

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Piirainen, Sirke; Lehtikoinen, Aleks; Husby, Magne; Kålås, John Atle; Lindström, Åke; Ovaskainen, Otso

Title: Species distributions models may predict accurately future distributions but poorly how distributions change : A critical perspective on model validation

Year: 2023

Version: Published version

Copyright: © 2023 The Authors. Diversity and Distributions published by John Wiley & Sons Ltd

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>


Please cite the original version:

Piirainen, S., Lehtikoinen, A., Husby, M., Kålås, J. A., Lindström, Å., & Ovaskainen, O. (2023). Species distributions models may predict accurately future distributions but poorly how distributions change : A critical perspective on model validation. *Diversity and Distributions*, 29(5), 654-665. <https://doi.org/10.1111/ddi.13687>

RESEARCH ARTICLE

Diversity and Distributions WILEY

Species distributions models may predict accurately future distributions but poorly how distributions change: A critical perspective on model validation

Sirke Piirainen^{1,2}  | Aleksi Lehikoinen¹  | Magne Husby^{3,4} | John Atle Kålås⁵ | Åke Lindström⁶ | Otso Ovaskainen^{7,8,9}

¹Zoological Unit, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

²Arctic Centre, University of Lapland, Rovaniemi, Finland

³Section of Science, Nord University, Levanger, Norway

⁴BirdLife Norway, Trondheim, Norway

⁵Norwegian Institute for Nature Research, Trondheim, Norway

⁶Department of Biology, Lund University, Lund, Sweden

⁷Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

⁸Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland

⁹Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence

Sirke Piirainen, Zoological Unit, Finnish Museum of Natural History, University of Helsinki, P. O. Box 17, Helsinki 00014, Finland.

Email: sirke.piirainen@helsinki.fi

Funding information

Academy of Finland, Grant/Award Number: 275606, 309581 and 326338; BiodivScen ERA-Net-COFUND programme; H2020 European Research Council, Grant/Award Number: grant agreement No 856506; ERC-synergy project LIF; Jane ja Aatos Erkon Säätiö; Koneen Säätiö, Grant/Award Number: 201903886; National Science Foundation, Grant/Award Number: CLO, ICER-1927646; Norges Forskningsråd, Grant/Award Number: Centres of Excellence Funding Scheme (223257) and NINA: 295767; Suomen Kulttuurirahasto; Svenska Forskningsrådet Formas, Grant/Award Number: 2018-02441

Editor: Boris Leroy

Abstract

Aim: Species distribution models (SDMs) are widely used to make predictions on how species distributions may change as a response to climatic change. To assess the reliability of those predictions, they need to be critically validated with respect to what they are used for. While ecologists are typically interested in how and where distributions will change, we argue that SDMs have seldom been evaluated in terms of their capacity to predict such change. Instead, typical retrospective validation methods estimate model's ability to predict to only one static time in future. Here, we apply two validation methods, one that predicts and evaluates a static pattern, while the other measures change and compare their estimates of predictive performance.

Location: Fennoscandia.

Methods: We applied a joint SDM to model the distributions of 120 bird species in four model validation settings. We trained models with a dataset from 1975 to 1999 and predicted species' future occurrence and abundance in two ways: for one static time period (2013–2016, 'static validation') and for a change between two time periods (difference between 1996–1999 and 2013–2016, 'change validation'). We then measured predictive performance using correlation between predicted and observed values. We also related predictive performance to species traits.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

Results: Even though static validation method evaluated predictive performance as good, change method indicated very poor performance. Predictive performance was not strongly related to any trait.

Main Conclusions: Static validation method might overestimate predictive performance by not revealing the model's inability to predict change events. If species' distributions remain mostly stable, then even an unfit model can predict the near future well due to temporal autocorrelation. We urge caution when working with forecasts of changes in spatial patterns of species occupancy or abundance, even for SDMs that are based on time series datasets unless they are critically validated for forecasting such change.

KEYWORDS

birds, climate change, Fennoscandia, forecasting, land use, model validation, prediction, species distribution modelling, species traits, temporal transferability

1 | INTRODUCTION

Understanding how species will respond to climate change is one of the current key challenges in ecology and nature conservation (Malhi et al., 2020; Sofaer et al., 2018). To gain such understanding, various types of species distribution models (hereafter SDMs) are widely used to make predictions on where the species occurs now, and projections on how species distributions are likely to change as a response to predicted climatic change (Huntley et al., 2007; Leathwick et al., 1996; Massimino et al., 2017). Thus, SDMs are important tools for ecologists and can have large implications for management and conservation (Kujala et al., 2013). SDMs can for example help to identify conservation needs, to define alternative conservation actions and to evaluate the effects of such actions (Guisan et al., 2013).

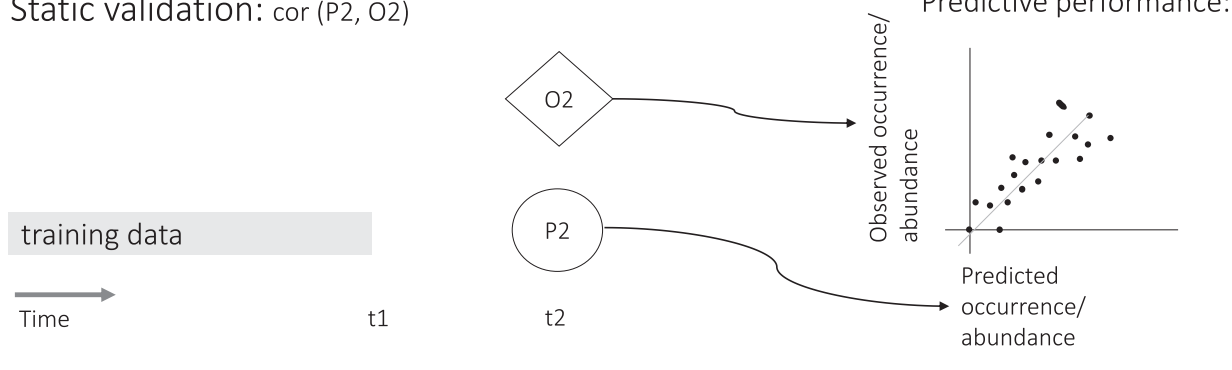
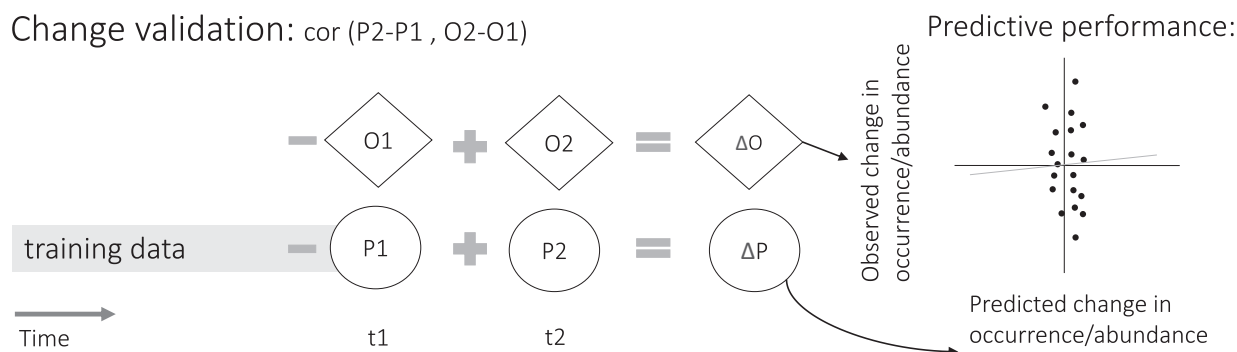
The utility of predictions clearly depends on how much they can be trusted, and to estimate reliability the predictions need to be critically validated with respect to what they are used for. In this paper, we seek to simulate a situation where an SDM is intended to be used in projecting community changes to a future time period approximately 20 years from the current moment. As Vaughan and Ormerod (2005) claim, the ideal way to measure a model's generalizability is to test the model by incorporating situations that resemble its future applications as closely as possible. We take this idea as read that if a model is intended to predict, for example 20 years ahead in future, then its validation should reflect this pattern by using data from the past to make predictions into present (a 'future' time period in relation to model fitting data), which lies approximately a similar length of time ahead. This means ignoring information from the latest part of the data for model fitting, but simulates the real-life data gap that an ecologist making projections into a more distant future is always faced with.

Here, we argue that while ecologists are typically interested in how species distributions will change in a changing climate, SDMs have seldom been critically evaluated in terms of their capacity to

predict such change. This has led to a situation where our ability to make reliable range change forecasts may often be overestimated.

Even though the benefits of using temporally independent data for testing SDM's predictive performance have been documented already decades ago (Araújo et al., 2005), still a typical SDM used for forecasting is not validated by means of such data (Uribe-Rivera et al., 2022). Typically, in those cases when temporally independent data exist and there has been an attempt to validate the model by means of retrospective forecasting, this has been done by splitting the dataset into two temporal blocks: the earliest records of the dataset form a training data block and the latest a test data block. There might or might not be a temporal gap between the training and testing datasets. Depending on the length of the gap, the test data can be seen as temporally independent data with respect to the training data. The model is then calibrated with training data and predictions are made to one static point in time (either to data from 1 year only or to a mean of multiple years). Lastly, these predictions are compared with observations ('static validation method' in Figure 1), and the predictive performance of the model is considered high if the predictions match with the observations. Several authors have used temporally independent static validation, for example on butterflies (Eskildsen et al., 2013), plants (Dobrowski et al., 2011) and birds (Araújo et al., 2005; Regos et al., 2019).

However, we argue that if the aim was to evaluate how well a model can predict changes in the patterns of occupancy or abundance of the species, then using a static validation method that only considers the spatial distribution of occupancy or abundance (predicts and validates a single static point in time) is not adequate. While the static method is able to evaluate how well the model predicts a future distribution on a general level, it does not evaluate whether or not the model is able to predict if and how the species distribution changes. This is because the species ranges may remain mostly stable and thus there can be a high correlation between locations where the species has occurred in the past and where it will occur in the near future (Rapacciuolo et al., 2014).

Static validation: $\text{cor}(P2, O2)$ Change validation: $\text{cor}(P2-P1, O2-O1)$ 

P1/2, prediction to period 1 or 2; O1/2, observation from period 1 or 2; $\Delta O/P$, observed or predicted change between time periods 1 and 2

FIGURE 1 Schematic picture of the two validation methods compared in the study. The evaluated model uses data from 1975 to 1999 for model training (indicated by the grey bar). Static validation method uses the model to predict a static pattern, species occurrence or abundance in only one, independent time period t2 (mean of 2013–2016, indicated by a circle P2). Predictive performance is calculated as the correlation between the prediction and the corresponding observed value (diamond O2) over study locations. Change method predicts both time periods t1 (1996–1999, circle P1) and t2 (circle P2) and then counts the predicted change (ΔP), that is difference in species occurrence or abundance between these two separate time periods, and compares it to the corresponding observed change (ΔO).

So even if the model fails to predict change events, which usually happen especially at the range margins, it may still predict a major part of the unchanged distribution correctly and thus achieve an overoptimistic evaluation of predictive performance when only a static validation method is applied (Rapacciuolo et al., 2012; Sofaer et al., 2018).

As a solution on how to effectively test the performance of SDMs intended for predicting biodiversity changes in future, we propose that the model validation method should be explicitly based on measuring and validating a change over a given time period. By 'change', we mean the difference in a population measure between two points in time, that is increase, decrease or stability in occurrence probability or abundance. Ideally, the time period over which a change is calculated should be independent in relation to the model training data ('change validation method' in Figure 1).

We base our change validation method on the work of Rapacciuolo et al. (2014) where the authors developed a temporal validation plot to visualize and measure the agreement between predicted and observed changes and applied it to two bird species from the Great Britain and a virtual species. The temporal validation plots solved the issue of generally static distributions by focussing the

assessment of model performance only on locations where range change events (a grid cell either gaining or losing occupancy) over time were either observed or predicted. By focussing on change events only, the authors were able to reveal aspects of the relationship between species' range change and model variables that might not have been identified through range-wide measures (including also grid cells with stable presences or absences).

While the need to validate the predicted change rather than validating the predicted future may sound self-evident, to our knowledge our work is one of only a few papers to apply such an approach in the context of climate change predictions (Johnston et al., 2013; Rapacciuolo et al., 2014). A greater proportion of studies have investigated models' ability to predict to new geographical areas by applying spatial validation methods (Bahn & McGill, 2013; Charney et al., 2021; Journé et al., 2020; Rousseau & Betts, 2022). While there is a certain analogy between spatial and temporal block cross-validation through the space-for-time substitution in distribution modelling (Blois et al., 2013), we, however, think that spatial validation methods are not best suited for testing a model, which is not intended to predict to new areas, but to a new time, and therefore, results obtained from spatial validation studies do not directly apply

in the context of temporal validation and making predictions to future time.

In multispecies studies, there is usually a high level of variability in predictive performance among the species (Araújo et al., 2005; Dobrowski et al., 2011; Sofaer et al., 2018; Venne & Currie, 2021), meaning that some species are easier to model than others. One reason for this variation can be ecological differences between species (Tessarolo et al., 2021). For instance, species traits such as range size, migratory behaviour, rarity and body size have been linked with model performance in birds (McPherson & Jetz, 2007). Studying the connection between species' traits and predictive performance can give us an understanding on how those traits affect modelling and how to interpret results.

In this study, we compared our 'change validation method' with the more routinely used 'static validation method' using large-scale monitoring data on 120 bird species from Finland, Sweden and Norway. We studied whether predictive performance differs between these two validation methods and whether the variation in predictive performance is linked with certain species' traits. Concerning the influence of species traits, we expected that common, large-sized, nontropical migrant species which prefer naturally patchy habitats show higher predictive performance than rare, small-sized, long-distance migratory species, which prefer broad and structurally more varied habitats such as forest, or are more influenced

by factors outside their breeding range (Johnston et al., 2014; Laaksonen & Lehikoinen, 2013; McPherson & Jetz, 2007).

2 | METHODS

2.1 | Bird data

We gathered data from national bird monitoring surveys performed in North Europe: Finland, Sweden and Norway (Lehikoinen et al., 2014; Lindström et al., 2015) covering circa 1900 km latitudinal gradient. In total the data consist of 2591 bird survey sites (Norway 486, Sweden 1197 and Finland 908, Figure 2). Data come from line transects (Finland and Sweden) and/or point count routes (Norway and Sweden), which cover the countries systematically at approximately 20–40 km distance of each other, as well as from more randomly located surveys. Survey sites have been visited once per year during the start of the breeding season, but not necessarily each year. All the data have been gathered by volunteers during 1975–2016. The earliest observations are from 1975, 1978 and 1991 from Sweden, Finland and Norway, respectively. In each survey, abundances of observed birds were recorded either as pairs (Finland and Norway) or as individuals (Sweden). We approximated that these measures correspond because in the Finnish and Norwegian survey

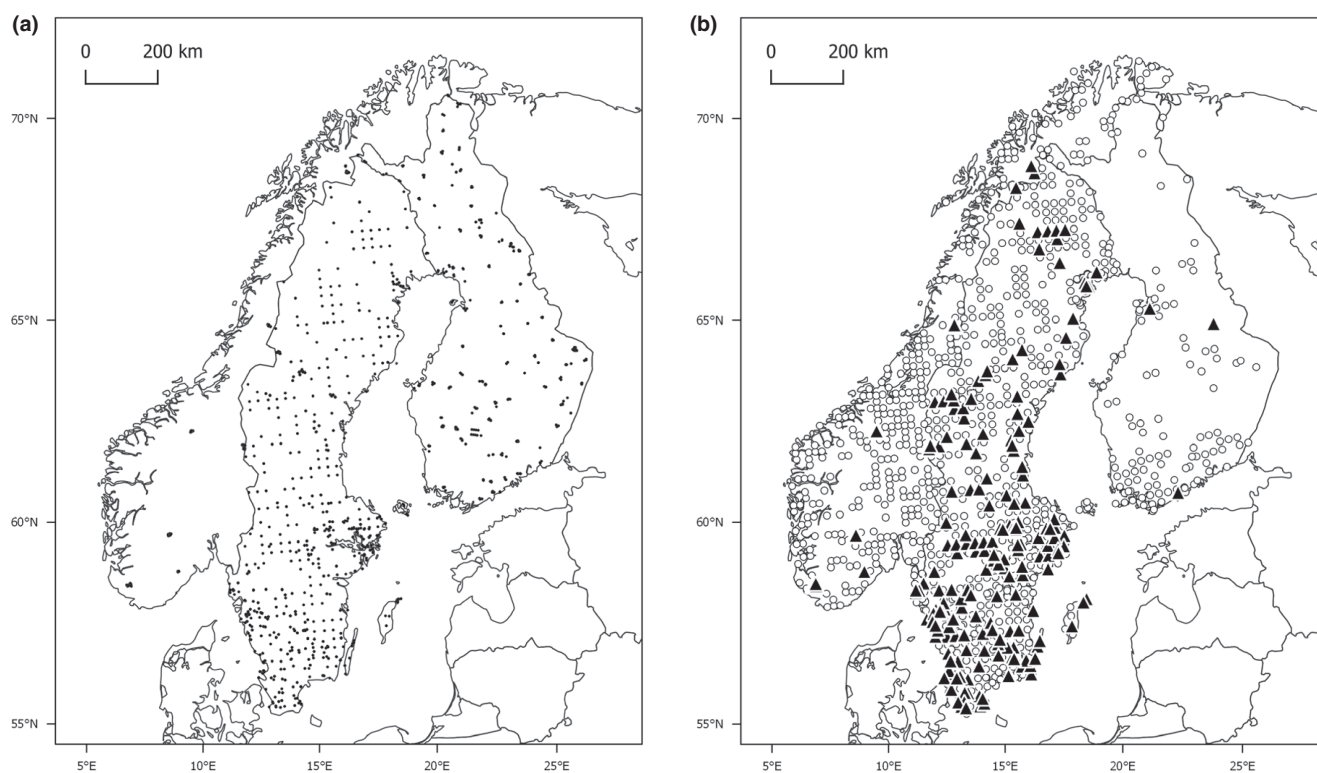


FIGURE 2 Bird survey sites in Norway, Sweden and Finland. (a) Survey sites which were visited between 1975 and 1999 and thus contribute to model training data. (b) Survey sites which were visited at least twice in each of the time periods t1 (1996–1999) and t2 (2013–2016) (which contribute to change validation method for occurrence) are indicated with filled triangles ($n = 264$). Survey sites which were visited at least twice in time period t2 (which contribute to static validation method for occurrence) are indicated with both filled triangles and open circles ($n = 1214$).

methods, observations of solitary males (the by far most common records) and females are in any case, interpreted as pairs. In total, 264 bird species were recorded. In Finland, waterbirds were included in the surveys from 2006 onwards. For each survey, a method (point or line) and effort (number of points belonging to a point count survey or length of line in metres for a line transect survey) of survey was recorded, as well as the starting point or centroid coordinates of the survey site.

Traits of each bird species (prevalence, body mass, migratory behaviour and habitat preference) are listed in Table S1.2: Appendix S1. Source data for species traits and more precise hypotheses concerning species traits are presented in Table S1.1: Appendix S1

2.2 | Environmental variables

To describe the habitat where birds were observed, we drew a 300-m-wide buffer around each line transect and group of points in point count route and determined the land cover inside the buffer from the 100-m resolution Corine land cover data (© European Union, Copernicus Land Monitoring Service 2022, European Environment Agency, EEA). In our study area, the Corine land cover survey had been conducted in the years 2000, 2006 and 2012. We extrapolated information from this data to cover the years 1975–2003, 2004–2009 and 2010–2016, respectively. We acknowledge that in the earliest period there is a maximum of 25-years of mismatch between land cover and bird data. As the long time series from the 1970s provides a strong calibration dataset for the future predictions, and as the land cover variables have changed less than the climatic variables that are of our primary interest, we accepted this mismatch. We classified land cover types into six classes and calculated their proportions (value range 0–1) on the buffer. The classes were urban and agricultural areas (Urb), broadleaved and mixed forests (Br), coniferous forests (Co), mountainous areas and shrublands (Op), marine water and beaches (Ma) and wetland, peatland and inland water areas (We). For climatic variables, we chose only mean daily temperature measurements, since previous studies have shown that temperature has much larger impact on North European bird populations compared with, for example rainfall (Tayleur et al., 2015). For each survey site, we determined mean temperatures of the current year spring (April and May), previous winter (December–February) and previous summer (June and July of the preceding year) from E-OBS weather data (version 15, 0.25-degree resolution) by the European Climate Assessment & Dataset project, ECA&D (Cornes et al., 2018). We selected these weather parameters because they reflect the conditions which the birds and their breeding environment have experienced by the time of the observation, after the past summer and winter and approximately at the start of the breeding season (Pearce-Higgins et al., 2015). Earlier studies have shown that climate variables, particularly temperature, tend to be more important in explaining abundances of bird species in Northern Europe compared with land use variables (Howard

et al., 2015), which supports our goal to study especially the impacts of changing climate.

2.3 | Species distribution models

In the analysis, we used the Hierarchical Modelling of Species Communities approach (HMSC; Ovaskainen et al., 2017), which belongs to the class of hierarchical Bayesian joint species distribution models (JSDMs; Warton et al., 2015). We considered each visit to each survey site (line transect or point count survey) as one sampling unit. We included in the HMSC analyses data on species abundances (matrix **Y** of sampling units times species) and environmental covariates (matrix **X** of sampling units times the included covariates).

As the response variable, that is components of the matrix **Y**, we considered the abundance of species, that is the vector of species counts for each sampling unit. To account for the zero-inflated nature of the data, we applied a hurdle approach, in which we first modelled species presence/absence data with a probit regression and then modelled abundance conditional on presence with a log-normal regression. We included all species that had the prevalence (fraction of surveys in which the species was observed) of at least 0.05, which limited the data to 120 species.

As fixed effects, as components of the matrix **X**, we included the mean (1) winter and (2) spring temperatures of the present year, (3) the mean summer temperature of the previous year, (4) an indicator variable describing the survey method and (5) survey effort (log-transformed), and (6–11) proportions of habitat classes within the 300-m buffer.

We fitted the HMSC model with the R-package Hmsc (Tikhonov et al., 2020) assuming the default prior distributions (Ovaskainen & Abrego, 2020, Chapter 8). We sampled the posterior distribution with four Markov Chain Monte Carlo (MCMC) chains, each of which was run for 1,500,000 iterations, of which the first 500,000 were removed as burn-in. The chains were thinned by 1000 to yield 1000 posterior samples per chain and so 4000 posterior samples in total. We examined MCMC convergence by examining the potential scale reduction factors (Gelman & Rubin, 1992) of the model parameters that measure species responses to the covariates included in the **X** matrix.

Since differences between modelling approaches are typically less impactful than other factors (especially modelling strategy, included covariates, and the study system) when evaluating temporal predictive performance (Rapacciuolo et al., 2012), we use a single model, instead of using consensus modelling approaches which put more emphasis on precision rather than accuracy (Dobrowski et al., 2011). We selected to use specifically HMSC because it performed generally the best in terms of predictive power in a recent comparison among single-species and JSDMs (Norberg et al., 2019).

One of the benefits of this multispecies approach is that when the model is estimating the environmental responses of each species, through its hierarchical structure it can borrow information across all species (Ovaskainen & Abrego, 2020). This is likely to increase

predictive performance, especially for rare species for which the data alone is not sufficiently informative for accurate parameterization. Accordingly, Norberg et al. (2019) showed that HMSC models that included such feature had better predictive performance compared with corresponding single-species models.

We report our data and modelling process according to the ODMAP protocol (Zurell et al., 2020) in Appendix S2. We emphasize that the details regarding model algorithms were secondary in this study as the focus was on the comparison of validation methods.

2.4 | Model validation methods

To test the ability of the model to make predictions for the future, we first parameterized the model using data until the year 1999 only (3486 visits to 901 unique routes). We then used the model to make predictions y^P for each survey site (see below for how the predictions were generated for the two components of the hurdle model), both for the period 1996–1999 (denoted by the vector y_{1999}^P) and for the period 2013–2016 (denoted by y_{2016}^P). Then, we computed analogous variables from the raw data to record the observed y^O (denoted by vectors y_{1999}^O and y_{2016}^O).

For the static validation method, we measured the model's ability to predict the observed distribution for each focal species as the Pearson correlation between y_{2016}^P and y_{2016}^O over the survey sites. For the change validation method, we computed the predicted change (denoted by the vector c^P) for each survey site as the difference between the predictions for the two time periods, that is $c^P = y_{2016}^P - y_{1999}^P$. To record the observed change (denoted by the vector c^O), we computed the difference between the true observations, that is $c^O = y_{2016}^O - y_{1999}^O$. We then computed the model's predictive performance for each species as the correlation between c^O and c^P .

To evaluate the presence-absence part of the hurdle model, we used as the measure y the probability of occurrence, that is proportion of years in which the survey site was occupied. For evaluating the abundance part of the hurdle model, we used the mean of log-transformed counts over the years. Predicted mean abundances were counted only for years corresponding to species' presence.

We chose correlation as our measure of predictive performance because we were interested in identifying locations or environmental conditions which will influence the species especially favourably or unfavourably, that is on measures of discrimination. Evaluating predictive performance in terms of accuracy, calibration and precision would clearly be relevant as well (Norberg et al., 2019), but for the above reason we considered discrimination to be of the highest relevance and focussed on that.

For occurrence data, we chose only survey sites that had been visited at least twice, either during only the latter time period (static method) or during both time periods (change method). This resulted in 1214 and 264 sites, respectively. For abundance, the data were further restricted to only those survey sites where the species was observed at least once. This resulted on average 422 sites per

species (range 67–1182) for the latter time period (static method) and 83 sites (range 3–262) for both time periods (change method). Despite the smaller sample size and geographical coverage of validation data for the change method (Figure 2), the validation data still cover both the northern range edge and the core breeding areas for over 90% of study species (Keller et al., 2020). Species with observations from <10 sites were excluded from the validation data. Thus, in the abundance data for the change method, hazel grouse (*Bonasa bonasia*) and rustic bunting (*Emberiza rustica*) were excluded.

To quantify the overall amount of change per species, we calculated species-specific mean observed and predicted changes from across all the study sites, which contributed to the change validation datasets ($n = 10$ –264 depending on species and type of data). To ask whether the model was able to predict which species increased and which decreased, we counted a correlation coefficient between the mean predicted and observed changes over the species.

2.5 | Effect of validation method

To investigate the differences in predictive performance between the two validation methods, we fitted a linear mixed-effects model, separately for occurrence and abundance datasets. For both occurrence and abundance data, we used the species-specific Pearson's correlation coefficient, that is predictive performance, as the response variable. As a fixed effect, we entered the validation method (categorical variable: static or change). Species was included as a random effect and species' phylogeny was modelled in the variance structure to account for if closely related species showed similar predictive performance. Phylogeny tree was acquired from birdtree.org (Jetz et al., 2012). The modelling was conducted using the R-package and function MCMCglmm (Hadfield & Nakagawa, 2010) with Gaussian error distribution using 1,003,000 iterations, where first 3000 were used for 'burning in' and thinning interval was 1000.

2.6 | Effect of species traits

To investigate how the predictive performance for each validation method was related to species prevalence (continuous variable), migration strategy (long or short migration or resident), body mass (continuous) and habitat preference (forests, wetlands, cultural environments, mountains and mires), we fitted a set of 17 competing models. Prevalence and body mass variables were log-transformed prior to analysis to reduce non-normality. Our full model included each species trait and its two-way interaction with validation method as fixed effects. The remaining 16 models included 14 models with reduced combinations of the two-way interactions between method and traits, one model with only the main effects of all variables and one model with only the main effect of validation method (all models are listed in Table S1.6: Appendix S1). Species was included as a random effect and species' phylogeny was modelled in the variance structure.

The models were fitted using the function MCMCglmm (Hadfield & Nakagawa, 2010) with Gaussian error distribution (1,003,000 iterations, 3000 warm-up iterations, thinning interval of 1000). For model selection, we used the multimodel inference approach (Burnham & Anderson, 2002) by calculating the corrected deviance information criterion (DIC) for each model and selecting the one with the lowest DIC (Spiegelhalter et al., 2002).

2.7 | Robustness test

We assessed the sensitivity of our results to the effect of using a single species instead of JSDM. To do so, we considered otherwise identical HMSC models as in the case of joint species distribution modelling but fitted the models separately for each of the 120 study species. We then calculated the same predictive performance metrics with static and change methods for each species. To investigate the effect of modelling approach (single vs. joint SDM), we fitted a linear mixed-effects model where both validation method and modelling approach were used as fixed effects to explain predictive performance. Species was included as a random effect and species' phylogeny was modelled in the variance structure. The model was fitted with the MCMCglmm -function (Hadfield & Nakagawa, 2010) with Gaussian error distribution (1,003,000 iterations, 3000 warm-up iterations, thinning interval of 1000).

3 | RESULTS

Both HMSC models, for presence/absence data and for abundance data, reached good MCMC convergence and thus their posterior distributions were adequately sampled (Table S1.3: Appendix S1). Species occurrences and abundances were more often statistically

connected with climatic variables rather than land use variables (Figures S1.1 and S1.2: Appendix S1).

The overall mean species-specific observed and predicted changes across all study sites varied almost symmetrically around zero, thus some species declining and some increasing in their overall distribution and abundance. The observed mean changes varied between -0.11 and 0.15 in the occurrence probability and between -0.74 and 0.95 in the log-transformed abundance. The predicted mean changes varied much less, between -0.05 and 0.10 in the occurrence probability, and between -0.30 and 0.25 in the log-transformed abundance (Figures S1.3 and S1.5: Appendix S1). For most species, observed changes were more negative (less increase or greater decrease) than what was predicted (Figures S1.4 and S1.6: Appendix S1). The correlation coefficient between mean observed and mean predicted changes was 0.21 ($p = .023$) for occurrence data and 0.25 ($p = .0056$) for abundance data (Figures S1.3 and S1.5: Appendix S1), meaning that the model predicted better than by random which species declined and which increased in general.

3.1 | Predictive performances

For occurrence data, the mean correlation coefficient for the static validation method was high, 0.54 (SD 0.18), whereas for the change method it was very low, 0.024 (SD 0.13 ; Figure 3, species-specific correlations in Table S1.2). For abundance data, the corresponding mean correlations were 0.31 (SD 0.19) for the static method and 0.003 (SD 0.18) for the change method (Figure 3, Table S1.2: Appendix S1). Thus, for both occurrence and abundance data, the static validation method resulted in much higher predictive performance compared with change method. When focussing on the static validation method, the mean predictive performances were higher for occurrence compared with abundance data, but for the change method no such difference was detected (Figure 3).

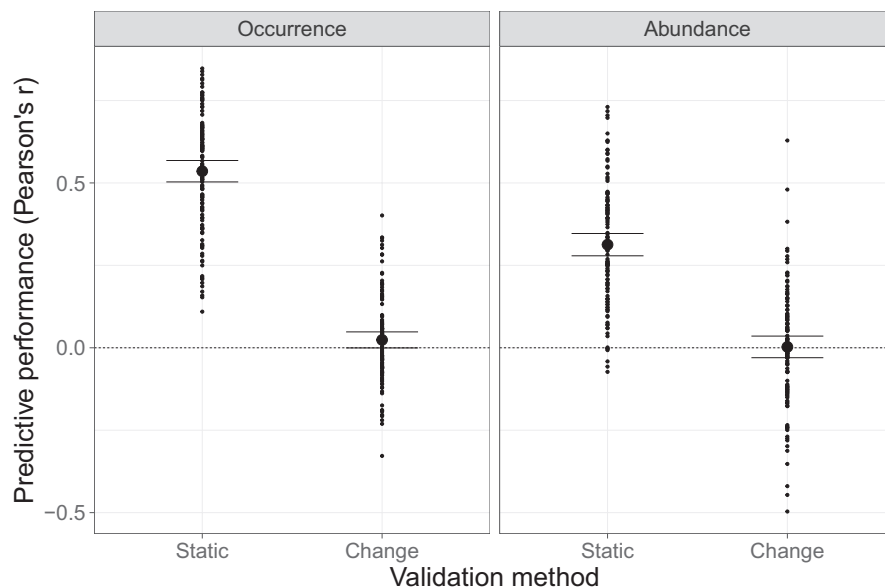


FIGURE 3 Predictive performance, measured as the correlation between observed and predicted values, for each species ($n = 118$ – 120). The two panels correspond to occurrence and abundance analyses, and for each panel, results using static and change validation methods are shown. The dot and the error bar show the overall mean for all species and its 95% confidence interval.

The posterior mean estimates for the effect of validation method were -0.51 , 95% CI $[-0.55, -0.48]$ for occurrence data and -0.31 , 95% CI $[-0.35, -0.27]$ for abundance data (Tables S1.4 and S1.5: Appendix S1).

3.2 | Species traits

Habitat preference and prevalence were the most important traits affecting predictive performance in occurrence data (Table S1.6: Appendix S1). However, these effects were strongly moderated by the validation method used, as evidenced by the significant interactions between the terms (Table S1.7: Appendix S1). Species which prefer cultural (urban and agricultural areas) and mountain and mire habitats were associated with higher predictive performance but only with the static method, while no such connection was detected with the change method (Figure S1.7: Appendix S1). Higher prevalence was also associated with higher predictive performance with the static method, whereas with the change method the positive association was very weak (Figure S1.8: Appendix S1).

In abundance data, habitat preference, prevalence and migration behaviour were the most important traits affecting predictive performance (Table S1.8: Appendix S1). More common, short-migrating species preferring cultural habitats were associated with higher performances when conducting validation with the static method (Table S1.9, Figures S1.9–S1.11: Appendix S1). However, again these connections were much weaker or nonexistent with the change method.

In summary, the effects of species traits were rather weak and less visible for change method, which suffers from low predictive performance in general.

3.3 | Single versus joint species distribution modelling

When comparing the results from single SDMs to joint SDMs, for static method the predictive performances were much lower (but still higher than change method), whereas for the change method the performances remained near zero (Table S1.10, Figure S1.12: Appendix S1). The model results showed that the single-species modelling approach had a negative effect on predictive performance. The posterior mean estimate for the effect of modelling approach was -0.13 , 95% CI $[-0.15, -0.10]$ (Table S1.11: Appendix S1). In summary, our robustness test indicated that the use of joint instead of single SDM does not influence the relative order of estimates of model performances between static and change validation methods.

4 | DISCUSSION

Our study aimed at developing the concept of SDM validation by comparing two different validation methods for assessing temporal

predictive performance, one that measures and predicts a static distribution in a population, that is only 1 year or mean of several years, and one that measures a change, that is difference in distribution between two separate time periods.

4.1 | Static validation ignores change events

Our results showed a major plummet in the estimate of predictive performance when change validation was applied, compared with the static method. This means that the interpretation of model performance changes drastically depending on which type of validation is applied, static or change. This highlights that, even if validation is done as recommended, with temporally independent data (Harris et al., 2018), the estimates of model's predictive performance might still be overly optimistic if evaluating future static distribution instead of change in distribution.

The main reason for this difference lies behind the fact that species ranges are often rather static, expanding or retracting mainly at their edges. This stability increases the possibility that, even as time passes, a species is still present, and in somewhat same quantities, in most parts of its previous range. Spurious species–environment associations during model calibration may not be detected by a static performance metric if the temporal validation compares only mostly unchanged areas. Our model, on average, predicted only very moderate changes across most of the species' range. It therefore failed especially at those locations where more pronounced changes had happened, but at the same time succeeded in correctly predicting little change in most locations where distributions remained rather stable. Several studies have found similar results where static temporal validation has shown good estimates but closer inspection of occurrences at sites where changes have happened, usually at the species' range margins, has revealed that predictions of change events have failed (Eskildsen et al., 2013; Rapacciuolo et al., 2012; Sofaer et al., 2018).

4.2 | Reasons for failing to predict change events

Why we failed to predict change could be that there were only minor changes happening in either bird distributions or in land cover or climate variables in the study area within the forecasted time period (approximately 16 years). The smaller the variation in the calibration data, the harder it will be for the model to estimate the underlying species–environment associations and thus to predict the spatial pattern of change events correctly. However, our bird data and climate records from Fennoscandia show that poleward shifts in both species' abundance (1.5 km/year) and climate (temperature, 7.4 km/year) have occurred in our study area (Lehikoinen & Virkkala, 2016). In addition, logging volumes have for instance increased c. 20% in the 21st century within the study area (Virkkala et al., 2020). Therefore, we cannot conclude that our SDMs would have failed to predict changes simply because nothing had changed.

A more likely explanation for failing to predict change involves missing information. Even if we used relatively long-term data, the variables which we thought would best associate with species distributions over the study area may not be the ones that are most important in driving the change. We suspect that our classification of the land cover variables (six categories) might have been too coarse, and thus, it might not have included enough information on fine-resolution changes in habitat quality, such as changes in the forest structure due to increased logging intensity (Virkkala et al., 2020), changes in agricultural landscape due to increased farming intensity (Laaksonen & Lehtikoinen, 2013) or degradation of peatlands due to draining (Fraixedas et al., 2017). While the level of resolution of the abiotic variables that we used in our model (temperature and land cover) probably was adequate to determine suitable sites for species, that is where the species was present or absent, the missing fine-resolution detail of the land cover variable may have hindered our predictions on species abundance and especially, where local changes have happened. Furthermore, different species may be affected by the same variables but at different spatial scales. For example, species with small territories are more likely affected by local changes in the habitat whereas species with larger territories can be more affected by landscape level changes (Table S1.1: Appendix S1).

Another source of prediction error could originate from the fact that we did not model the whole species' ranges. Undoubtedly, it would be preferable to calibrate an SDM with information from the species' full environmental space (Owens et al., 2013), albeit global calibration might not in certain cases always improve predictive performance either (El-Gabbas & Dormann, 2018; Giovanelli et al., 2010). However, we note that our study area covered a circa 1900 km latitudinal gradient, which is a large spatial coverage and relatively broad in SDM studies in general. Further, for about 10% of the species, mainly northern boreal species, our study area did cover full climatic conditions of their range, yet they (as mainly rare species) did not show particularly high predictive performance in the species trait analysis.

In addition, our test data for change validation method were geographically biased towards southern Sweden (Figure 2). However, we assume that this geographical bias was negligible because most of our study species had a southerly distribution and because this sample still covers mostly the same temperature (cold to warm) and land cover (from agricultural to forested and mountain landscape) gradients as a sample with a more northerly coverage would. One gradient that is not covered with this sampling is the increasing human population density towards southern latitudes. This could be important because in Norway, Husby et al. (2021) found that farmland birds have a stronger decline near urban areas than in rural areas. We acknowledge the possibility that omitting this effect of urban landscape may have distorted our results, yet we assume its effect to be marginal as only a small proportion of the study species breed in farmlands.

Lastly, an ongoing debate questions the use of correlative SDMs for projections of future distribution changes altogether (Dormann et al., 2012; Ehrlén & Morris, 2015; Zurell et al., 2009). In essence,

correlative SDMs have not been originally designed to generate accurate projections of change, as opposed to more mechanistic, process-based models, for example dynamic occupancy models, which directly model occupancy states in space-time (Kéry et al., 2013). Despite their limitations, correlative SDMs are still often preferred because their data requirements are more feasible compared with alternative mechanistic approaches. This may result in a larger spatial coverage of predictions, but of poorer quality.

4.3 | Influence of species traits

Regarding species traits, our results were not clearly consistent throughout the validation methods and data types, thus not showing strong support for our hypotheses. The hypothesis concerning lower prevalence leading to lower predictability gained support in both occurrence and abundance data, but significantly only when associated with prediction performances calculated with the static method. Higher predictive performance of more common species could be related to better detection probability compared with rarer species, which would decrease bias in the data (Johnston et al., 2014; McPherson & Jetz, 2007).

4.4 | Importance for conservation

Species distribution models can be used in various ways in conservation ecology (Guisan et al., 2013; Sofaer et al., 2019). However, if the models are blind to the factors that drive changes, and the spatial predictions derived from them completely miss the exact locations where changes are likely to happen, then one should not base any local conservation action, such as establishing finely targeted new protected areas for endangered species, on these projections. When data allow, an alternative might be to use, for example species-specific dynamic occupancy models, yet also their predictions need to be equally rigorously validated. Although our model failed to predict local change events, it still fared decently in predicting overall which species declined and which increased at the whole study area level. Therefore, these predictions can still prove to be helpful when preparing conservation actions over larger spatial scales.

5 | CONCLUSION

In conclusion, as SDMs are becoming increasingly important tools to predict future distributions of species and as time series are getting longer, also enabling validation by measuring change, we should pay more attention to the reliability of the predictions. We recommend that when the aim was to make future forecasts of changes in species distributions, the models used for the purpose are validated using temporally independent data and with methods that are able to evaluate not only the model's ability to make future predictions on a general level but also its ability to

predict change events in the spatial pattern of species occupancy or abundance. We remind that the choice of modelling approach and validation method should be determined by the intended application of the model. We also conclude that the field should proceed towards more critical model validations and thus take the challenge of making time series datasets more broadly available and attempting to make better models that can be shown to be successful in predicting also fine-scaled change events in species distributions.

ACKNOWLEDGEMENTS

We acknowledge the tremendous fieldwork effort made by the volunteers. This research was partly funded through the 2017–2018 Belmont Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND programme, and with the funding organizations Academy of Finland (Helsinki: 326338), Swedish Research Council (Lund: 2018-02441), the Research Council of Norway (RCN, NINA: 295767) and the National Science Foundation (CLO, ICER-1927646). The Swedish Bird Survey is supported by grants from the Swedish Environmental Protection Agency, with additional financial and logistic support from the Regional County Boards (Länsstyrelsen). The censuses were carried out within the framework of the Linnaeus-project Centre for Animal Movement Research (CAnMove) and the strategic research environment Biodiversity and Ecosystem Services in a Changing Climate (BECC). The Norwegian Climate and Environment Ministry and the Norwegian Environment Agency finance the Norwegian bird monitoring. Additionally, AL was funded by the Academy of Finland (project 275606). SP was funded by the Finnish Cultural Foundation and the Kone Foundation (grant number 201903886). OO was funded by Academy of Finland (grant no. 309581), Jane and Aatos Erkko Foundation, Research Council of Norway through its Centres of Excellence Funding Scheme (223257), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 856506; ERC-synergy project LIFEPLAN). We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<http://www.uerra.eu>) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (<https://www.ecad.eu>). We thank also the anonymous reviewers.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13687>.

DATA AVAILABILITY STATEMENT

The original bird survey dataset includes 63 species whose occurrence and location information are classified as 'sensitive' according to Scandinavian laws and regulations and thus cannot be openly shared as such. Therefore, in the openly shared dataset, the

coordinates for the survey sites have been made coarser. These data, as well as R-code and additional data for analysis, are available in Dryad at <https://doi.org/10.5061/dryad.bzkh189br>. The original, accurate data are available from the corresponding author upon reasonable request. The newest versions of the E-OBS weather data are freely available from the European Climate Assessment & Dataset project webpage at <https://www.ecad.eu/download/ensembles/download.php#datafiles>. The Corine land cover data are freely available from the Copernicus Programme webpage at <https://land.copernicus.eu/pan-european/corine-land-cover>.

ORCID

Sirke Piirainen  <https://orcid.org/0000-0002-2568-8134>

Aleksi Lehtikainen  <https://orcid.org/0000-0002-1989-277X>

REFERENCES

- Araújo, M. B., Whittaker, R. J., Ladle, R. J., & Erhard, M. (2005). Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, 14(6), 529–538. <https://doi.org/10.1111/J.1466-822X.2005.00182.X>
- Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122(3), 321–331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>
- Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(23), 9374–9379. <https://doi.org/10.1073/pnas.1220228110>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
- Charney, N. D., Record, S., Gerstner, B. E., Merow, C., Zarnetske, P. L., & Enquist, B. J. (2021). A test of species distribution model transferability across environmental and geographic space for 108 Western North American tree species. *Frontiers in Ecology and Evolution*, 9, 689295. <https://doi.org/10.3389/fevo.2021.689295>
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409. <https://doi.org/10.1029/2017JD028200>
- Dobrowski, S. Z., Thorne, J. H., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M., & Swanson, A. K. (2011). Modeling plant ranges over 75 years of climate change in California, USA: Temporal transferability and species traits. *Ecological Monographs*, 81(2), 241–257. <https://doi.org/10.1890/10-1325.1>
- 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>
- Ehrlén, J., & Morris, W. F. (2015). Predicting changes in the distribution and abundance of species under environmental change. *Ecology Letters*, 18(3), 303–314. <https://doi.org/10.1111/ele.12410>
- El-Gabbas, A., & Dormann, C. F. (2018). Wrong, but useful: Regional species distribution models may not be improved by range-wide data under biased sampling. *Ecology and Evolution*, 8(4), 2196–2206. <https://doi.org/10.1002/ece3.3834>
- Eskildsen, A., le Roux, P. C., Heikkinen, R. K., Høye, T. T., Kissling, W. D., Pöyry, J., Wisz, M. S., & Luoto, M. (2013). Testing species distribution models across space and time: High latitude butterflies and recent warming. *Global Ecology and Biogeography*, 22(12), 1293–1303. <https://doi.org/10.1111/geb.12078>

- Fraixedas, S., Lindén, A., Meller, K., Lindström, Å., Keiřs, O., Kålås, J. A., Husby, M., Leivits, A., Leivits, M., & Lehikoinen, A. (2017). Substantial decline of Northern European peatland bird populations: Consequences of drainage. *Biological Conservation*, 214, 223–232. <https://doi.org/10.1016/j.biocon.2017.08.025>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Giovanelli, J. G. R., de Siqueira, M. F., Haddad, C. F. B., & Alexandrino, J. (2010). Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods. *Ecological Modelling*, 221(2), 215–224. <https://doi.org/10.1016/j.ecolmodel.2009.10.009>
- 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>
- Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, 23(3), 494–508. <https://doi.org/10.1111/j.1420-9101.2009.01915.x>
- Harris, D. J., Taylor, S. D., & White, E. P. (2018). Forecasting biodiversity in breeding birds using best practices. *PeerJ*, 6, e4278. <https://doi.org/10.7717/peerj.4278>
- Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2015). The drivers of avian abundance: Patterns in the relative importance of climate and land use. *Global Ecology and Biogeography*, 24(11), 1249–1260. <https://doi.org/10.1111/geb.12377>
- Huntley, B., Green, R., Collingham, Y., & Willis, S. (2007). *A climatic atlas of European breeding birds*. Lynx Edicions.
- Husby, M., Hoset, K. S., & Butler, S. (2021). Non-random sampling along rural-urban gradients may reduce reliability of multi-species farmland bird indicators and their trends. *Ibis*, 163(2), 579–592. <https://doi.org/10.1111/ibi.12896>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424), 444–448. <https://doi.org/10.1038/nature11631>
- Johnston, A., Ausden, M., Dodd, A. M., Bradbury, R. B., Chamberlain, D. E., Jiguet, F., Thomas, C. D., Cook, A. S. C. P., Newson, S. E., Ockendon, N., Rehfisch, M. M., Roos, S., Thaxter, C. B., Brown, A., Crick, H. Q. P., Douse, A., McCall, R. A., Pontier, H., Stroud, D. A., ... Pearce-Higgins, J. (2013). Observed and predicted effects of climate change on species abundance in protected areas. *Nature Climate Change*, 3, 1055–1061. <https://doi.org/10.1038/nclimate2035>
- Johnston, A., Newson, S. E., Risely, K., Musgrove, A. J., Massimino, D., Baillie, S. R., & Pearce-Higgins, J. (2014). Species traits explain variation in detectability of UK birds. *Bird Study*, 61(3), 340–350. <https://doi.org/10.1080/00063657.2014.941787>
- Journé, V., Barnagaud, J.-Y., Bernard, C., Crochet, P.-A., & Morin, X. (2020). Correlative climatic niche models predict real and virtual species distributions equally well. *Ecology*, 101(1), e02912. <https://doi.org/10.1002/ecy.2912>
- Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanese, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M. V., Bauer, H.-G., & Foppen, R. P. B. (2020). *European Breeding Bird Atlas 2: Distribution, abundance and change*. European Bird Census Council & Lynx Edicions.
- Kéry, M., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2013). Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, 40(8), 1463–1474. <https://doi.org/10.1111/jbi.12087>
- Kujala, H., Moilanen, A., Araújo, M. B., & Cabeza, M. (2013). Conservation planning with uncertain climate change projections. *PLoS One*, 8(2), e53315. <https://doi.org/10.1371/journal.pone.0053315>
- Laaksonen, T., & Lehikoinen, A. (2013). Population trends in boreal birds: Continuing declines in agricultural, northern, and long-distance migrant species. *Biological Conservation*, 168, 99–107. <https://doi.org/10.1016/j.biocon.2013.09.007>
- Leathwick, J. R., Whitehead, D., & McLeod, M. (1996). Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: A modelling approach. *Environmental Software*, 11(1), 81–90. [https://doi.org/10.1016/S0266-9838\(96\)00045-7](https://doi.org/10.1016/S0266-9838(96)00045-7)
- Lehikoinen, A., Green, M., Husby, M., Kålås, J. A., & Lindström, Å. (2014). Common montane birds are declining in northern Europe. *Journal of Avian Biology*, 45(1), 3–14. <https://doi.org/10.1111/j.1600-048X.2013.00177.x>
- Lehikoinen, A., & Virkkala, R. (2016). North by North-West: Climate change and directions of density shifts in birds. *Global Change Biology*, 22(3), 1121–1129. <https://doi.org/10.1111/gcb.13150>
- Lindström, Å., Green, M., Husby, M., Kålås, J. A., & Lehikoinen, A. (2015). Large-scale monitoring of waders on their boreal and arctic breeding grounds in northern Europe. *Ardea*, 103(1), 3–15. <https://doi.org/10.5253/ardea.v103i1.a1>
- Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., & Knowlton, N. (2020). Climate change and ecosystems: Threats, opportunities and solutions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1794), 20190104. <https://doi.org/10.1098/rstb.2019.0104>
- Massimino, D., Johnston, A., Gillings, S., Jiguet, F., & Pearce-Higgins, J. (2017). Projected reductions in climatic suitability for vulnerable British birds. *Climatic Change*, 145(1), 117–130. <https://doi.org/10.1007/s10584-017-2081-2>
- McPherson, J. M., & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30(1), 135–151. <https://doi.org/10.1111/j.2006.0906-7590.04823.x>
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., ... Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89(3), e01370. <https://doi.org/10.1002/ecm.1370>
- Ovaskainen, O., & Abrego, N. (2020). *Joint species distribution modelling: With applications in R*. Cambridge University Press.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5), 561–576. <https://doi.org/10.1111/ele.12757>
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10–18. <https://doi.org/10.1016/j.ecolmodel.2013.04.011>
- Pearce-Higgins, J. W., Eglinton, S. M., Blaise, M., & Chamberlain, D. E. (2015). Drivers of climate change impacts on bird communities. *Journal of Animal Ecology*, 84(4), 943–954. <https://doi.org/10.1111/1365-2656.12364>
- Rapacciuolo, G., Roy, D. B., Gillings, S., Fox, R., Walker, K., & Purvis, A. (2012). Climatic associations of British species distributions show good transferability in time but low predictive accuracy for range change. *PLoS One*, 7(7), 1–11. <https://doi.org/10.1371/journal.pone.0040212>
- Rapacciuolo, G., Roy, D. B., Gillings, S., & Purvis, A. (2014). Temporal validation plots: Quantifying how well correlative species distribution

- models predict species' range changes over time. *Methods in Ecology and Evolution*, 5(5), 407–420. <https://doi.org/10.1111/2041-210X.12181>
- Regos, A., Gagne, L., Alcaraz-Segura, D., Honrado, J. P., & Domínguez, J. (2019). Effects of species traits and environmental predictors on performance and transferability of ecological niche models. *Scientific Reports*, 9(1), 4221. <https://doi.org/10.1038/s41598-019-40766-5>
- 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>
- Sofaer, H. R., Jarnevich, C. S., & Flather, C. H. (2018). Misleading prioritizations from modelling range shifts under climate change. *Global Ecology and Biogeography*, 27(6), 658–666. <https://doi.org/10.1111/GEB.12726>
- Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., Edwards, T. C., Guala, G. F., Howard, T. G., Morissette, J. T., & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. *Bioscience*, 69(7), 544–557. <https://doi.org/10.1093/BIOSCI/BIZ045>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Tayleur, C., Caplat, P., Massimino, D., Johnston, A., Jonzén, N., Smith, H. G., & Lindström, Å. (2015). Swedish birds are tracking temperature but not rainfall: Evidence from a decade of abundance changes. *Global Ecology and Biogeography*, 24(7), 859–872. <https://doi.org/10.1111/geb.12308>
- Tessarolo, G., Lobo, J. M., Rangel, T. F., & Hortal, J. (2021). High uncertainty in the effects of data characteristics on the performance of species distribution models. *Ecological Indicators*, 121, 107147. <https://doi.org/10.1016/j.ecolind.2020.107147>
- Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehtikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution*, 11(3), 442–447. <https://doi.org/10.1111/2041-210X.13345>
- Uribe-Rivera, D. E., Guillera-Aroita, G., Windecker, S. M., Plischoff, P., & Wintle, B. A. (2022). The predictive performance of process-explicit range change models remains largely untested. *Ecography*, e06048. <https://doi.org/10.1111/ecog.06048>
- Vaughan, I. P., & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, 42(4), 720–730. <https://doi.org/10.1111/j.1365-2664.2005.01052.x>
- Venne, S., & Currie, D. J. (2021). Can habitat suitability estimated from MaxEnt predict colonizations and extinctions? *Diversity and Distributions*, 27(5), 873–886. <https://doi.org/10.1111/DDI.13238>
- Virkkala, R., Lehtikoinen, A., & Rajasärkkä, A. (2020). Can protected areas buffer short-term population changes of resident bird species in a period of intensified forest harvesting? *Biological Conservation*, 244, 108526. <https://doi.org/10.1016/j.biocon.2020.108526>
- Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*, 30(12), 766–779. <https://doi.org/10.1016/j.tree.2015.09.007>
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, 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., Jeltsch, F., Dormann, C. F., & Schröder, B. (2009). Static species distribution models in dynamically changing systems: How good can predictions really be? *Ecography*, 32(5), 733–744.

BIOSKETCH

Sirke Piirainen is a doctoral researcher at the Finnish Museum of Natural History at the University of Helsinki. Her main areas of research are the effects of land use and climate change on the persistence of bird populations in Fennoscandia. The team of co-authors is interested in macroecology, species distribution modelling, population ecology and global change.

Author contributions: S.P., A.L. and O.O. conceived the ideas. S.P., A.L., M.H., J.A.K. and Å.L. gathered the data. S.P. and O.O. analyzed the data. S.P. led the writing. All authors critically reviewed the manuscript and approved the final version.

SUPPORTING INFORMATION

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

How to cite this article: Piirainen, S., Lehtikoinen, A., Husby, M., Kålås, J. A., Lindström, Å., & Ovaskainen, O. (2023). Species distributions models may predict accurately future distributions but poorly how distributions change: A critical perspective on model validation. *Diversity and Distributions*, 00, 1–12. <https://doi.org/10.1111/ddi.13687>