Biological Conservation*, *134*, 48–63. https://doi.org/10.1016/j.biocon.2006.08.017

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Amatulli, G., Domisch, S., Tuanmu, M. N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*, 180040. https://doi.org/10.1038/sdata.2018.40

Araujo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology*, *11*, 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, *93*, 1527–1539. https://doi.org/10.1890/11-1930.1

Ardron, J. A., Possingham, H. P., & Klein, C. J. (2008). *Marxan good practices handbook*. Vancouver, BC: Pacific Marine Analysis and Research Association.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*, 277–284. https://doi.org/10.1111/geb.12693

Bacheler, N. M., Berrane, D. J., Mitchell, W. A., Schobernd, C. M., Schobernd, Z. H., Teer, B. Z., & Ballenger, J. C. (2014). Environmental conditions and habitat characteristics influence trap and video detection probabilities for reef fish species. *Marine Ecology Progress Series*, *517*, 1–14. https://doi.org/10.3354/meps11094

Ball, I. R., Possingham, H. P., & Watts, M. (2009). Marxan and relatives: Software for spatial conservation prioritisation. In A. Moilanen, K. A. Wilson & H. P. Possingham (Eds.), *Spatial conservation prioritisation: Quantitative methods and computational tools* (pp. 185–195). Oxford: Oxford University Press.

Barnes, M., Glew, L., Wyborn, C., & Craigie, I. D. (2018). Prevent perverse outcomes from global protected area policy. *Nature Ecology and Evolution*, *2*, 759–762. https://doi.org/10.1038/s41559-018-0501-y

Beale, C. M., & Lennon, J. J. (2012). Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 247–258. https://doi.org/10.1098/rstb.2011.0178

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, *36*, 192–236. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Beyer, H. L., Dujardin, Y., Watts, M. E., & Possingham, H. P. (2016). Solving conservation planning problems with integer linear programming. *Ecological Modelling*, *328*, 14–22. https://doi.org/10.1016/j.ecolmodel.2016.02.005

Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Bastos, R. P., & Pinto, M. P. (2006). Challenging wallacean and linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, *12*, 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x

Bivand, R., Hauke, J., & Kossowski, T. (2013). Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, *45*, 150–179.

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). New York, NY: Springer.

Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of the American Statistical Association*, *63*, 1–36. https://doi.org/10.18637/jss.v063.i18

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Cabeza, M. (2003). Habitat loss and connectivity of reserve networks in probability approaches to reserve design. *Ecology Letters*, *6*, 665–672. https://doi.org/10.1046/j.1461-0248.2003.00475.x

Carr, M. H., Neigel, J. E., Estes, J. A., Andelman, S., Warner, R. R., & Largier, J. L. (2003). Comparing marine and terrestrial ecosystems: Implications for the design of coastal marine reserves. *Ecological Applications*, *13*, S90–S107. https://doi.org/10.1890/1051-0761(2003)013[0090:CMATEI]2.0.CO;2

CBD (2010). Report of the tenth meeting of the conference of the parties to the convention on biological diversity. In: The tenth meeting of the Conference of the Parties to the Convention on Biological Diversity, Nagoya, Japan, NEP/CBD/COP/10/27. Retrieved from https://www.cbd.int/cop10/doc.

Crase, B., Liedloff, A., Vesk, P. A., Fukuda, Y., & Wintle, B. A. (2014). Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. *Global Change Biology*, *20*, 2566–2579. https://doi.org/10.1111/gcb.12598

Daigle, R. M., Metaxas, A., Balbar, A., McGowan, J., Treml, E. A., Kuempel, C. D., … Beger, M. (2018). Operationalizing ecological connectivity in spatial conservation planning with Marxan Connect. *Biorxiv*. https://doi.org/10.1101/315424

De Marco, P., Diniz-Filho, J. A. F., & Bini, L. M. (2008). Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, *4*, 577–580. https://doi.org/10.1098/rsbl.2008.0210

Domisch, S., Amatulli, G., & Jetz, W. (2015). Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scientific Data*, *2*, 150073.

Domisch, S., Wilson, A. M., & Jetz, W. (2016). Model-based integration of observed and expertbased information for assessing the geographic and environmental distribution of freshwater species. *Ecography*, *39*, 1078–1088. https://doi.org/10.1111/ecog.01925

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, *16*, 129–138. https://doi.org/10.1111/j.1466-8238.2006.00279.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Leitão, P. J., (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Esselman, P. C., & Allan, J. (2011). Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. *Freshwater Biology*, *56*, 71–88. https://doi.org/10.1111/j.1365-2427.2010.02417.x

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, *6*, 424–438. https://doi.org/10.1111/2041-210X.12242

Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, *27*, 245–256. https://doi.org/10.1111/geb.12684

Game, E., & Grantham, H. (2008). *Marxan user manual: For Marxan version 1.8.10*. St. Lucia, Qld, Australia, Vancouver, BC, Canada: University of Queensland, Pacific Marine Analysis and Research Association.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. https://doi.org/10.1111/ele.12189

Gurobi Optimization, I. (2017). *Gurobi optimizer reference manual*. Retrieved from http://www.gurobi.com.

Hampe, A. (2004). Bioclimate envelope models: What they detect and what they hide. *Global Ecology and Biogeography*, *13*, 469–471. https://doi.org/10.1111/j.1466-822X.2004.00090.x

Hanson, J., & Watts, M. (2015). *Marxan: decision support tools for reserve selection in R using Marxan*. R package version 1.0. 1.

Hanson, J., Schuster, R., Morrell, N., Strimas-Mackey, M., Watts, M., Arcese, P., …Possingham, H. P. (2017). prioritizr: Systematic conservation prioritization in R. R package version 3.0.1.1.

Hermoso, V., Filipe, A. F., Segurado, P., & Beja, P. (2018). Freshwater conservation in a fragmented world: Dealing with barriers in a systematic planning framework. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *28*, 17–25. https://doi.org/10.1002/aqc.2826

Hermoso, V., Linke, S., Prenda, J., & Possingham, H. (2011). Addressing longitudinal connectivity in the systematic conservation planning of fresh waters. *Freshwater Biology*, *56*, 57–70. https://doi.org/10.1111/j.1365-2427.2009.02390.x

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., & van Etten, J. (2018). *raster: Geographic data analysis and modeling*. R package version, 2.

Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., … Possingham, H. (2009). Incorporating ecological and evolutionary processes into continental scale conservation planning. *Ecological Applications*, *19*, 206–217. https://doi.org/10.1890/07-1684.1

Latimer, A. M., Wu, S., Gelfand, A. E., & Silander, J. A. (2006). Building statistical models to analyze species distributions. *Ecological Applications*, *16*, 33–50. https://doi.org/10.1890/04-0609

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*, 1659–1673.

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*, 2171–2186.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., … Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*, 494–502. https://doi.org/10.1890/100125

Lehtomäki, J., & Moilanen, A. (2013). Methods and workflow for spatial conservation prioritization using Zonation. *Environmental Modelling and Software*, *47*, 128–137. https://doi.org/10.1016/j.envsoft.2013.05.001

Lockwood, M. (2010). Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *Journal of Environmental Management*, *91*, 754–766. https://doi.org/10.1016/j.jenvman.2009.10.005

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Academic Press: Elsevier.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*, 243–253. https://doi.org/10.1038/35012251

Mateo, R. G., Felicísimo, Á. M., Pottier, J., Guisan, A., & Muñoz, J. (2012). Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, *7*, e32586.

McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the anthropocene. *Trends in Ecology and Evolution*, *30*, 104–113. https://doi.org/10.1016/j.tree.2014.11.006

McGowan, J., Hines, E., Elliott, M., Howar, J., Dransfield, A., Nur, N., & Jahncke, J. (2013). Using seabird habitat modeling to inform marine spatial planning in central California's National Marine Sanctuaries. *PLoS ONE*, *8*, e71406. https://doi.org/10.1371/journal.pone.0071406

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, *19*, 992–1006. https://doi.org/10.1111/ele.12624

Nhancale, B. A., & Smith, R. J. (2011). The influence of planning unit characteristics on the efficiency and spatial pattern of systematic conservation planning assessments. *Biodiversity and Conservation*, *20*, 1821–1835. https://doi.org/10.1007/s10531-011-0063-7

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography*, *12*, 361–371.

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*, 9–13.

Petitot, M., Manceau, N., Geniez, P., & Besnard, A. (2014). Optimizing occupancy surveys by maximizing detection probability: Application to amphibian monitoring in the mediterranean region. *Ecology and Evolution*, *4*, 3538–3549. https://doi.org/10.1002/ece3.1207

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752. https://doi.org/10.1126/science.1246752

Pressey, R. L., Cabeza, M., Watts, M. E., Cowling, R. M., & Wilson, K. A. (2007). Conservation planning in a changing world. *Trends in Ecology and Evolution*, *22*, 583–592. https://doi.org/10.1016/j.tree.2007.10.001

QGIS-Development-Team. (2017). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from https://www.qgis.org.

R-Core-Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Record, S., Fitzpatrick, M. C., Finley, A. O., Veloz, S., & Ellison, A. M. (2013). Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. *Global Ecology and Biogeography*, *22*, 760–771. https://doi.org/10.1111/geb.12017

Roberge, J. M., & Angelstam, P. (2004). Usefulness of the umbrella species concept as a conservation tool. *Conservation Biology*, *18*, 76–85. https://doi.org/10.1111/j.1523-1739.2004.00450.x

Rondinini, C., Di Marco, M., Chiozza, F., Santulli, G., Baisero, D., Visconti, P., ... Boitani, L. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 2633–2641. https://doi.org/10.1098/rstb.2011.0113

Sbrocco, E. J., & Barber, P. H. (2013). Marspec: Ocean climate layers for marine spatial ecology. *Ecology*, *94*, 979–979. https://doi.org/10.1890/12-1358.1

Schinegger, R., Pletterbauer, F., Melcher, A., & Schmutz, S. (2016). Metadata describing the European Fish Index Plus (EFI+) database. *Freshwater Metadata Journal*, *17*, 1–12. https://doi.org/10.15504/fmj.2016.17

SEAMAP-SA. (2017). SEAMAP-SA Data Management Work Group. SEAMAP-SA online database. Retrieved from http://www.dnr.sc.gov/SEAMAP/data.html.

Smith, R. J., Di Minin, E., Linke, S., Segan, D. B., & Possingham, H. P. (2010). An approach for ensuring minimum protected area size in systematic conservation planning. *Biological Conservation*, *143*, 2525–2531. https://doi.org/10.1016/j.biocon.2010.06.019

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A. C., & Guisan, A. (2014). Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, *5*, 947–955. https://doi.org/10.1111/2041-210X.12203

Tuanmu, M. N., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, *23*, 1031–1045. https://doi.org/10.1111/geb.12182

Tulloch, A. I., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K. M. P., ... Rhodes, J. R. (2016). Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. *Biological Conservation*, *199*, 157–171. https://doi.org/10.1016/j.biocon.2016.04.023

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-oracle: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*, 272–281. https://doi.org/10.1111/j.1466-8238.2011.00656.x

van Teeffelen, A. J., Cabeza, M., & Moilanen, A. (2006). Connectivity, probabilities and persistence: Comparing reserve selection strategies. *Biodiversity & Conservation*, *15*, 899–919. https://doi.org/10.1007/s10531-004-2933-8

Veach, V., Di Minin, E., Pouzols, F. M., & Moilanen, A. (2017). Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions*, *23*, 715–726. https://doi.org/10.1111/ddi.12571

Vieilledent, G., Latimer, A., Gelfand, A., Merow, C., Wilson, A., Mortier, F., & Silander, J. Jr (2014). hsdm: Hierarchical bayesian species distribution models. R package version, 1.

Weeks, R. (2017). Incorporating seascape connectivity in conservation prioritisation. *PLoS ONE*, *12*, e0182396. https://doi.org/10.1371/journal.pone.0182396

White, J. W., Schroeger, J., Drake, P. T., & Edwards, C. A. (2014). The value of larval connectivity information in the static optimization of marine reserve design. *Conservation Letters*, *7*, 533–544. https://doi.org/10.1111/conl.12097

Zupancic, P. (2015). BioFresh (data publisher) and Primoz Zupancic (data provider). Distribution data of Western Balkan freshwater fish mobilised through BioFresh. Retrieved from http://data.freshwaterbiodiversity.eu.

BIOSKETCH

Sami Domisch is an aquatic ecologist who is interested in freshwater biodiversity and biogeography. His research is based on spatial freshwater biodiversity science and biodiversity informatics, and how these tools can be further developed to yield robust estimates regarding freshwater biodiversity and conservation planning. The authors consist of a multidisciplinary team with a wide range of research interests, such as freshwater biodiversity, spatial modelling, river restoration, floodplain ecology and systematic conservation planning.

Author contributions: SD, MF, SCJ and SDL designed the study; SD performed the modelling and spatial prioritization; and SD, MF and SDL analysed output data. TH, FP and AW collected and prepared data. SD, MF and SDL wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.