Received: 3 February 2025 | Accepted: 10 August 2025
DOI: 10.1111/1365-2664.70177

BRITISH ECOLOGICAL SOCIETY Journal of Applied Ecology

RESEARCH ARTICLE

Leveraging long-term data to improve biodiversity monitoring with species distribution models

Nicolò Anselmetto[1,2] | Matteo Garbarino[1] | Matthew J. Weldy[3,4] | David M. Bell[3] | Christopher Daly[5] | Clinton W. Epps[4] | Nina Ferrari[2] | Hankyu Kim[6] | Joseph A. LaManna[7] | Damon B. Lesmeister[3,4] | Brooke E. Penaluna[3] | Mark Schulze[2] | Madison Sutton[7] | Marie I. Tosa[4] | Matthew G. Betts[2]

[1]Department of Agricultural, Forest, and Food Science, University of Torino, Grugliasco, Italy; [2]Department of Forest Ecosystems & Society, Oregon State University, Corvallis, Oregon, USA; [3]Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon, USA; [4]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, Oregon, USA; [5]PRISM Climate Group, Northwest Alliance for Computational Science and Engineering, College of Engineering, Oregon State University, Corvallis, Oregon, USA; [6]Department of Forest and Wildlife Ecology, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA and [7]Biological Sciences, Marquette University, Milwakee, Wisconsin, USA

Correspondence
Nicolò Anselmetto
Email: nicolo.anselmetto@unito.it

Funding information
National Science Foundation, Grant/Award Number: LTER8 DEB-2025755; Connecting Europe Facility, Grant/Award Number: INEA/CEF/ICT/A2018/1815462

Handling Editor: Pieter De Frenne

Abstract

1. Correlative species distribution models (SDMs) are essential tools in conservation biology, global change assessment and reserve prioritization, linking species occurrences with environmental conditions. These models often rely on coarse-scale spatial and temporal predictors, overlooking fine-scale environmental conditions experienced by organisms. Additionally, SDMs commonly use short-term occurrence data to make long-term predictions, which can reduce reliability.
2. We hypothesized that long-term and finer temporal resolution data would provide more accurate predictions by capturing population variability under microclimatic conditions. Using data from 37 bird species in the H. J. Andrews Experimental Forest (Oregon, USA), we built SDMs with a 10-year (2010–2019) dataset of breeding season observations at 184 sites. We evaluated four modelling frameworks that differed in temporal extents (short-term [1 year] vs. long-term [10 years]) and resolution (fine vs. coarse) of environmental data. Predictors included hourly microclimate temperatures beneath the forest canopy and LiDAR-derived vegetation variables. We evaluated interannual transferability and compared model performance based on temporal extent, resolution and species traits.
3. Temporally dynamic (long-term) models with higher resolution microclimate data outperformed static and short-term models (AUC and TSS difference ~ 0.06, difference in unreliability index of ~0.04) and were more accurate and spatially consistent, particularly for migratory species. Increased temporal resolution

improved performance for small-bodied species, whereas long-lived, larger species performed similarly in short- and long-term models.

4. *Synthesis and applications*. To our knowledge, this is the first empirical study to demonstrate the benefits of long-term dynamic SDMs with spatially matched predictor variables. If predicting the future of biodiversity under land-use and climate change is the goal, practitioners should consider additional investment in multi-year biodiversity monitoring rather than single-year 'snapshots' of species distributions.

KEYWORDS
bird distribution, correlative models, dynamic models, long-term observations, microclimate, old-growth forests, SDMs

1 | INTRODUCTION

Species distribution models (SDMs) have emerged as indispensable tools in applied ecology. They spatially assess species distributions using a mechanistic or, more commonly, correlative approach (Elith & Leathwick, 2009; Zurell et al., 2020, 2022). Applications include conservation biology (Guisan et al., 2013; Zurell et al., 2022), global change assessment (Newbold et al., 2020; Wenger et al., 2011), species interactions (Wenger et al., 2011) and spatial reserve prioritization (Bicknell et al., 2017; Carroll et al., 2010). Despite their value in predicting the potential impacts of global change, SDMs typically rely on coarse-scale spatiotemporal data due to their widespread availability (Araújo et al., 2019). However, this approach overlooks the fine-scale ecological conditions experienced by many species, creating a mismatch between the scale of predictor data and the fine-grained, long-term needs of conservation decisions (Guisan et al., 2013; Mateo et al., 2019). Consequently, only 1%–5% of conventional correlative SDM studies are relevant for local and regional management decisions (Araújo et al., 2019; Guisan et al., 2013). However, local and landscape-scale management decisions increasingly require accurate models that can predict biodiversity responses to changing site-level conditions (Harris & Betts, 2023; Lembrechts et al., 2019).

The goal of SDMs is often to extrapolate to new locations and time periods under environmental change (Araújo et al., 2019; Engler et al., 2017; Rousseau & Betts, 2022) to inform conservation and management planning. However, conventional SDMs are typically static in time, integrating species occurrence and environmental data collected at a single point in time (i.e. short temporal extent) or averaged over several years or seasons (i.e. coarse temporal resolution; Milanesi et al., 2020; Zurell et al., 2022). These models assume that species are in equilibrium with their environment, occupying all suitable habitats, and that habitat relationships remain unchanged over time (Zurell et al., 2022). However, mobile species commonly violate this assumption, as they often exhibit significant year-to-year variations in their distribution patterns, due to behavioural flexibility (Chávez-González et al., 2020; Frey, Hadley, & Betts, 2016), variation in resource availability and population fluctuations (Yegorova et al., 2013). Short-term data generally do not capture this temporal variability. When the ecological goal is to predict where a species occurs at a variety of scales and through time, alternatives to correlative SDMs exist in the form of process-based (i.e. mechanistic) models, such as ecophysiological models (Kearney & Porter, 2009), population models (Akçakaya, 2000), individual-based population models (Railsback & Harvey, 2002) and dynamic occupation models (MacKenzie et al., 2003). However, these approaches require much more a priori knowledge and experimental data, making them resource-intensive (Briscoe et al., 2021, 2023; Dormann et al., 2012; Tourinho & Vale, 2023).

Birds are an ideal taxon for studying biogeography, global change and ecology in general due to their ecological diversity, sensitivity to environmental changes and the abundance in occurrence data (Anderle et al., 2024; Bonnet et al., 2002; Gregory & van Strien, 2010; Zurell et al., 2022). Their life history traits, particularly body size and migratory behaviour, make them well-suited for examining responses to microclimatic variation. Indeed, large-scale climate and land-use dynamics strongly influence general bird distribution patterns, but forest and landscape characteristics mediate local responses (Northrup et al., 2019; Rigo et al., 2024). Smaller-bodied species, with a higher mass-specific metabolic rate (Brown et al., 2004; Kleiber, 1932) and surface-to-volume ratio (i.e. poorer thermoregulatory ability), are more vulnerable to microclimatic fluctuations in time (Henry et al., 2023; McKechnie & Wolf, 2010). Conversely, larger bodied species exhibit greater physiological resilience to temperature extremes and respond more slowly to small microclimatic variations (Kleiber, 1932; McKechnie & Wolf, 2010). Migratory birds have a high movement capacity and are able to engage in adaptive habitat selection decisions on a year-by-year basis, making them more sensitive to interannual fluctuations (Doligez et al., 2002; Vitasse et al., 2021). On the other hand, resident and partially migratory species respond more quickly to intraannual changes in environmental conditions (Boyles et al., 2011; Cox & Cresswell, 2014). For these reasons, temporally explicit models accounting for local microclimate variation should be more reliable for informing habitat

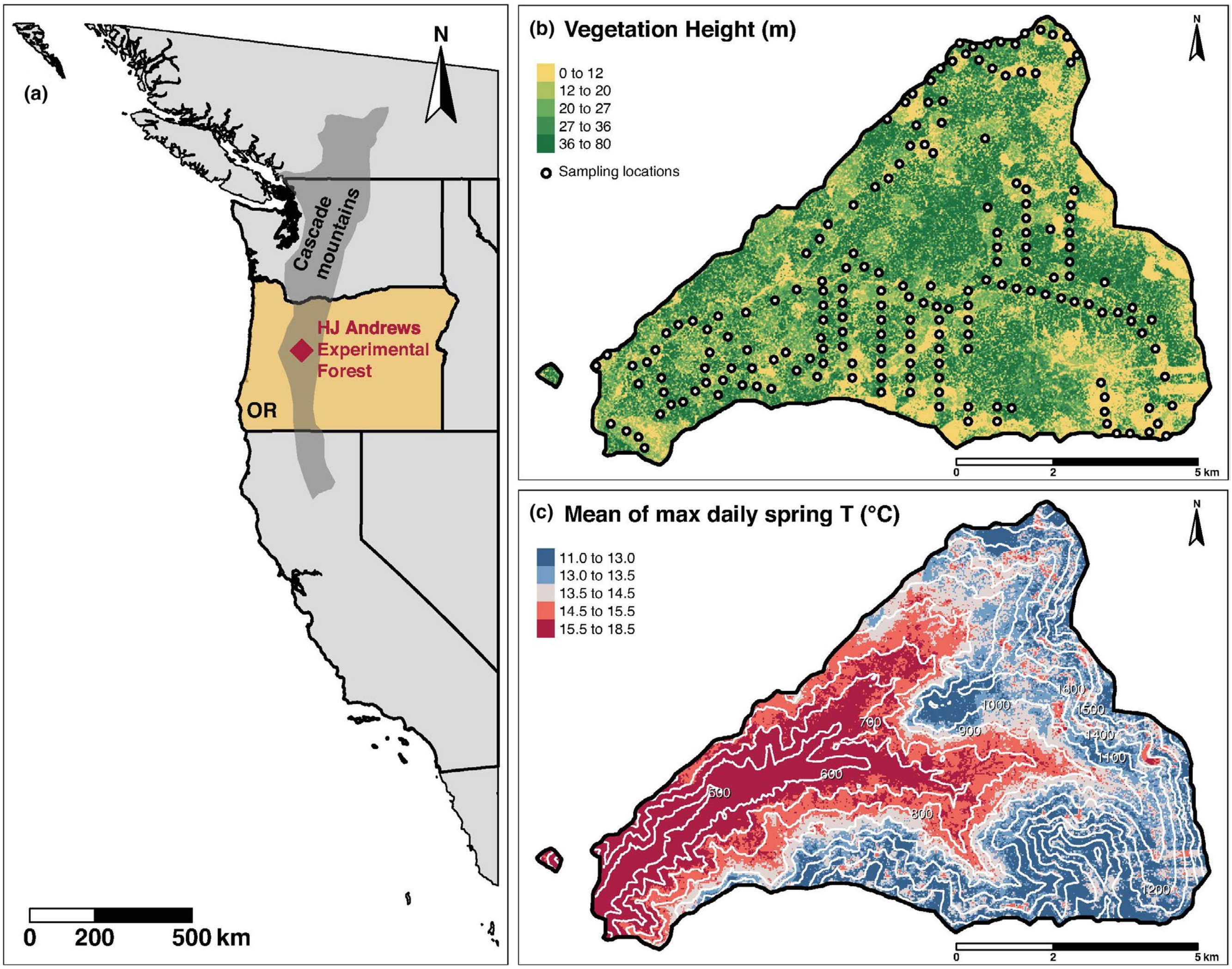
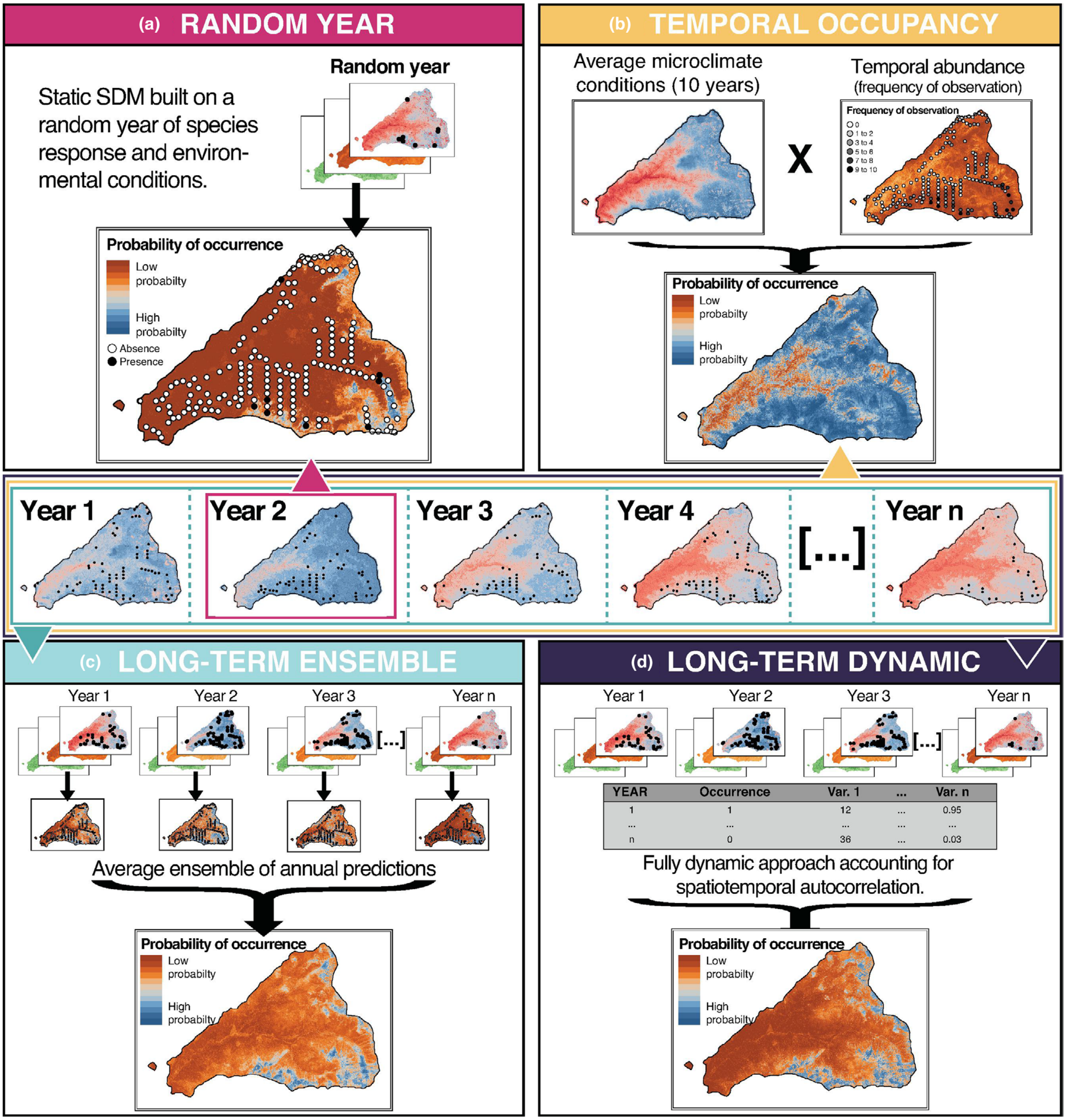


FIGURE 1 (a) Location of the H. J. Andrews Experimental Forest (HJA) in the Cascade Range, USA. (b) Map of vegetation height (from 1-m canopy height model) and sampling locations for bird occurrence and temperature data loggers. (c) Mean of maximum daily spring temperatures and elevation contours.

management and conservation, such as identifying the location and variability of refugia under global changes (Elo et al., 2023; Lembrechts et al., 2019).

Our overall objective was to test the degree to which long-term distributional data improve SDMs' capacity to predict new locations in different time periods. We used long-term datasets on bird distributions (37 species) and under-canopy temperatures collected over 10 years in a forest mosaic across a 900-m elevational gradient at the H.J. Andrews Experimental Forest (HJA) in the Cascades Range (western Oregon, USA; Figure 1). The dataset provides a rare opportunity to assess how long-term population dynamics and fine-scale microclimatic variation influence SDM predictions in time (i.e. over 10 years) and space (i.e. across the HJA forest mountain landscape). We evaluated how SDMs that vary in temporal extent (short-term vs. long-term) and resolution (fine vs. coarse) to determine their ability to forecast bird distributions at the landscape scale over time. Specifically, we compared four SDM frameworks (Figure 2): (i) a single random year of species observations and environmental conditions (reflecting the short temporal extent of most SDMs), (ii) predictor variables with microclimate conditions averaged over 10 years, and the proportion of time a species is observed at a given site across all years (representing models using climate normal and response data collected over several years), (iii) a long-term ensemble of annual models with model predictions averaged across years and (iv) a long-term dynamic SDM approach that accounts for spatiotemporal autocorrelation.

Our central hypothesis was that SDMs using long-term data, by capturing interannual population and distribution dynamics, should provide more reliable predictions to new locations and time periods. However, we also expected that this effect could be mediated by species longevity and movement capacity (i.e. life history and ecological traits). We expected that long-term dynamic models should be particularly important for migratory species with high interannual variation in population sizes and that fine-scale spatiotemporal predictors would improve niche modelling for small-bodied species with relatively short lifespans.



Static SDM built on a random year of species response and environmental conditions.

YEAR	Occurrence	Var. 1	...	Var. n
1	1	12	...	0.95
...	...	...	...	...
n	0	36	...	0.03

Fully dynamic approach accounting for spatiotemporal autocorrelation.

FIGURE 2 Conceptual diagram showing the different modelling frameworks applied in the study. (a) Random year: Static SDMs built using a random year of occurrences and environmental variables. (b) Temporal occupancy: Static SDMs built using the temporal occupancy (no. annual observations/no. years) of a species for each sampling location as weights within the BRT model and the average environmental conditions as predictors. (c) Long-term ensemble: Dynamic SDMs built as an average of annual models. (d) Long-term dynamic SDMs: Fully dynamic models built using all the observations and environmental conditions accounting for spatiotemporal autocorrelation. Coloured boxes around the input data indicate the temporal resolution and extent of the predictors.

2 | MATERIALS AND METHODS

2.1 | Study area

We collected long-term bird and microclimate data at the H.J. Andrews Experimental Forest (HJA) in Oregon's Cascade Range (44.23° N, 122.188° W; Figure 1), encompassing 6400 ha. The forest ranges in elevation from 407 to 1632 m a.s.l., with dominant species including Douglas fir (*Pseudotsuga menziesii*) and western hemlock (*Tsuga heterophylla*) at lower elevations, and Pacific silver fir (*Abies amabilis*), noble fir (*Abies procera*), and mountain hemlock (*Tsuga mertensiana*) at higher elevations. Most of the landscape consists of old-growth forest stands with complex vertical structures and age-class distributions, while 25% of the forest is second-growth, regenerating after clearcutting or partial retention harvests (Kim et al., 2022; Figure 1b). A small portion of high elevation is covered by mountain meadows and shrublands. The climate in the HJA is marine temperate, with an annual mean temperature of 9°C and annual precipitation ranging from 1660 to 2810 mm and mostly occurring from October to April. Local microclimates are influenced by elevation, topography and vegetation structure (Figure 1c; Frey, Hadley, Johnson, et al., 2016; Wolf et al., 2021).

We used bird and microclimate sampling locations ($n = 184$) (Betts et al., 2023; Schulze et al., 2023), as previously described by Frey, Hadley, and Betts (2016), Frey, Hadley, Johnson, et al. (2016), Wolf et al. (2021), and Kim et al. (2022). Points were stratified across elevation gradients (460–1558 m), vegetation structure (plantations $n = 66$ vs. primary forest $n = 118$) and distance from the roads.

2.2 | Overview of SDM framework

SDMs followed the Overview, Data, Model, Assessment, and Prediction (ODMAP) protocol (Zurell et al., 2020) (refer to the Table S1.1 for additional details). We assumed that (i) microclimate and forest structure drive the distribution of our focal bird species within the landscape (Kim et al., 2022), (ii) bird surveys provide adequate and representative data with negligible detection errors and unbiased species identification and (iii) vegetation structure remains stable over the 10 years. All analyses were performed in R version 4.2.3 (R Core Team, 2023) (see Table S1.2 for the specific R packages).

2.3 | Species data

We used a 10-year avian point count inventory collected (2010–2019), covering 184 points (Betts et al., 2023; Frey, Hadley, & Betts, 2016; Kim et al., 2022). Surveys, conducted from May to July, lasted 10 min between 5:00 and 10:30 AM (Figure 1b). Surveyors visited each point up to six times per year from 2010 to 2013 and up to four times per year from 2014 to 2019 (see Kim et al., 2022 for additional details). As this study involved only field observations of birds and did not require animal handling, no ethical approval for animal use was necessary. Site-use permissions were approved for use at the HJ Andrews Experimental Forest. We included 37 species (49% of the 75 species detected in total) for which model calibration and validation over time were possible. We excluded species that were not observed for at least 3 years. We used presence/absence as the response variable and considered a species present at each site when detected at least once during point count surveys within the breeding season, when they are territorial and least likely to move around as vagrants.

We obtained traits (i.e. morphological, ecological, and life history traits) from the AVONET dataset (Tobias et al., 2022), and used them to relate the performance of the SDM to the ecology of the species. A principal component analysis (PCA) was used to reduce the dimensionality of trait data. We retained the first principal component ($p < 0.001$, 73.3% of variance explained), strongly correlated with body size in the analysis. Further information regarding species codes, names, traits and PCA results can be found in Table S1.3 and Figure S1.1.

2.4 | Environmental predictors

2.4.1 | Microclimate data

We used 184 HOBO temperature dataloggers (168 UA-002-64 and 16 U22-001 (Onset Computer Corporation)) affixed to a custom PVC radiation shield mounted on a fibreglass post at 1.5 m height above-ground and facing south to collect data every 20 min year-round (see Frey, Hadley, Johnson, et al., 2016).

Using data from 2009 to 2019, we derived 56 temperature metrics (e.g. monthly minimum, mean, maximum, growing degree days and cooling degree days) from daily minimum, mean and maximum temperatures of the understory. We calculated these metrics using monthly values from July of the previous year until June of the reference year to capture relevant microclimate patterns 1 year before species observations. For example, when referring to the 2015 microclimate, we considered data from 1 July 2014 to 30 June 2015. To create comprehensive microclimate maps for the landscape, we followed the procedure outlined by Wolf et al. (2021) using boosted regression trees (BRT) models with different microclimate metrics as response variables and vegetation, elevation and microtopography as predictors (Frey, Hadley, Johnson, et al., 2016; Wolf et al., 2021). We generated these maps at 25-m resolution. More details on microclimate variables can be found in the Table S1.4.

2.4.2 | Vegetation structure

We obtained 10 vegetation variables from a LiDAR flight (Oregon Lidar Consortium, 2016) operated between May and June 2016 using a Leica ALS80 sensor and capturing an average of 12.64 points per square meter (Oregon Lidar Consortium, 2016). The derived variables encompassed canopy cover, canopy point density and several height metrics. We resampled the vegetation structure raster data at 25 m resolution to ensure spatial consistency with microclimate data (see the Table S1.4 for additional information on vegetation structure).

2.4.3 | Variable pre-selection

We narrowed an initial pool of 66 variables by examining the correlation between variables ($r > 0.9$) and the variance inflation factor (VIF) (Naimi et al., 2014). The algorithm finds a pair of variables with the highest linear correlation (greater than the threshold) and then excludes the variable with the highest VIF in the pair. We performed this operation for all 10 years and retained variables eligible for model calibration for more than 5 years. Our aim was to speed up the computation time and avoid highly correlated variables, even though BRTs are considered robust to multicollinearity (Pichler & Hartig, 2023). We obtained a final set of 41 variables (32 microclimate and 9 vegetation variables; Table S1.4).

2.5 | Species distribution modelling

We used boosted regression trees (BRTs) through the dismo R package v1.3–14 (Hijmans et al., 2023) and the dynamicSDM R package v1.3.2 (Dobson et al., 2023) to model species distribution. BRTs are machine learning models in which multiple decision trees are sequentially trained on the residuals of the preceding tree. This boosting forward-fitting technique enhances the diversity of trees in an ensemble, improving predictive performance and resulting in an additive regression approach that is widely recognized as an effective method in SDMs (Elith et al., 2008).

All models treated response variables (bird distributions) as dichotomous (presence/absence) but varied in the way long-term data were summarized. To assess the interplay between species occurrence and environmental conditions, we employed four distinct modelling frameworks (i.e. random year, temporal occupancy, long-term ensemble and long-term dynamic) based on their temporal extent and resolution, from fully static, short-term models to fully dynamic, long-term models. See Table S1.5 for the list of settings of BRTs for dismo and dynamicSDM packages used in the calibration.

2.5.1 | Random year

This framework is based on a randomly selected year from those available for calibration, creating a snapshot in time. We used the random year approach to reflect studies that obtain data for SDMs from short-term data collection (Figure 2a) and as a proxy of the typical use of occurrence data from sources like GBIF, which are often modelled as a function of average climate conditions without accounting for interannual variability.

2.5.2 | Temporal occupancy

The temporal occupancy framework is based on Snell Taylor et al. (2021) (Figure 2b). This combines average environmental conditions (temporal extent = 10 years, temporal resolution = 1 year) with dynamic species responses (temporal extent = 10 years, temporal resolution = 10 years). By calibrating BRT models using the proportion of times a species was observed over 10 years at a given site (i.e. temporal abundance) as a weight, this approach enables the inclusion of long-term data but does not match the interannual population fluctuations directly to microclimate changes. This makes the temporal occupancy approach more similar to a static approach rather than a dynamic approach. This approach is accessible for annual monitoring datasets. For this approach, we excluded five species (*Catharus ustulatus*, *Cyanocitta stelleri*, *Poecile rufescens*, *Regulus satrapa* and *Setophaga occidentalis*) that had a cumulative 100% prevalence over the 10-year period and no absences were obtained. Therefore, the temporal occupancy framework was applied only to 32 different bird species.

2.5.3 | Long-term ensemble

We built long-term ensemble models by creating year-specific models with year-specific response and predictor variables and then averaging model predictions across all years (Figure 2c). This approach requires both observations and predictor values to be collected at the same temporal resolution over the long term. This ensemble modelling strategy provides a nuanced understanding of temporal dynamics.

2.5.4 | Long-term dynamic

For long-term dynamics models (Milanesi et al., 2020), we split the long-term dataset into five spatiotemporal blocks that account for both temporal and spatial autocorrelation; we then calibrated models using default BRT settings. During calibration, each unique block is excluded in a spatiotemporal cross-validation procedure (Bagchi et al., 2013). In particular, we split blocks through the spatiotemp_block of the dynamicSDM R package v. 1.3.2 (Dobson et al., 2023) using forest structure as the spatial layer for spatial sampling unit splitting (split degree = 1). The model uses each spatiotemporal block as the test dataset and the others as training data. The calibration procedure returns a list of fitted values equal to the length of the blocks. We then calculated both the mean and the uncertainty of the predictions from the five resulting BRT models. As with the long-term ensemble approach, this framework incorporates fine temporal resolutions (1 year) and long temporal extents (10 years), but trains models in such a way that directly incorporates interannual variation (i.e. models are built to minimize error when predicting to different years and spatial blocks in the landscape) (Figure 2d).

2.6 | Model performance and spatial predictions

One challenge associated with validating SDMs is that hold-out (test) data may be correlated with training data due to spatial or temporal autocorrelation (Dormann et al., 2007; Rousseau & Betts, 2022; Valavi et al., 2019). For instance, if a bird species persists at a site due to site fidelity, holding out a single random year from a model trained using the same sites could result in inflated prediction success. To rigorously evaluate model accuracy and calibration while reducing the influence of both spatial and temporal autocorrelation, we implemented a spatiotemporal validation strategy with two main components (Figure S1.2). First, we used a leave-one-year-out cross-validation (e.g. Roberts et al., 2017; Wenger & Olden, 2012), where models were trained on all but one focal year and tested on that held-out year. However, to eliminate the potential bias caused by temporal autocorrelation, we applied an additional validation scheme in which we held out 30% of spatial locations in the single focal year as an independent test set. Models were trained on data from the remaining 70% of locations and from all years but the focal test year. This means that (i) no test location was included in model training, (ii) no data from the test year were used in training and (iii) the test year had no overlap in time or space with the training set. Also, as the bird territory sizes are smaller than the distance between points, no individuals were simultaneously present in test and training data. This procedure allowed us to assess predictive performance on spatially and temporally independent data, reducing the inflation of performance metrics due to individual persistence, spatial autocorrelation or repeated environmental conditions across years. It is important to note that since our objective was not to extrapolate beyond the boundaries of our study region (the HJA), we considered our spatial blocking approach to be sufficient (Rousseau & Betts, 2022; Valavi et al., 2019; Wang et al., 2023).

We calculated several metrics to assess the performance of the models both in terms of accuracy/discrimination and calibration/generalizability (i.e. the agreement between predicted probabilities of occurrence and observation of presence and absence) (Table S1.6). For the former, we employed AUC (Area Under the receiving operator Curve), TSS (True Skill Statistic), *F*1 score, Cohen's kappa, sensitivity and specificity as proxies of model accuracy. We used the value that maximized the TSS as thresholds for threshold-dependent metrics. We also assessed model calibration and generalizability using point biserial correlation through the Pearson's correlation coefficient (*r*), unreliability index, log likelihood, Brier score and slope and intercept of the calibration line (Kenney & Keeping, 1962; Miller et al., 1991).

The primary output of our modelling consisted of predictions about the relative probability of occurrence across the HJA landscape. For dynamic frameworks (i.e. long-term ensemble and long-term dynamic models), we also derived the uncertainty based on the 5th and 95th percentiles of the predictive probability distribution as a bootstrap estimate of uncertainty across the k-fold predictions.

We tested the degree to which the variation in performance among the four frameworks is explained by species' life history traits such as migratory habit (resident, $n=11$; partially migratory, $n=5$; or migratory, $n=21$) and body size (derived through principal component analysis of species' life history traits). We used linear mixed-effect models on performance results to test for differences among modelling frameworks as a function of migratory habit and body size. In all analyses of trait data, we incorporated nested random effects (species within families) to account for phylogenetic similarity between species and graphically checked assumptions of residuals' normality and homoscedasticity. We performed Tukey's post-hoc tests on significant effects ($p<0.05$) with Bonferroni adjustment. We also used Spearman's rank (r_s) to quantify the correlations among spatial predictions (fitted values) from different modelling frameworks. We chose Spearman's rank correlation because we obtained non-normally distributed probabilities, requiring non-parametric tests.

3 | RESULTS

3.1 | Model performance and life history traits

Models utilizing long-term data, particularly dynamic ones (long-term ensemble and long-term dynamic models), had better performance than short-term models and static approaches (random year and temporal occupancy) in terms of both accuracy (e.g. AUC, TSS, *F*1 score) and calibration (e.g. Pearson's *r*, unreliability index, Brier score) across 37 species according to the spatiotemporal validation procedure (Figures 3–5, Figures S2.1 and S2.2). The mean AUC of long-term dynamic models was 0.69, 9% higher than the AUC of random year models (0.63). Similar differences existed for TSS (0.71 vs. 0.66), Pearson's *r* (0.27 vs. 0.19), unreliability index (0.14 vs. 0.18) and Brier score (0.17 vs. 0.18). In general, model performance for models with long-term dynamic frameworks was acceptable to good for most of the species (24 species with AUC > 0.6, TSS > 0.3, $r>0.2$ and low unreliability index), with seven showing excellent accuracy results (AUC > 0.75, TSS > 0.7 and $r>0.4$) (Figure S2.2). Notably, for random year models, 48% of species have validation results that overlap AUC = 0.5 (random performance) but for long-term models this is only 20% (Figure 3). Model performance metrics improved for 86.5% of the species when using long-term dynamic models rather than static models. Relative performance (especially in terms of accuracy) increased as the modelling framework incorporated more complex temporal data, but calibration was lower for the temporal occupancy models rather than random year for most of the species, and long-term ensemble and long-term dynamics had similar Brier score, unreliability index and log likelihood but different calibration slopes and Pearson's *r* (Figure S2.2).

Migratory species had the highest model performance for most of the metrics, followed by residents. Decreases in performance between long-term dynamic and random year models were more severe for resident species; on average, random year models for residents had AUCs that were 0.08 lower than long-term dynamic models (5th percentile = −0.04, 95th percentile = 0.19; Figure 4a) and values of unreliability higher than 0.04 (−0.24, 0.13), indicating both poorer accuracy and calibration (Figure 4c and Table S2.1). The lowest drops in AUC (0.04), TSS (0.05), *F*1 (0.03) and *r* (0.09) were for migratory species. We compared AUC and unreliability index to visualize the trade-offs between accuracy and calibration of the different models (Figure 4b). The centroids of the dynamic models (long-term ensemble and long-term dynamic models) stood out as substantially improved over more static modelling frameworks.

Body size was not significantly related to any performance metrics but the Brier score ($p=0.04$), which showed improving

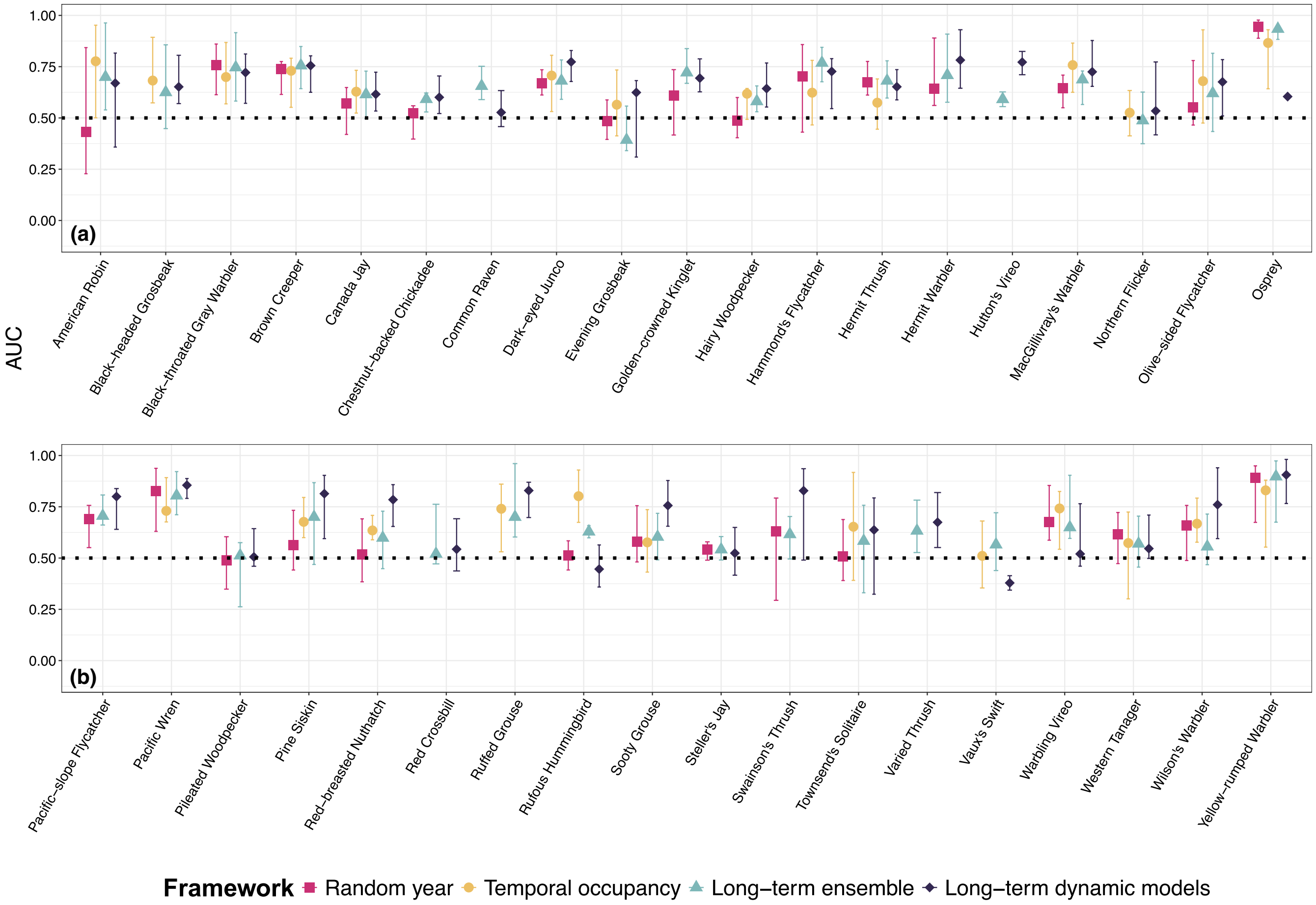


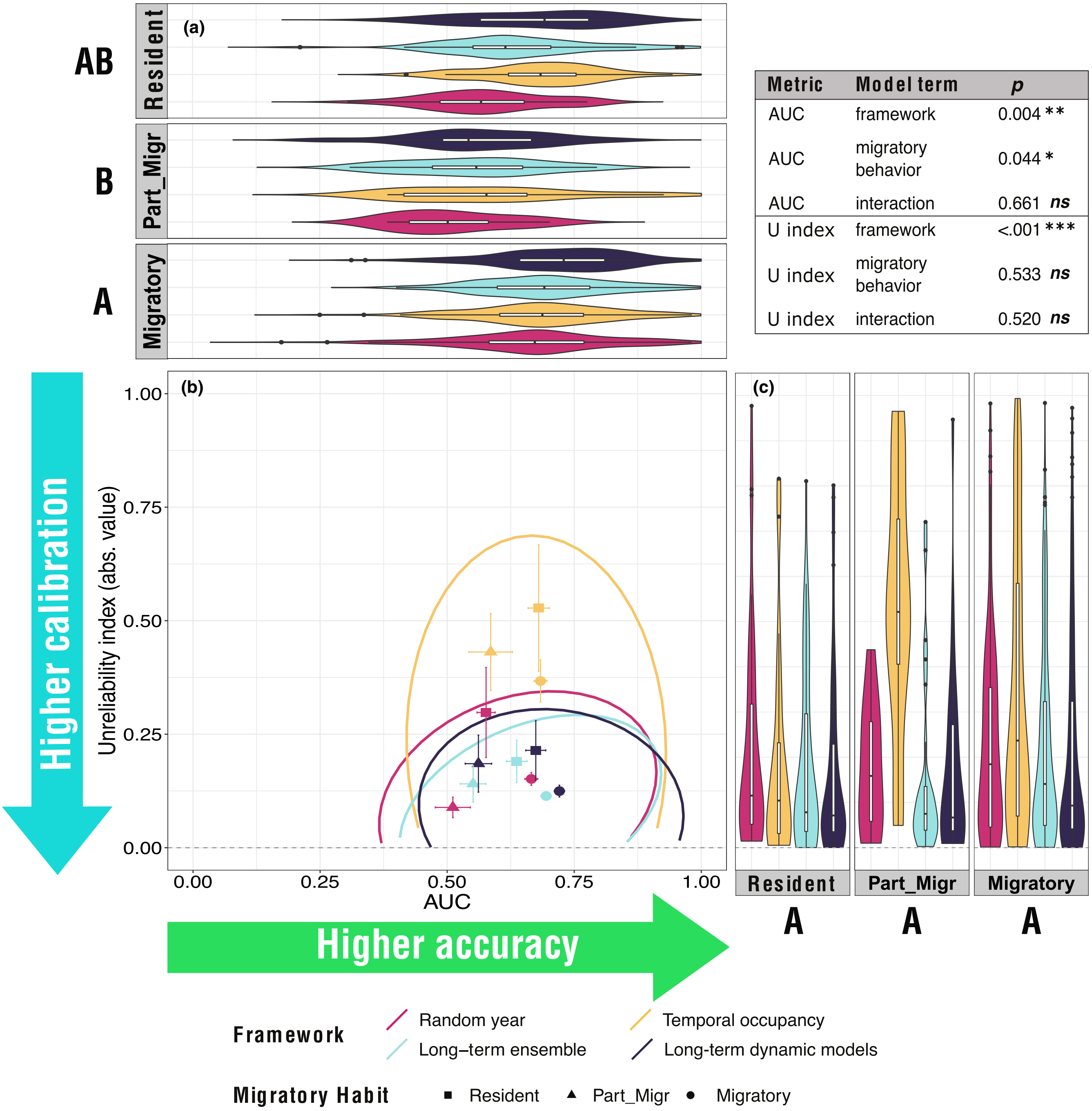
FIGURE 3 Model performance (AUC) according to the spatiotemporal validation procedure across 37 bird species under four different modelling frameworks. (a and b) show the median (symbols) and 5th–95th percentile range (vertical lines) of AUC values for each species. The horizontal dashed line represents AUC = 0.50, indicating random prediction performance. For approximately half of the species, the random year models yielded AUC values crossing this threshold, whereas only one fifth of species showed such results with long-term dynamic models. Species names and codes are provided in Table S1.3.

calibration for larger birds (Figure 5). The interaction between body size and the modelling framework was statistically significant for TSS, Cohen's kappa, *r* and the unreliability index. Smaller birds had the highest accuracy with dynamic models, while larger birds had similar performance across the modelling frameworks, with temporal occupancy and random year models performing better than dynamic models for some metrics.

3.2 | Spatial predictions

We compared the spatial predictions for the probability of occurrence (i.e. fitted values) across modelling frameworks, body sizes and migratory habits. Migratory species showed the greatest degree of similarity among modelling frameworks (mean $r_s = 0.70 \pm 0.03$), followed by resident ($r_s = 0.58 \pm 0.04$) and partially migratory ($r_s = 0.43 \pm 0.06$) species. We observed the greatest similarity between model spatial predictions for the two long-term dynamic approaches for migratory species ($r_s = 0.78 \pm 0.04$) and the lowest between the two static models for partially migratory species ($r_s = 0.29 \pm 0.09$). The long-term ensemble framework showed the most similarities to the other models ($r_s = 0.73 \pm 0.04$ with temporal occupancy and 0.63 ± 0.04 with long-term dynamic and random years). Similarly, spatial predictions were not significantly correlated with body size ($\beta = 0.021 \pm 0.013$; $p = 0.23$), and the interaction term between body size and modelling framework was not statistically significant ($p = 0.06$) (see Table S2.5 for fixed effects of the interactions).

We also visually compared spatial predictions for the 37 birds (Figure 6 and Figures S2.4–S2.10). The two static frameworks showed less variability in spatial patterns (i.e. less heterogeneity in prediction) of occurrence probabilities, with the random year framework returning low occurrence probabilities and temporal occupancy returning high values, as expected. We observed higher probability gradients (i.e. spatial variations in patterns of predicted probability of occurrence) across the landscape in dynamic models; the two dynamic models also showed the greatest similarity in model predictions.



Metric	Model term	p
AUC	framework	0.004 **
AUC	migratory behavior	0.044 *
AUC	interaction	0.661 *ns*
U index	framework	<.001 ***
U index	migratory behavior	0.533 *ns*
U index	interaction	0.520 *ns*

FIGURE 4 Results of the spatiotemporal validation of the different models according to the four modelling frameworks and three migratory habits (resident, partially migratory, and migratory). (a) Violin plots of AUC results, (b) mean (centroids) and 95th-percentile ellipsoids of AUC and unreliability index and (c) violin plots of unreliability index for each migratory habit. Letters indicate results of post-hoc tests of linear mixed-effect models (see Tables S2.2 and S2.3; Figures S2.1 and S2.2 for further details). Uppercase letters indicate significant differences for AUC (panel a) and unreliability index (panel c) among the migratory habits. We observed an improvement in model performance with increasing temporal complexity (dynamicity) of the models.

4 | DISCUSSION

To our knowledge, ours is the first study to test for potential performance boosts of dynamic long-term SDMs on an empirical dataset with spatially matched predictor variables (i.e. microclimate data at the scale of individual sample points). Model performance was influenced by temporal extent and resolution, with single-year models performing worst for every metric. Among the other models, long-term dynamic models outperformed temporal occupancy and long-term ensemble approaches for most metrics.

Vertebrate populations are known to fluctuate interannually (Lack, 1954). Behavioural plasticity allows vagile animals to switch

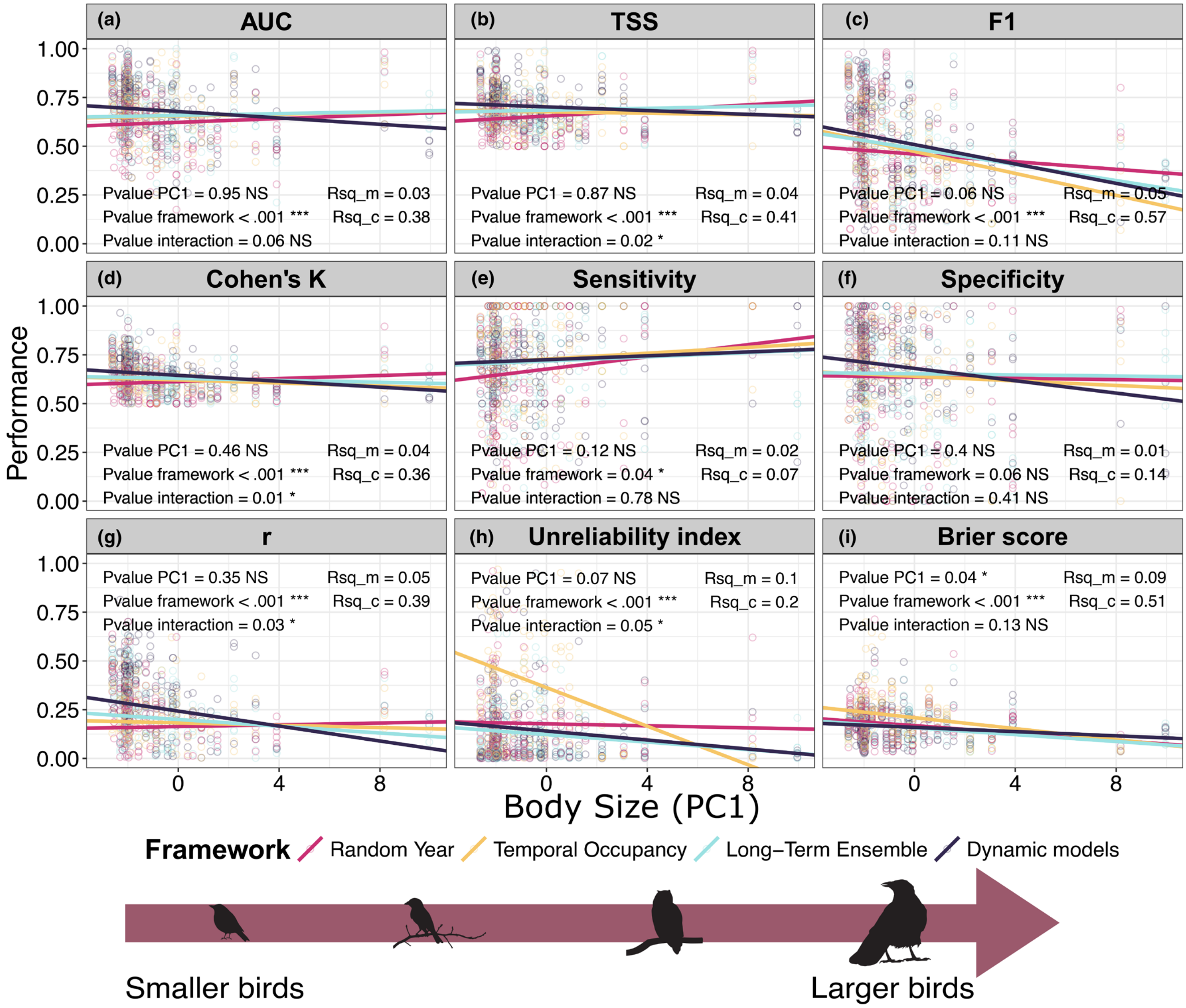


FIGURE 5 Relationship between model performance from the spatiotemporal validation according to the four modelling frameworks and the body size of the species (see Figure S1.1). The panels indicate (a) AUC, (b) TSS, (c) *F1* score, (d) Cohen's kappa, (e) sensitivity, (f) specificity, (g) point biserial correlation (Pearson's correlation; *r*), (h) unreliability index and (i) Brier score. *p* values and R^2 are reported in the panels. The arrow below indicates the body size gradient across the *x*-axis. The importance of dynamic models emerges, especially for short-lived smaller birds.

spatial locations over time (Fretwell & Lucas, 1969). Given this interannual dynamism in distributions, it is not surprising that models incorporating long-term data with interannual dynamics performed the best when predicting to new periods. Milanesi et al. (2020) demonstrated similar improvements in dynamic SDMs using simulated data. Long-term models are less biased by short-term population fluctuations that can cause mismatches between habitat quality and species distributions. For instance, random mortality of individuals and inefficient habitat selection behaviour can cause animals to select suboptimal habitats (Van Horne, 1983). Over long periods of time, stochastic events causing such mismatches are more likely to be averaged out. Therefore, single-year models are more prone to extrapolation errors, especially when test years feature environmental extremes that are not present in training datasets. Extrapolation is well known to cause model misspecification and, therefore, reduce accuracy (Betts et al., 2006; Rousseau & Betts, 2022). This likely explains why some models in our study—especially those calibrated according to the random year framework—had very poor performance, with, for example, AUC < 0.5. On the contrary, long-term dynamic models are trained using data across all years in spatial blocks, thus much less likely to result in extrapolation errors. Interestingly, seven out of the nine species for which long-term models either failed or were worse than random year models were for species with large home ranges or transient behaviour (e.g. common raven, osprey, red crossbill, Vaux's swift).

Migratory species had the highest model performance across all models. We speculate that this could be due to the high movement capacity of these species, which enables them to engage in adaptive

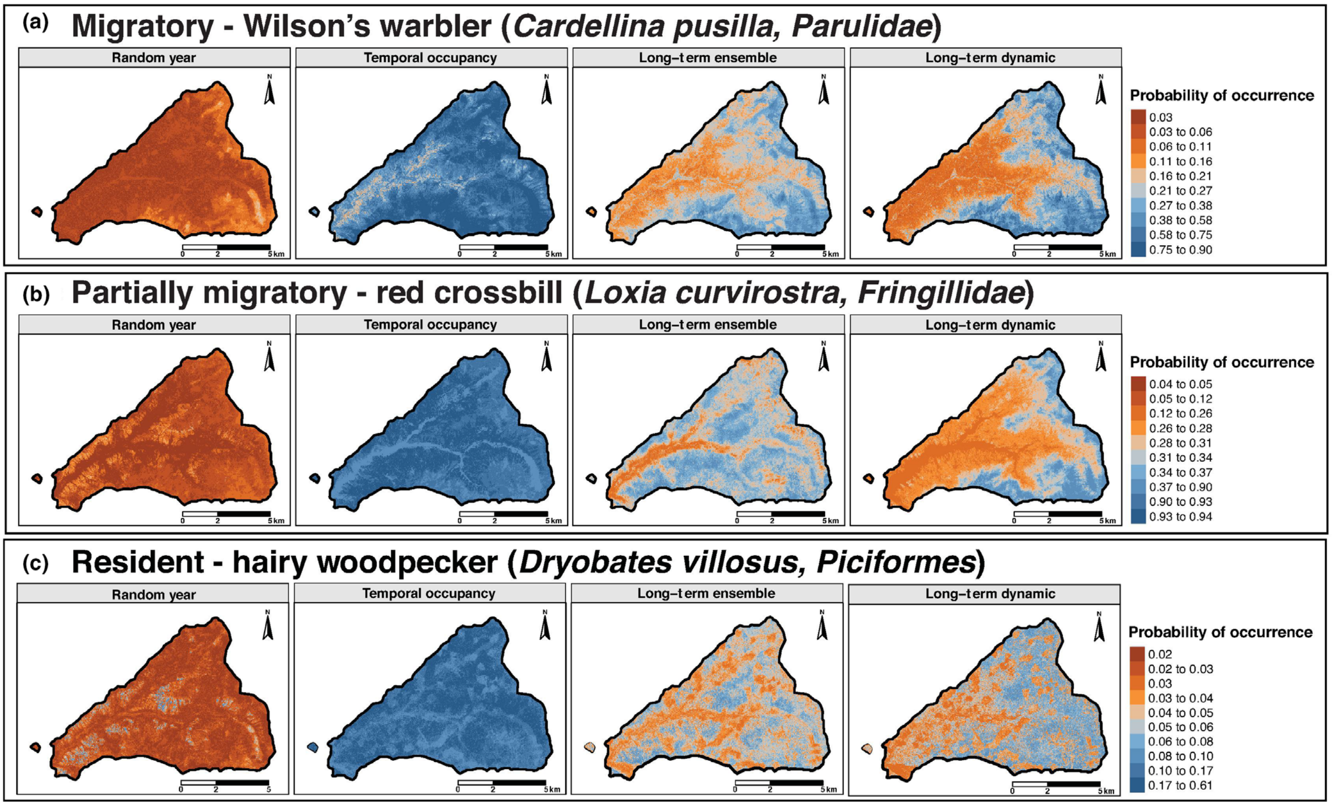


FIGURE 6 Spatial predictions of probability of occurrence for three example species, (a) Wilson's warbler, (b) red crossbill and (c) hairy woodpecker according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks. Maps for the other species can be found in the Supporting Information.

habitat selection decisions on a year-by-year basis. Indeed, some migratory passerines are known to have efficient algorithms for selecting habitat (Betts et al., 2008; Doligez et al., 2002). For resident species, temporal occupancy models provided significant improvements over random year models or surpassed long-term ensemble models for some metrics, while further gains from dynamic approaches were smaller. This pattern of performance suggests diminishing returns between model complexity and performance for residents, but not for migratory species. Regarding body size, smaller birds benefited more from dynamic models, likely because their greater sensitivity to microclimate fluctuations necessitates higher mobility.

For several decades, conservation and applied ecology has relied on SDMs to project the effects of global change on species distributions (Tourinho & Vale, 2023; Zurell et al., 2022). The rapid increase in data availability at different spatial and temporal scales is pushing ecological models to promising frontiers (Araújo et al., 2019; Tosa et al., 2021; Tourinho & Vale, 2023; Zurell et al., 2020, 2022). Although process-based and hybrid models are growing in popularity, they require detailed demographic and dispersal data, which remain scarce. Consequently, correlative SDMs still dominate the field and can be considered flexible across taxa and landscape types. However, most SDM studies use short-term data or ignore temporal scales in modelling, resulting in tools with limited applicability to biodiversity conservation and landscape planning, especially in a time of rapid global change.

Out of 82 papers gleaned from Web of Science (2022–2024) on correlative SDMs (search terms in title, abstract and keywords: ("species distribution model*" OR "environmental niche model*" OR "habitat suitability model*" OR sdm* OR enm* OR hsm*) AND ("vertebrat*" OR "bird*" OR "mammal*" OR "reptil*" OR "amphib*" OR "chordata")) in the five top journals dealing with this topic (*Diversity and Distribution*, *Journal of Biogeography*, *Journal for Nature Conservation*, *Global Ecology and Conservation* and *Science of the Total Environment*) only four papers (<5%) applied an approach that can be considered to be long term and dynamic (i.e. Elo et al., 2023; Lindenmayer et al., 2022; Stevens et al., 2022; Thomsen et al., 2023). Roughly one-third of studies failed to match the temporal extent of species occurrences and predictors. It is promising that we found some papers using long-term observations, but few authors incorporate these directly in the models to improve their calibration and reliability. However, we recognize the cost and labour associated with collecting these long-term data.

We recommend prioritizing long-term data in SDMs whenever possible, especially when fine-spatial scale variations, such as microclimate data, are available. The question is, how long is long enough? No universal rules exist—and it is beyond the scope of this paper to provide a clear answer. The ideal temporal extent likely varies based on taxa life expectancy, population dynamics, life history traits and climate variability. Mobile, short-living taxa (e.g. birds, small mammals) that are more coupled to shorter climate conditions than stable and long-living ones (e.g. trees, lichens) may require shorter temporal extents but higher temporal resolution. Generally, dynamic SDMs should be preferred whenever a collection of data exists over multiple years (i.e. long-term). Our findings highlight the potential of dynamic SDMs to improve predictions of species distributions under global change, making them important for applied ecology, such as conservation planning and habitat management in rapidly changing environments. Dynamic approaches might be of particular interest in regions and landscapes with greater environmental variability and mosaic complexity (e.g. mountain ranges, ecotones and transition zones) and when dealing with global change (Bradter et al., 2022; Milanesi et al., 2020). Indeed, recent and predicted future shifts in climate and land use are characterized by rapid dynamics and large intra- and interannual variations (IPCC, 2023). When only short-term data are available, it is important to temporally match predictors and responses.

Managers and policymakers increasingly require reliable distribution models at sufficiently fine spatial scales to be relevant in conservation planning. For instance, long-term dynamic SDMs will be more reliable for locating microclimate refugia (Milanesi et al., 2020; Wolf et al., 2021), fine-scale biodiversity hotspots and refugia (Lembrechts et al., 2019; Lenoir et al., 2017), and for monitoring habitat changes over the long term in relation to management actions (Jefferys et al., 2024; Malchow et al., 2024). By integrating interannual variability through nuanced modelling frameworks and fine-scale microclimatic drivers, we demonstrate a practical and scalable application of what will increasingly be necessary in a rapidly changing biosphere: the incorporation of environmental dynamism and microclimate into ecological models (e.g. Bradter et al., 2022; Elo et al., 2023; Lembrechts et al., 2019; Milanesi et al., 2020).

AUTHOR CONTRIBUTIONS

Nicolò Anselmetto and Matthew G. Betts conceived the ideas and designed the methodology together with Matteo Garbarino; Nicolò Anselmetto led the methodological process with support from Matteo Garbarino, Matthew J. Weldy, David M. Bell, Clinton W. Epps, Nina Ferrari, Damon B. Lesmeister, Mark Schulze, Marie I. Tosa and Matthew G. Betts; Nicolò Anselmetto analysed the data with support from Matthew J. Weldy and Matthew G. Betts; Nicolò Anselmetto led the writing of the manuscript with support from Matteo Garbarino, Matthew J. Weldy, David M. Bell, Christopher Daly, Clinton W. Epps, Nina Ferrari, Hankyu Kim, Joseph A. LaManna, Damon B. Lesmeister, Brooke E. Penaluna, Mark Schulze, Madison Sutton, Marie I. Tosa and Matthew G. Betts. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

Data were provided by the H.J. Andrews Experimental Forest and Long Term Ecological Research (LTER) programme, administered cooperatively by Oregon State University, the USDA Forest Service Pacific Northwest Research Station and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the grant LTER8 DEB-2025755. Nicolò Anselmetto was funded by the European Union through the Connecting Europe Facility grant number INEA/CEF/ICT/A2018/1815462 during the research period. We would also like to thank the many field assistants who collected 14 years of point count data across 184 points in challenging conditions. Open access publishing facilitated by Universita degli Studi di Torino, as part of the Wiley - CRUI-CARE agreement.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

Example data and code are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.j3tx95xqb (Anselmetto et al., 2025). Raw bird point count data (code SA024) and microclimate data (code MS045) are available at the H.J. Andrews LTER database (https://andlter.forestry.oregonstate.edu/data/). LiDAR data are available at the Oregon Department of Geology and Mineral Industries Oregon Lidar Consortium website (Quad name: Carpenter Mountain, Mackenzie River and Blue River; https://www.oregon.gov/dogami/lidar/Pages/index.aspx).

ORCID

Nicolò Anselmetto https://orcid.org/0000-0003-4311-0528
Matteo Garbarino https://orcid.org/0000-0002-9010-1731
Matthew J. Weldy https://orcid.org/0000-0002-8627-2466
Clinton W. Epps https://orcid.org/0000-0001-6577-1840
Nina Ferrari https://orcid.org/0000-0003-4205-9708
Hankyu Kim https://orcid.org/0000-0001-8682-8535
Joseph A. LaManna https://orcid.org/0000-0002-8229-7973
Damon B. Lesmeister https://orcid.org/0000-0003-1102-0122
Brooke E. Penaluna https://orcid.org/0000-0001-7215-770X
Mark Schulze https://orcid.org/0000-0003-0438-4091
Madison Sutton https://orcid.org/0009-0001-2416-2966
Marie I. Tosa https://orcid.org/0000-0002-9020-5098
Matthew G. Betts https://orcid.org/0000-0002-7100-2551

REFERENCES

Akçakaya, H. R. (2000). Viability analyses with habitat-based metapopulation models. *Population Ecology*, *42*(1), 45–53. https://doi.org/10.1007/s101440050043

Anderle, M., Brambilla, M., Angelini, L., Guariento, E., Paniccia, C., Plunger, J., Seeber, J., Stifter, S., Tappeiner, U., Tasser, E., & Hilpold, A. (2024). Efficiency of birds as bioindicators for other taxa in mountain farmlands. *Ecological Indicators*, *158*, 111569. https://doi.org/10.1016/j.ecolind.2024.111569

Anselmetto, N., Garbarino, M., Weldy, M. J., Bell, D. M., Daly, C., Epps, C. W., Ferrari, N., Kim, H., LaManna, J. A., Lesmeister, D. B., Penaluna, B. E., Schulze, M., Sutton, M., Tosa, M. I., & Betts, M. G. (2025). Data from: Leveraging long-term data to improve biodiversity monitoring with species distribution models. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.j3tx95xqb

Araújo, M. B., Anderson, R. P., Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Bagchi, R., Crosby, M., Huntley, B., Hole, D. G., Butchart, S. H. M., Collingham, Y., Kalra, M., Rajkumar, J., Rahmani, A., Pandey, M., Gurung, H., Trai, L. T., van Quang, N., & Willis, S. G. (2013). Evaluating the effectiveness of conservation site networks under climate change: Accounting for uncertainty. *Global Change Biology*, *19*(4), 1236–1248. https://doi.org/10.1111/gcb.12123

Betts, M. G., Diamond, A. W., Forbes, G. J., Villard, M. A., & Gunn, J. S. (2006). The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, *191*(2), 197–224. https://doi.org/10.1016/j.ecolmodel.2005.04.027

Betts, M. G., Frey, S. J., & Kim, H. (2023). Forest-wide bird survey at 183 sample sites the Andrews Experimental Forest from 2009 to present ver 5. Environmental Data Initiative. https://doi.org/10.6073/pasta/161101744422c412802e8287dbd8b663

Betts, M. G., Hadley, A. S., Rodenhouse, N., & Nocera, J. J. (2008). Social information trumps vegetation structure in breeding-site selection by a migrant songbird. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1648), 2257–2263. https://doi.org/10.1098/rspb.2008.0217

Bicknell, J. E., Collins, M. B., Pickles, R. S. A., McCann, N. P., Bernard, C. R., Fernandes, D. J., Miller, M. G. R., James, S. M., Williams, A. U., Struebig, M. J., Davies, Z. G., & Smith, R. J. (2017). Designing protected area networks that translate international conservation commitments into national action. *Biological Conservation*, *214*, 168–175. https://doi.org/10.1016/j.biocon.2017.08.024

Bonnet, X., Shine, R., & Lourdais, O. (2002). Taxonomic chauvinism. *Trends in Ecology & Evolution*, *17*(1), 1–3. https://doi.org/10.1016/S0169-5347(01)02381-3

Boyles, J. G., Seebacher, F., Smit, B., & McKechnie, A. E. (2011). Adaptive thermoregulation in endotherms may alter responses to climate change. *Integrative and Comparative Biology*, *51*(5), 676–690. https://doi.org/10.1093/icb/icr053

Bradter, U., Johnston, A., Hochachka, W. M., Soultan, A., Brommer, J. E., Gaget, E., Kålås, J. A., Lehikoinen, A., Lindström, Å., Piirainen, S., Pavón-Jordán, D., Pärt, T., Øien, I. J., & Sandercock, B. K. (2022). Decomposing the spatial and temporal effects of climate on bird populations in northern European mountains. *Global Change Biology*, *28*(21), 6209–6227. https://doi.org/10.1111/gcb.16355

Briscoe, N. J., Morris, S. D., Mathewson, P. D., Buckley, L. B., Jusup, M., Levy, O., Maclean, I. M. D., Pincebourde, S., Riddell, E. A., Roberts, J. A., Schouten, R., Sears, M. W., & Kearney, M. R. (2023). Mechanistic forecasts of species responses to climate change: The promise of biophysical ecology. *Global Change Biology*, *29*(6), 1451–1470. https://doi.org/10.1111/gcb.16557

Briscoe, N. J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A. K., Kéry, M., Schmid, H., & Guillera-Arroita, G. (2021). Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. *Global Change Biology*, *27*, 4269–4282. https://doi.org/10.1111/gcb.15723

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*(7), 1771–1789. https://doi.org/10.1890/03-9000

Carroll, C., Dunk, J. R., & Moilanen, A. (2010). Optimizing resiliency of reserve networks to climate change: Multispecies conservation planning in the Pacific northwest, USA. *Global Change Biology*, *16*(3), 891–904. https://doi.org/10.1111/j.1365-2486.2009.01965.x

Chávez-González, E., Vizentin-Bugoni, J., Vázquez, D. P., MacGregor-Fors, I., Dáttilo, W., & Ortiz-Pulido, R. (2020). Drivers of the structure of plant–hummingbird interaction networks at multiple temporal scales. *Oecologia*, *193*(4), 913–924. https://doi.org/10.1007/s00442-020-04727-4

Cox, D. T., & Cresswell, W. (2014). Mass gained during breeding positively correlates with adult survival because both reflect life history adaptation to seasonal food availability. *Oecologia*, *174*, 1197–1204. https://doi.org/10.1007/s00442-013-2859-5

Dobson, R., Challinor, A. J., Cheke, R. A., Jennings, S., Willis, S. G., & Dallimer, M. (2023). dynamicSDM: An R package for species geographical distribution and abundance modelling at high spatiotemporal resolution. *Methods in Ecology and Evolution*, *14*, 2041-210X.14101. https://doi.org/10.1111/2041-210X.14101

Doligez, B., Danchin, E., & Clobert, J. (2002). Public information and breeding habitat selection in a wild bird population. *Science*, *297*(5584), 1168–1170. https://doi.org/10.1126/science.1072838

Dormann, C. F., McPherson, J. M., Araújo, B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*(5), 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, *39*(12), 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.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*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elo, M., Kajanus, M. H., Tolvanen, J., Devictor, V., Forsman, J. T., Lehikoinen, A., Mönkkönen, M., Thorson, J. T., Vollstädt, M. G. R., & Kivelä, S. M. (2023). Do large-scale associations in birds imply biotic interactions or environmental filtering? *Journal of Biogeography*, *50*(1), 169–182. https://doi.org/10.1111/jbi.14520

Engler, J. O., Stiels, D., Schidelko, K., Strubbe, D., Quillfeldt, P., & Brambilla, M. (2017). Avian SDMs: Current state, challenges, and opportunities. *Journal of Avian Biology*, *48*(12), 1483–1504. https://doi.org/10.1111/jav.01248

Fretwell, S. D., & Lucas, H. L. (1969). On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica*, *19*, 16–36. https://doi.org/10.1007/BF01601953

Frey, S. J. K., Hadley, A. S., & Betts, M. G. (2016). Microclimate predicts within-season distribution dynamics of montane forest birds. *Diversity and Distributions*, *22*(9), 944–959. https://doi.org/10.1111/ddi.12456

Frey, S. J. K., Hadley, A. S., Johnson, S. L., Schulze, M., Jones, J. A., & Betts, M. G. (2016). Spatial models reveal the microclimatic buffering capacity of old-growth forests. *Science Advances*, *2*(4), e1501392. https://doi.org/10.1126/sciadv.1501392

Gregory, R. D., & van Strien, A. (2010). Wild bird indicators: Using composite population trends of birds as measures of environmental health. *Ornithological Science*, *9*(1), 3–22. https://doi.org/10.2326/osj.9.3

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435. https://doi.org/10.1111/ele.12189

Harris, S. H., & Betts, M. G. (2023). Selecting among land sparing, sharing and triad in a temperate rainforest depends on biodiversity and timber production targets. *Journal of Applied Ecology*, *60*(4), 737–750. https://doi.org/10.1111/1365-2664.14385

Henry, E., Santini, L., Huijbregts, M. A., & Benítez-López, A. (2023). Unveiling the environmental drivers of intraspecific body size variation in terrestrial vertebrates. *Global Ecology and Biogeography*, *32*(2), 267–280. https://doi.org/10.1111/geb.13621

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2023). dismo: Species distribution modeling. R package version 1.3–14. https://CRAN.R-project.org/package=dismo

IPCC. (2023). Sections. In Core Writing Team, H. Lee, & J. Romero (Eds.), *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 35–115). IPCC. https://doi.org/10.59327/IPCC/AR6-9789291691647

Jefferys, K. M., Betts, M. G., Robinson, W. D., Jefferys, K., Betts, M., Curtis, J., Hallman, T., Smith, A., Strevens, C., & Aguirre-Gutiérrez, J. (2024). Breeding habitat loss linked to declines in rufous hummingbirds. *Avian Conservation and Ecology*, *19*, art2. https://doi.org/10.5751/ACE-02681-190202

Kearney, M., & Porter, W. (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters*, *12*(4), 334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Kenney, J. F., & Keeping, E. S. (1962). *Root mean square. "Mathematics of Statistics"* (3rd ed.) (pp. 59–60). Van Nostrand.

Kim, H., McComb, B. C., Frey, S. J., Frey, S. J. K., Bell, D. M., & Betts, M. G. (2022). Forest microclimate and composition mediate long-term trends of breeding bird populations. *Global Change Biology*, *28*(21), 6180–6193. https://doi.org/10.1111/gcb.16353

Kleiber, M. (1932). Body size and metabolism. *Hilgardia*, *6*(11), 315–353.

Lack, D. (1954). *The natural regulation of animal numbers*. Oxford University Press.

Lembrechts, J. J., Nijs, I., & Lenoir, J. (2019). Incorporating microclimate into species distribution models. *Ecography*, *42*(7), 1267–1279. https://doi.org/10.1111/ecog.03947

Lenoir, J., Hattab, T., & Pierre, G. (2017). Climatic microrefugia under anthropogenic climate change: Implications for species redistribution. *Ecography*, *40*(2), 253–266. https://doi.org/10.1111/ecog.02788

Lindenmayer, D. B., Blanchard, W., Bowd, E., Scheele, B. C., Foster, C., Lavery, T., McBurney, L., & Blair, D. (2022). Rapid bird species recovery following high-severity wildfire but in the absence of early successional specialists. *Diversity and Distributions*, *28*(10), 2110–2123. https://doi.org/10.1111/ddi.13611

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*(8), 2200–2207. https://doi.org/10.1890/02-3090

Malchow, A. K., Fandos, G., Kormann, U. G., Grüebler, M. U., Kéry, M., Hartig, F., & Zurell, D. (2024). Fitting individual-based models of spatial population dynamics to long-term monitoring data. *Ecological Applications*, *34*(4), e2966. https://doi.org/10.1002/eap.2966

Mateo, R. G., Gastón, A., Aroca-Fernández, M. J., Broennimann, O., Guisan, A., Saura, S., & García-Viñas, J. I. (2019). Hierarchical species distribution models in support of vegetation conservation at the landscape scale. *Journal of Vegetation Science*, *30*(2), 386–396. https://doi.org/10.1111/jvs.12726

McKechnie, A. E., & Wolf, B. O. (2010). Climate change increases the likelihood of catastrophic avian mortality events during extreme heat waves. *Biology Letters*, *6*(2), 253–256. https://doi.org/10.1098/rsbl.2009.0702

Milanesi, P., Della Rocca, F., & Robinson, R. A. (2020). Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. *Ecology and Evolution*, *10*(2), 1087–1092. https://doi.org/10.1002/ece3.5938

Miller, M. E., Hui, S. L., & Tierney, W. M. (1991). Validation techniques for logistic regression models. *Statistics in Medicine*, *10*(8), 1213–1226. https://doi.org/10.1002/sim.4780100805

Naimi, B., Hamm, N. A., Groen, T. A., Hamm, N. A. S., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling. *Ecography*, *37*, 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Newbold, T., Oppenheimer, P., Etard, A., & Williams, J. J. (2020). Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology & Evolution*, *4*(12), 1630–1638. https://doi.org/10.1038/s41559-020-01303-0

Northrup, J. M., Rivers, J. W., Yang, Z., & Betts, M. G. (2019). Synergistic effects of climate and land-use change influence broad-scale avian population declines. *Global Change Biology*, *25*(5), 1561–1575. https://doi.org/10.1111/gcb.14571

Oregon Lidar Consortium. (2016). 2016 McKenzie River Lidar Project lidar QC acceptance report. https://www.oregongeology.org/pubs/ldq/reports/2016_OLC_McKenzie_River_Acceptance_Report.pdf

Pichler, M., & Hartig, F. (2023). Can predictive models be used for causal inference? arXiv Preprint arXiv:2306.10551. https://doi.org/10.48550/arXiv.2306.10551

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Railsback, S. F., & Harvey, B. C. (2002). Analysis of habitat-selection rules using anindividual-based model. *Ecology*, *83*(7), 1817–1830. https://doi.org/10.1890/0012-9658(2002)083[1817:AOHSRU]2.0.CO;2

Rigo, F., Paniccia, C., Anderle, M., Chianucci, F., Obojes, N., Tappeiner, U., Hilpold, A., & Mina, M. (2024). Relating forest structural characteristics to bat and bird diversity in the Italian Alps. *Forest Ecology and Management*, *554*, 121673. https://doi.org/10.1016/j.foreco.2023.121673

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rousseau, J. S., & Betts, M. G. (2022). Factors influencing transferability in species distribution models. *Ecography*, *2022*(7), e06060. https://doi.org/10.1111/ecog.06060

Schulze, M. D., Betts, M. G., & Frey, S. J. (2023). Air temperature at core phenology sites and additional bird monitoring sites in the Andrews Experimental Forest, 2009 to present ver 5. Environmental Data Initiative. https://doi.org/10.6073/pasta/cbf6e8a684a074cca95b5f422ca5f0f2

Snell Taylor, S., Di Cecco, G., & Hurlbert, A. H. (2021). Using temporal occupancy to predict avian species distributions. *Diversity and Distributions*, *27*(8), 1477–1488. https://doi.org/10.1111/ddi.13296

Stevens, B. S., Conway, C. J., Luke, K., Weldon, A., Hand, C. E., Schwarzer, A., Smith, F., Watson, C., & Watts, B. D. (2022). Large-scale distribution models for optimal prediction of eastern black rail habitat within tidal ecosystems. *Global Ecology and Conservation*, *38*, e02222. https://doi.org/10.1016/j.gecco.2022.e02222

Thomsen, S., Loos, J., Stewart, F. A., & Piel, A. K. (2023). Wildlife habitat association over a twelve-year period (2008–2020) in the greater Mahale ecosystem, western Tanzania. *Journal for Nature Conservation*, *75*, 126464. https://doi.org/10.1016/j.jnc.2023.126464

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., … Schleuning, M. (2022). AVONET: Morphological, ecological and geographical data for all birds. *Ecology Letters*, *25*(3), 581–597. https://doi.org/10.1111/ele.13898

Tosa, M. I., Dziedzic, E. H., Appel, C. L., Urbina, J., Massey, A., Ruprecht, J., Eriksson, C. E., Dolliver, J. E., Lesmeister, D. B., Betts, M. G., Peres, C. A., & Levi, T. (2021). The rapid rise of next-generation natural history. *Frontiers in Ecology and Evolution*, *9*, 698131. https://doi.org/10.3389/fevo.2021.698131

Tourinho, L., & Vale, M. M. (2023). Choosing among correlative, mechanistic, and hybrid models of species' niche and distribution. *Integrative Zoology*, *18*(1), 93–109. https://doi.org/10.1111/1749-4877.12618

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/2041-210X.13107

Van Horne, B. (1983). Density as a misleading indicator of habitat quality. *The Journal of Wildlife Management*, *47*, 893–901. https://doi.org/10.2307/3808148

Vitasse, Y., Ursenbacher, S., Klein, G., Bohnenstengel, T., Chittaro, Y., Delestrade, A., Monnerat, C., Rebetez, M., Rixen, C., Strebel, N., Schmidt, B. R., Wipf, S., Wohlgemuth, T., Yoccoz, N. G., & Lenoir, J. (2021). Phenological and elevational shifts of plants, animals and fungi under climate change in the European Alps. *Biological Reviews*, *96*(5), 1816–1835. https://doi.org/10.1111/brv.12727

Wang, J., Hopkins, L., Hallman, T., Robinson, W. D., & Hutchinson, R. (2023). Cross-validation for geospatial data: Estimating generalization performance in geostatistical problems. *Transactions on Machine Learning Research*, 1–40.

Wenger, S. J., Isaak, D. J., Luce, C. H., Neville, H. M., Fausch, K. D., Dunham, J. B., Dauwalter, D. C., Young, M. K., Elsner, M. M., Rieman, B. E., Hamlet, A. F., & Williams, J. E. (2011). Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(34), 14175–14180. https://doi.org/10.1073/pnas.1103097108

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*(2), 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Wolf, C., Bell, D. M., Kim, H., Nelson, M. P., Schulze, M., & Betts, M. G. (2021). Temporal consistency of undercanopy thermal refugia in old-growth forest. *Agricultural and Forest Meteorology*, *307*, 108520. https://doi.org/10.1016/j.agrformet.2021.108520

Yegorova, S., Betts, M. G., Hagar, J., & Puettmann, K. J. (2013). Bird-vegetation associations in thinned and unthinned young Douglas-fir forests 10 years after thinning. *Forest Ecology and Management*, *310*, 1057–1070. https://doi.org/10.1016/j.foreco.2013.06.052

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261–1277. https://doi.org/10.1111/ecog.04960

Zurell, D., König, C., Malchow, A., Kapitza, S., Bocedi, G., Travis, J., & Fandos, G. (2022). Spatially explicit models for decision-making in animal conservation and restoration. *Ecography*, *2022*(4), ecog.05787. https://doi.org/10.1111/ecog.05787

SUPPORTING INFORMATION

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

Table S1.1. ODMAP protocol for the study (sensu Zurell et al., 2020).

Table S1.2. List of R packages used in the analysis with citations.

Table S1.3. List of the 37 species used in model calibration, validation and prediction with their code, common and scientific name, family and order, migratory habit and information about models fitting (LTE = long-term ensemble, RY = random year, TEMP_OCC = temporal occupancy, DYNA = long-term dynamic). Taxonomic information derived from BirdLife International (2020).

Table S1.4. List of environmental variables used in the models with variable group (either microclimate or vegetation structure) and data source, unit and brief description and ecological meaning.

Table S1.5. List of default settings of BRTs for dismo and dynamicSDM packages used in the calibration.

Table S1.6. List and description of the performance metrics used in the study grouped by metric type (accuracy or calibration) and describing the range of values and a summary of their description and interpretation.

Figure S1.1. Principal component (PC) analysis on birds life history traits. The first PC axis was negatively associated with several body lengths measurements and body mass and explains the 73.3% of the total variance. The second PC axis was related to species movement and dispersal capacity (migratory habit and hand-wing index, HWI) and shows no significance according to null distributions built through random permutations. Therefore, we only used PC1 and interpreted it as a proxy of body size. Copyright-free images were downloaded from https://pixabay.com/.

Figure S1.2. Spatiotemporal validation strategy used for assessing model performance. (A) Temporal component: each fold holds out 1 year for testing (red), while training occurs on all remaining years (blue); in the example, the model for the target species was performed over the entire timespan (10 years: 2010–2019). This procedure iterates across all years to evaluate model generalizability in time. (B) Spatial component: 70% of sites were used for training (blue), and 30% of spatially independent sites were reserved as a test set (red). This design ensured both spatial and temporal independence of the test data.

Table S2.1. Drop-off of the different validation metrics grouped by migratory habit calculated as the difference in performance between long-term dynamic and random year models for the main metrics.

Table S2.2. Results of linear mixed models and post-hoc comparisons on the most important performance metrics and migratory habit. Letters after the migratory habit indicate significant differences for each performance metric among the migratory habits, while letters in the rows indicate significant differences among the frameworks for each performance metric. Letters in the cells indicate statistically significant differences among the modelling frameworks within the migratory habits for each performance metric. Letters represent pairwise comparisons with a confidence level of 0.95 (alpha = 0.05).

Table S2.3. Results of linear mixed models and post-hoc comparisons on performance metrics and body size (PC1). Letters in the rows indicate significant differences among the frameworks for each performance metric. Letters represent pairwise comparisons with a confidence level of 0.95 (alpha = 0.05).

Table S2.4. Results of linear mixed models and post-hoc comparisons on Spearman's rank correlation and migratory habit. Letters indicate significant differences among the frameworks and represent pairwise comparisons with a confidence level of 0.95 (alpha = 0.05).

Table S2.5. Fixed effects of the linear mixed effect models assessing the relation between Spearman's rank correlation and body size.

Figure S2.1. Results of temporal leave-one-year-out cross-validation of the different models according to the four modelling frameworks and three different movement behaviours (resident, partially migratory and migratory) of birds. Violin plots of (A) TSS, (B) *F*1 score, (C) Cohen's kappa, (D) sensitivity, (E) specificity, (F) Pearson's *r*, (G) Brier score and (H) log likelihood. Similarly to Figure 4 (AUC and unreliability index), we observed gradients of improvement in performance as a result of increases in the temporal complexity (dynamicity) of the models.

Figure S2.2. Trade-offs and performance patterns of species distribution modelling frameworks across accuracy, discrimination and calibration metrics retrieved fromfrom our the spatiotemporal validation in which random years and 30% of locations were held out as test data. Principal component analysis (PCA) biplot showing the distribution of the four different modelling frameworks based on 12 different performance metrics. on the different performance metrics. PC1 (35.8% of variation explained) mostly captures variation in accuracy and discrimination, while PC2 (22.1%) is more correlated to calibration quality. Arrows represent the contribution and direction of each metric; ellipses represent 95% confidence regions for each framework. A clear gradient is visible from the most static model (ry; random year) to the most complex (dyna; long-term dynamics) in terms of increasing accuracy and discrimination along PC1 (*x*-axis), while the pattern for calibration along PC2 (*y*-axis) is more nuanced.

Figure S2.3. Variable importance according to migratory habit (rows: resident, partially migratory and migratory) and modelling framework (columns: random year, temporal occupancy, long-term ensemble and long-term dynamic models). Variables were grouped into categories: CDD (cooling degree days) and GDD (growing degree days), Tmean = monthly or seasonal mean temperatures, tvar = standard deviation of temperature, Tmax = monthly or seasonal maximum temperatures, Tmin = monthly or seasonal minimum temperatures, vegetation = LiDAR-derived vegetation variables.

Figure S2.4. Spearman's rank correlations testing for consistency in spatial predictions of probability of occurrence among modelling frameworks based on body size (PC1) (upper right corner) and migratory habit (lower left corner). Letters indicate post-hoc results on linear mixed-effect models and indicate significant differences among the pairs of frameworks. Migratory species had the greatest degree of similarity between modelling frameworks. Long-term ensemble framework was the most similar to all the other models.

Figure S2.5. Spatial predictions of probability of occurrence for (A) American robin (AMRO), (B) black-headed grosbeak (BHGR), (C) brown creeper (BRCR), (D) black-throated grey warbler (BTYW) and (E) chestnut-backed chickadee (CBCH) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.6. Spatial predictions of probability of occurrence for (A) common raven (CORA), (B) dark-eyed junco (DEJU), (C) evening grosbeak (EVGR), (D) golden-crowned kinglet (GCKI) and (E) Canada jay (CAJA) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.7. Spatial predictions of probability of occurrence for (A) hammond's flycatcher (HAFL), (B) hermit thrush (HETH), (C) hermit warbler (HEWA), (D) Hutton's vireo (HUVI) and (E) MacGillvray's warbler (MGWA) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.8. Spatial predictions of probability of occurrence for (A) Northern flicker (NOFL), (B) olive-sided flycatcher (OSFL), (C) osprey (OSPR), (D) Pacific wren (PAWR) and (E) pine siskin (PISI) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.9. Spatial predictions of probability of occurrence for (A) pileated woodpecker (PIWO), (B) Pacific-slope flycatcher (PSFL), (C) red-breasted nuthatch (RBNU), (D) ruffed grouse (RUGR) and (E) rufous hummingbird (RUHU) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.10. Spatial predictions of probability of occurrence for (A) sooty grouse (SOGR), (B) Steller's jay (STJA), (C) Swainson's thrush (SWTH), (D) Townsend's solitaire (TOSO) and (E) Vaux's swift (VASW) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.11. Spatial predictions of probability of occurrence for (A) varied thrush (VATH), (B) Warbling vireo (WAVI), (C) Western tanager (WETA) and (D) yellow-rumped warbler (YRWA) according to the four modelling frameworks. The colour scale is different for each species to improve visualization and comparison between frameworks.

Figure S2.12. Principal component analysis (PCA) summarizing the results of performance and correlations in spatial predictions for all the species grouped by migratory habit. (A) Results of PC1 versus PC2, (B) results of PC1 versus PC3. The full shapes with bigger size represent the centroids of the three migratory habits resident, partial migratory and migratory.

How to cite this article: Anselmetto, N., Garbarino, M., Weldy, M. J., Bell, D. M., Daly, C., Epps, C. W., Ferrari, N., Kim, H., LaManna, J. A., Lesmeister, D. B., Penaluna, B. E., Schulze, M., Sutton, M., Tosa, M. I., & Betts, M. G. (2025). Leveraging long-term data to improve biodiversity monitoring with species distribution models. *Journal of Applied Ecology*, *00*, 1–16. https://doi.org/10.1111/1365-2664.70177