DR. NATALIE JANE BRISCOE (Orcid ID : 0000-0003-0049-8956)
DR. DAMARIS ZURELL (Orcid ID : 0000-0002-4628-3558)
DR. CHRISTIAN KOENIG (Orcid ID : 0000-0003-0585-5983)
DR. GUILLERMO FANDOS (Orcid ID : 0000-0003-1579-9444)

Article type : Primary Research Articles

# **Title**: Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds.

**Running Title:** Testing dynamic occupancy model predictions

**Authors:** Natalie J. Briscoe[1*], Damaris Zurell[2,3], Jane Elith[1], Christian Koenig[2,3] , Guillermo Fandos[2,3], Anne-Kathleen Malchow[2,3], Marc Kéry[4], Hans Schmid[4], Gurutzeta Guillera-Arroita[1]

**Institutional affiliations:**

1. School of BioSciences, University of Melbourne, Parkville, VIC, 3010, Australia
2. Geography Dept., Humboldt-University Berlin, Berlin, Germany
3. Inst. for Biochemistry and Biology, Potsdam University, Potsdam, Germany
4. Swiss Ornithological Institute, Sempach, Switzerland

***Corresponding author:** Natalie Briscoe (nbriscoe@unimelb.edu.au)

**Keywords:** species distribution models, predictive performance, model evaluation, multiseason occupancy models, detection probability, species trends

## Abstract

Predictions of species' current and future ranges are needed to effectively manage species under environmental change. Species ranges are typically estimated using correlative species distribution models (SDMs), which have been criticized for their static nature. In contrast, dynamic occupancy models explicitily describe temporal changes in species' occupancy via colonisation and local extinction probabilities, estimated from time series of occurrence data. Yet, tests of whether these models improve predictive accuracy under current or future conditions are rare. Using a long-term dataset on 69 Swiss birds, we tested whether dynamic occupancy models improve predictions of distribution changes over time compared to SDMs. We evaluated the accuracy of spatial predictions and their ability to detect population trends. We also explored how predictions differed when we accounted for imperfect detection and parameterised models using calibration datasets of different time series lengths. All model types had high spatial predictive performance when assessed across all sites (mean AUC > 0.8), with flexible machine-learning SDM algorithms outperforming parametric static and dynamic occupancy models. However, none of the models performed well at identifying sites where range changes are likely to occur. In terms of estimating population trends, dynamic occupancy models performed best, particularly for species with strong population changes and when fit with sufficient data, while static SDMs performed very poorly. Overall, our study highlights the importance of considering what aspects of performance matter most when selecting a modelling method for a particular application and the need for further research to improve model utility. While dynamic occupancy models show promise for capturing range dynamics and inferring population trends when fitted with sufficient data, computational constraints on variable selection and model fitting can lead to reduced spatial accuracy of predictions, an area warranting more attention.

## Introduction

Species are shifting their ranges in response to changes in land-use and climate. Range shifts have been empirically documented for a variety of taxa (Parmesan & Yohe, 2003). However, empirical data typically only capture information about species occurrence or abundance in a subset of places and times. To derive a more complete picture of where species are likely to occur both now and in the future, we typically rely on species distribution models (Guisan & Thuiller, 2005).

Correlative (static) species distribution models (SDMs) describe the likely occurrence or abundance of a species as a function of environmental covariates (Elith & Leathwick, 2009). In addition to predicting the probability of species occurring in unsampled regions or under altered environmental conditions, SDMs are often used to understand and infer possible drivers of species distributions (Kelly et al., 2017) and to predict species' range or population trends or future extinction risk (Keith et al., 2014; Zhang et al., 2017). However, there is growing recognition that SDMs can provide misleading predictions, particularly when used to forecast species responses to environmental change (Sequeira et al., 2018; Sofaer et al., 2018) or when applied to species that are not in equilibrium with their environment (Elith et al., 2010). Researchers have argued that explicitly modelling the processes that underlie changes in distribution and abundance across the landscape may be important for forecasting range dynamics accurately (Briscoe et al., 2019; Joseph, 2020; Urban et al., 2016; Yackulic et al., 2015; Zurell, 2017). Furthermore, by explicitly identifying and testing mechanisms that drive range dynamics at the population or individual level, such process-explicit models can provide critical insights and inform effective management of species under environmental change (Fordham et al., 2013; Merow et al., 2017).

In addition to accounting for dynamics, several studies have highlighted the importance of accounting for observation errors when inferring range shifts or using occurrence data to model species distributions (Kujala et al., 2013; MacKenzie et al., 2002). Even when data are collected using well-planned survey designs and focus on sessile organisms (Chen et al., 2013), species are not always detected when they are present, and variation in detectability across environments can lead to misleading predictions that more closely resemble where we are able to detect the species rather than where the species truly occurs (Kéry, 2011; Lahoz-Monfort et al., 2014).

Many different methods have been proposed for incorporating processes explicitly into models of range dynamics (Briscoe et al., 2019); however, data to adequately fit these models are often lacking (Urban et al., 2016). Process-explicit models can also be very complex, often requiring additional expertise, species-specific knowledge and computational resources that make their implementation challenging (Hefley et al., 2017). Further, several model-comparison studies have shown that more complex process-explicit models do not always improve predictions compared to simpler correlative SDMs, although this depends on the underlying dynamics and time horizon (Briscoe et al., 2016; Fordham et al., 2018; Zurell et al., 2016).

Among the suite of available modelling tools, dynamic occupancy models may represent a good ‘trade-off’ between model complexity and realism: they use the same *type* of data as static SDMs (occurrence data, compared to, e.g. abundance or demographic data required by other process-explicit models), but describe more closely the processes that underpin range dynamics by explicitly modelling species colonisation and local extinction events at sites using data collected across time (Kéry et al., 2013; MacKenzie et al., 2003). Unlike static correlative SDMs, dynamic occupancy models do not assume that species are in equilibrium with their environment. They can also account for imperfect detection if data are collected so they are informative about the observation process, e.g. if repeat surveys or time-to-detection data are available within sampling seasons (Guillera-Arroita, 2017); but can otherwise be fitted without the observation component (Kéry & Royle 2021). These methods, also known as multiseason occupancy models, have typically not been included in studies comparing correlative SDMs with a range of process-explicit methods (Fordham et al., 2018; Pagel & Schurr, 2012; Zurell et al., 2016). But single-species studies have highlighted their utility for predicting range dynamics across both local and broad spatial scales (Clement et al., 2016; Heard et al., 2015; Santika et al., 2014).

Here, we aim to evaluate, across many species, the potential advantages of i) accounting for dynamics via explicitly modelling colonisation and extinction processes, and ii) accounting for imperfect detection when describing and predicting spatio-temporal patterns of species distributions. To this end, we fitted four different types of models: static correlative species distribution models (SDMs) and dynamic occupancy models (DOMs) that do not account for imperfect detection, as well as their counterparts, static occupancy detection models (SODMs) and dynamic occupancy detection models (DODMs) that do account for imperfect species detection. We expected that dynamic occupancy detection models would perform better when species are not in equilibrium, and when detection probability is low. To reflect the breadth of modelling approaches currently used, we included correlative SDMs with different abilities to capture simple to complex covariate relationships and with contrasting variable selection approaches.

Both data limitations and the lack of independent model evaluation have been highlighted as key challenges hampering broader and improved use of process-explicit models of range dynamics (Briscoe et al., 2019; Urban et al., 2016). We address this gap by evaluating predictive performance of dynamic occupancy detection models (and related simpler methods) fitted to varying amounts of data. We used the Swiss national common breeding bird survey, *Monitoring Häufige Brutvögel* (MHB) (see also, for instance Kéry & Schmid, 2006; Schmid et al., 2004), which provides suitable data for accounting for detection (repeat surveys in each year) over a time period (14 years) during which range changes are known to have occurred for at least some species (Knaus et al., 2018). The dataset includes species with different ecologies that show a variety of responses to recent changes in climate and land-use (Knaus et al., 2018; Maggini et al., 2011). Critically, the timespan over which data were collected allowed us to use different subsets of the dataset for model fitting (3-10 years), while setting aside the remaining years (4-11) as an independent dataset on which to evaluate model predictions.

## Methods

### Species data

We used data from the Swiss breeding bird survey from 2000-2013. The survey data were collected within the MHB (*Monitoring Häufige Brutvögel*; the Swiss common breeding bird survey) which covers 267 sampling sites (1 km$^2$ cell) every year. We used additional data from the Swiss Biodiversity Monitoring (BDM), which covers an additional 267 cells in a five-year rotating scheme (approx. one-fifth each year; 2001-2013), to test model predictions at independent sites. Within-season replicate surveys (2-3) at each 1 km$^2$ sample site allowed us to fit occupancy models that accounted for imperfect detection. Data were cleaned to exclude observations from outside each species' breeding season (described in Kéry & Schmid, 2006) and surveys without dates recorded (date was used as covariate to model detection). From this 'cleaned' MHB dataset, we included in our analyses all species (excluding waterfowl) that were observed in at least 50 sites across all time-periods used (n = 69 species). Waterfowl were excluded because we had insufficient environmental data for their key habitat (water).

Since there was more than one observation per year, methods were needed to deal with the within-year repeat visits when fitting models assuming perfect detection. We considered two approaches: (a) "collDet", collapsing all within-year data at a given site to one value, with a presence assigned if the species was observed in any survey, and absence otherwise; (b) "det2", using the observed record from the second survey of the year (peak breeding season), mimicking datasets with only one visit per year. We tested both approaches on the static SDMs. As results were not very sensitive to the approach used (see Results), we fitted the dynamic occupancy models, which take longer to run, only with "collDet". "collDet" datasets were always used for model evaluation.

The BDM data were used for independent evaluation only and were not included in the training data. In the MHB dataset, high mountain sites are underrepresented, while in the BDM data set, sites occur closer to their proportion in the Swiss landscape. Thirteen species were not detected in the BDM dataset, so independent evaluation was not possible for these species.

### Static site covariates

We used static land use layers summarising the proportion of each type of landcover in each 1 km$^2$ cell. Swiss land use statistics were available for two timepoints: collected over the years 1992-97 and

2004-2009 (GEOSTAT). We used the land use data from the later years (2004-2009) as these overlapped most closely with our occurrence data. We collapsed the general land-use categories (n=17, Aggregation NOAS04_17) into 10 broad classes that captured the key aspects likely to influence bird species (Table S1.1, Fig. S1.1).

### Dynamic site covariates

We used monthly CHELSA timeseries climate data (http://chelsa-climate.org) to derive relevant yearly climate predictors that may influence bird range dynamics. These included mean annual temperature (bio1), temperature seasonality (bio4), yearly annual precipitation (bio12), precipitation seasonality (bio15), as well as precipitation in the wettest and driest months (bio13, bio14). We also included maximum temperature in the warmest month (bio5) and minimum temperature of the coldest month (bio6) to account for the effect of temperature extremes.

Mean and standard deviation of NDVI was calculated each year using monthly NDVI data obtained from Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (https://lpdaac.usgs.gov/products/mod13a3v006/). These variables were included to quantify shifts in the tree-line over time (not captured by the static land-use layers), and to help differentiate forest types (e.g. non-deciduous needleleaf differs less across the year so standard deviation is likely to be lower) and regions that show stable or highly variable patterns of productivity.

### Survey covariates

When modelling detection, we considered date (day of the year on which the survey was conducted) with a quadratic term as a predictor for detection probability because it influences detection of bird species in this dataset (Kéry & Royle, 2016). We also allowed for an interaction between date and elevation, to account for variation in seasonal patterns of detection across elevations (e.g. see Kéry et al. 2013). Models with interaction terms between survey date and elevation also included the main effect terms.

### Variable pre-filtering

From the full set of 20 potential site covariates, we selected a subset of 14 variables that were not substantially correlated (|*r*| < 0.7) and that captured the key axes of environmental variation across the study region and over all years – a pragmatic approach when modelling a diverse range of species. To do this, in every year we performed a Principle Component Analysis (PCA) to assess how much environmental variation each variable captured across the study region. Then, for each principal axis, starting with the first, we selected the variable with the highest loading and assigned it a rank based on the variation explained by that principle axis (i.e. the variable with the highest loading on the first axis had a rank of one). Across the 20 PCA axes, this gave a set of 20 variables with rank 1-20, but there may be duplicates (since the same variable might load first on different axes). Duplicates were removed. For the set of variables selected, we then assessed all pairwise correlations using Spearman's rank correlation. Pairs with an absolute value above 0.7 were considered correlated (Dormann et al., 2013) and therefore the variable with the lower rank was removed, with the exception that priority was given to variables that quantified yearly changes (climate and NDVI) because of our focus on modelling how species' ranges change over time. We retained the 14 top-ranked variables summed across years. The final set of candidate predictors included six dynamic variables (bio4,5,12,14,15, standard deviation of NDVI) and eight static landcover variables (LC5, 6, 8, 9, 10,11, 12, 13_17). Continuous predictors (climate, elevation, date and NDVI variables) were standardized before analyses by subtracting the mean and dividing by the standard deviation.

### Model fitting

Five modelling approaches were used to cover axes of dynamics (yes or no), detection error (yes or no) and model structure (Table 1). All models were fitted to three datasets (hereafter *training datasets*) which included data from the first three years (2000-2002), the first five years (2000-2004), and the first ten years (2000-2009) of surveys, respectively. Dynamic site covariates were all sampled at the year matching the year of observation. Predictions were made to all years (2000 to 2013) using year-specific dynamic covariates.

We aimed to fit the models in ways that would be typical for good usage of the methods, and feasible across multiple species. For all models other than Boosted Regression Trees (BRTs), we further reduced the set of predictors by fitting univariate models of occurrence, including linear and quadratic terms, and selecting the environmental variables from the four top models as ranked using AIC. For models that account for imperfect detection, univariate models of occurrence included a full model for detectability:

$$p \sim date + date^2 + elev + date{:}elev + date{:}elev^2. \quad \text{eq1}$$

This sub-setting of potential covariates was necessary to reduce the number of candidate models, which scales to the $2^n$ where n is the number of terms included in the model; for dynamic occupancy detection models, this includes each term considered for initial occupancy, colonisation, extinction and detection (e.g. for 4 environmental co-variates and 3 survey co-variate sets we fitted 12,288 DODMs). We used a ‘data driven’ approach for model selection (preselecting predictors based on model fitting) because, with the exception of survey covariates, we did not have *a priori* hypotheses about the variables driving each process, and our aim was to fit the best model rather than to make inferences about the importance of individual variables.

In contrast, we offered the full set of 14 candidate variables for BRTs, following common practice in applying machine learning methods (Merow et al., 2014). Our approach meant that there were differences in the steps taken to identify the best models when using different approaches, which could influence how well they performed. However, these differences were driven by the practical realities and features of the underlying methods, and thus are justifiable. The ability of BRTs to contend with a relatively large set of candidate variables is a strength of this approach, and thus it may unfairly disadvantage them if models were only fit to a much smaller subset of variables. We further explored the impact of these decisions by fitting BRTs with only the subset of variables used to fit DODMs.

Dynamic and static occupancy models were fit using the *colext* and *occu* functions in the R package 'unmarked' (Fiske et al. 2019), which uses maximum likelihood methods for parameter estimation. For each species, we fit all candidate models where either initial occupancy, colonisation and extinction (DODMs) or occupancy (SODMs) were modelled as a function of the four selected environmental variables, including both linear and quadratic terms (i.e. we tested every combination of retained predictors in each model component). Detection was modelled in three ways: as specified above (eq1), as a function of date only (linear and quadratic terms), or assumed to be constant. We fitted all models, with the constraint that linear and quadratic terms for a variable were always included together, and selected the 'best' model using AIC. When fitting static occupancy models, each site-by-year combination was treated as a site (i.e. we temporally-matched each species observation with the environmental conditions at that site in the survey year). To fit dynamic occupancy models (DOMs) ignoring imperfect detection, we re-ran the analysis as above but forcing the detection probability parameter to be one by "tricking" the software. To achieve this, we duplicated the aggregated collDet data (i.e. whether the species was ever detected at the site in each year) to fit models, therefore mimicking perfect detection (as in e.g. Kéry et al 2013). We then followed the same variable selection strategy as above.

GLMMs were fitted using *glmer* in the package lme4 (Bates et al., 2015, p. 4) with year as a random effect to model potential differences in general suitability between consecutive years. As for the static and dynamic occupancy detection models, we fitted all models with four environmental covariates and linear and quadratic terms, choosing the most parsimonious using AIC.

One challenge when fitting complex models with multiple covariates via maximum likelihood estimation is that the numerical optimization algorithms sometimes may identify local rather than global optima (Guillera-Arroita et al., 2014; Kéry & Royle, 2021). To reduce the risk that static and dynamic occupancy models and GLMMs used for prediction suffer from this problem, we fitted all models with 10 sets of starting values for the regression parameters (uniformly distributed between -1.5 and 1.5) and kept the overall best model with the lowest AIC.

BRTs were fitted using packages: gbm and dismo – specifically the *gbm.step* function in dismo (Hijmans et al., 2017). Since there are multiple observations per site given repeat sampling in different years, the internal cross-validation used to tune the complexity of BRTs was structured so that all records for a site were within the same fold (Read et al., 2011). This means that the internal estimate of predictive performance, which is used to tune the model, is based on sites not included in the training dataset. Following the methods and explanations of Elith et al. (2008), we aimed to fit at least 1000 trees in each BRT model. To achieve this, we set individual values for the learning rate (lr) and tree complexity (tc) on a per-species basis (based on the number of records per species), testing different values of lr and tc if the number of trees in the final model was below 1000 or over 10000. Final learning rates varied from 0.005 to 0.0005, and tree complexity from 1 to 2, which resulted in relatively simple trees.

We provide detailed model descriptions in Appendix S3 following the ODMAP protocol (Zurell et al., 2020), including details of computational time.

### Model prediction

For static methods, occupancy predictions were generated by simply applying the selected best model to all years of environmental data (2000-2013) to compute site occupancy probability. For dynamic methods, predictions were generated by simulating the likely future occupancy status of all cells in the landscape. To do this, we ran 5000 simulations using the probabilistic estimates of initial occupancy, colonisation and extinction functions from the fitted model, drawing the occurrence state of each site (0/1) accordingly at each time step. Models were run for the entire study period (2000-2013).

### Model evaluation

Model evaluation focused on testing both the spatial predictions of models and assessing how well they captured range-wide temporal trends in occupancy and abundance (overview in Fig. 1). To evaluate the spatial predictive ability of our models, we compared our occurrence predictions to the observed data. As noted by Kéry *et al.* (2013), when data are such that cumulative detection is imperfect, this comparison indicates the model's ability to predict the combination of occurrence probability and detection probability (Yackulic et al., 2012), but does not directly measure how well the model predicts species occurrences. For methods that account for imperfect detection (DODMs and SODMs), we therefore evaluated models using predictions of both occurrence probability and the combined outcome of occurrence probability and detection probability (Lahoz-Monfort et al., 2014) when survey information were available.

We evaluated *spatial predictions* of models using mean yearly AUC (Area under the ROC Curve) and % deviance explained, using both the entire MHB dataset and only the test years not used for fitting models (i.e. temporally independent data). AUC provides a measure of model discrimination (whether occupied sites are consistently ranked higher than unoccupied sites), whilst % deviance explained compares the deviance explained by the model with that of a null model that estimates species prevalence in each year (i.e. the proportion of sites where the species was detected). We constrained % deviance explained to 0-100% to avoid negative values when models perform worse than the null. We obtained these metrics first using all surveyed sites and then focusing only on sites that showed some occupancy change (i.e. were both occupied and unoccupied across the study period), in order to more specifically assess how well models predicted range changes. For this, we also calculated the correlation between the change in predicted occupancy from year 1 to year 14 and the corresponding change in observed occupancy status, across all sites. Finally, we calculated AUC and % deviance explained using detection – nondetection data from BDM surveys (not used for model fitting), at sites at least 2km away from sites used for model fitting (i.e. considered to be spatially independent data).

We evaluated the model's ability to capture and predict species' *temporal trends* using the C-index (Harrell et al., 1982; hereafter temporal AUC). The C-index is a generalisation of AUC, originally developed for survival data (Cook, 2007). Here we applied it to trends in populations. It quantifies the probability that the ranking of pairs of predictions matches the rankings of the observations used for testing. We compared predicted species trends summarized using yearly sums of the estimated probability of occupancy across the landscape (the *expected* area of occupancy, AOO), with trend estimates produced by the annual Swiss Breeding bird indices, which provide an index of relative changes in population size (Knaus et al., 2020), for the full period 2000-2013. This is an interesting comparison because, despite their limitations (see Discussion), SDM outputs are sometimes used to infer trends in abundance. We also did these calculations comparing predicted trends to the number of survey cells observed to be occupied (the *observed* area of occupancy) in the MHB dataset. We calculated C-index values using the *rcorr.cens* function in the Hmisc package (Harrell, 2019).

Selected variables, fitted responses and spatial predictions were also assessed via visual checks for a representative sub-set of species showing varying population trends and ecologies. Although we tried to ensure good model fits, this was not possible for all species and data combinations, particularly when shorter training datasets were used.

### Comparative analyses of model performance

Finally, we explored how modelling choices and species' range characteristics affect model performance. These analyses (details in Appendix S2) included (1) modelling the relationship between performance (spatial and temporal AUCs) and aspects relating to modelling approach (method, training dataset) and range characteristics (prevalence, population trend and range trend); and (2) evaluating paired differences in model performance (AUC) and model agreement (mean difference in cell-based estimates of occurrence) for both static (SODMs/GLMMs or BRTs) and dynamic models (DODMs/DOMs) in relation to both species' detection (mean detection from raw occurrence data) and training dataset. To generate indices of population trends for the above analyses, we fitted generalised linear models as a function of year to observed Swiss-wide population trends 2000-2013, as provided by the Swiss Bird Index (Knaus et al., 2020). We estimated regression coefficients (b), which we also used to classify species into broad categories for visualisation and interpretation purposes, as follows: strongly increasing ($b >= 0.05$), increasing ($0.05 < b > =0.01$), stable ($0.01 < b >= -0.01$) or decreasing ($b < -0.01$), with *b* set to 0 when year did not have a significant effect (i.e. there was no clear yearly trend).

To explore implications of using different models to guide spatial conservation decisions about where to protect habitat for species, we also calculated, for each pair of models, the ratio between the sum of predictions from one method across the top 5% of cells based on predicted occupancy as identified by that method and across the top 5% of cells as identified by the second method of the pair. Here, values close to 1 would suggest similar expected conservation outcomes of targeting sites selected by each of the models.

## Results

### *Model performance*

Figure 1 exemplifies the analyses and typical results for species that displayed temporal population trend, specifically the declining Eurasian Skylark. Spatial AUC values were highest for BRTs, which more accurately identified occupied habitat (top row of spatial predictions). However, temporal AUC (C-Index) was highest for DODMs (when fitted to 10 years of data), which accurately captured population declines (bottom panel, observed population trends shown in black), but not with high spatial accuracy. Across species, there was little difference between spatial AUCs of static SDMs fitted with the collDet and det2 datasets, however models fitted with collDet data had higher average % deviance explained (Fig. S1.2). For simplicity, we only present results from models built using collDet data in subsequent analyses.

All methods had relatively high discrimination for most species, as measured by spatial AUC calculated *across the range* using temporally and spatially independent test datasets, with mean AUC for all methods >0.8 regardless of the training dataset used to fit the model (Fig. 2, Fig. S1.3). Model performance was most variable across methods and species when the shorter training datasets (3 years) were used to fit models. BRTs consistently had the highest spatial AUCs, with all other methods having similar mean performance except for DODMs, which had lower mean AUCs when only 3 years of data were used. Similar patterns were observed for % deviance explained, although overall model performance was relatively low and more variable (Fig. 2b, Fig. S1.3b). BRTs ranked as the top model for 84-97% of species based on spatial AUC, and 65-94% based on % deviance explained (Fig. S1.4). Spatial model performance metrics were mostly consistent within species: species that were modelled well by one method or training dataset were typically modelled well by all and vice versa (Fig. S1.5). Pairwise comparisons of predictions at the top 5% of cells selected by each model suggested moderate to good agreement between models, with the average of ratios of probability sums ranging from 0.85 - 0.98 for models using the same methods but different training datasets, and from 0.80 - 0.95 for models using the same training datasets but different methods (Table S1.2). In contrast, spatial predictive ability was much lower when looking only at sites that showed some occupancy change (Fig. 4; Fig. S1.6-7). Mean spatial AUC was between 0.64-0.68 for all methods, except BRT, again the best performing method, which achieved 0.69-0.73. Deviance also indicated poor fit between predictions and observations at sites with occupancy changes, and correlations between predicted and observed changes were low.

Tests of model ability to capture species' temporal trends showed a different pattern, with DODMs and DOMs having the highest temporal AUC values when fitted with 10 years of data (Fig. 3, Fig. S1.8). This pattern was driven by their better ability to predict trends in species that have undergone population changes over this period. Consistent with this, DODMs were ranked the top model for the majority of species (30-49%), followed by DOMs (20-24%) and then BRTs (10-16%) (Fig. S1.9). In comparison to spatial AUC, temporal AUC values varied substantially across training data sets and particularly across different models (Fig. S1.10).

### *Drivers of differences in model performance*

The results of our comparative performance analyses (Appendix S2) showed that spatial and temporal AUC values were most strongly influenced by species' prevalence, followed by population trend and then modelling method. Prevalence generally had a positive effect on spatial AUC, and a negative effect on temporal AUC. For temporal AUC, performance increased for species with more positive population trends. Modelling method had a stronger effect on temporal AUC performance compared to spatial AUC, and had a noticeable interaction with population trend, with DODM and DOM performance more affected by trends (as also seen in Fig. 3).

Comparisons of model performance for model pairs that did and did not account for imperfect detection in the model fitting (DODMs v DOMs; SODMs v GLMMs; SODMs v BRTs) did not identify a consistent effect of species detectibility (Appendix S2), likely due to the overall high per-visit detection probabilities achieved in the Swiss survey data (minimum 0.46, median 0.77). With three repeat visits to a site, only three species (Eurasian Sparrowhawk, Hawfinch, and Long-tailed Bushtit) would have an overall cumulative detection probability < 0.9. Consistent with this, spatial AUC of models that accounted for imperfect detection was very similar when calculated based on the projected probability of occurrence, compared to the combined probabilities of occurrence and detection (Fig. S1.11), with a few notable exceptions for species where detection probability varied strongly with elevation. However, % deviance explained was higher when detection probabilities were accounted for, particularly in SODMs. Differences in the predicted area of occupancy (AOO) (%) between all model pairs that did and did not account for imperfect detection were correlated with detection probability ($r$ = -0.43 – -0.51), with models that accounted for detection yielding larger estimates of AOO for species with lower detection probabilities. For most species, differences were relatively low (mean: 10-12%, median: 5-8%), however for species with lower detection probabilities, estimates of AOO differed by up to 85% (Eurasian Sparrowhawk, p = 0.46). Selected environmental predictors often differed between paired models that did and did not account for detection; variables included in models for the same species and using the same training dataset differed for between 52-56% of species for DODMs and DOMs, and 65-81% of species for SODMs and GLMMs (Figs. S1.12-15).

The strong model performance of BRTs predicting spatial patterns appeared to be at least partly driven by the more flexible approach to variable selection: when we re-fit BRTs using the same settings, but restricting models to only the four selected environmental variables used to fit DODMs for each species and training dataset combination, model performance decreased (Fig. S16), although it remained slightly higher on average than that for the other methods.

An analysis of novel environmental space across the study region using Multivariate Environmental Similarity Surfaces (MESS; Elith et al. 2010) highlighted the potential for extrapolation, particularly when models were fitted to three years of data (Fig. S12). However, the variables showing novel values (bio4 and bio15) were rarely selected as covariates (Figs. S13-15).

## Discussion

Species distribution models are used to inform a wide range of conservation and management decisions, including identifying potential and high-quality habitat for threatened species, and inferring future population trends based on changes in potential habitat (Guisan et al., 2013; Hof et al., 2018). As one of the first studies to apply dynamic occupancy detection models (MacKenzie et al., 2003) to predict the distributions of a large number of species, our study highlights methodological challenges and opportunities to improve the use of these methods for forecasting range dynamics. Focusing on model predictive ability over relatively short time frames that are relevant for many management decisions and using real data for a suite of bird species, we found that models that explicitly account for colonisation-extinction dynamics were better able to capture population trends when given enough calibration data but yet had lower spatial predictive accuracy than the most flexible static SDMs. Critically, none of the considered methods performed well at predicting the occupancy of sites that experienced change. These results underline the importance of considering what aspects of model performance matter most when applying distribution modelling to specific conservation and management problems and ensuring that model evaluation targets these aspects (Rapacciuolo, 2019).

Across species, all of the models we considered showed relatively good spatial predictive performance across the species range, with BRTs being consistently the best at capturing species' distributions across space. These differences in spatial model performance between BRTs and other modelling methods seemed to be at least partly due to the different approach we took to variable selection and model fitting: when BRTs were fitted with the selected four environmental co-variates used to fit DODMs (rather than using the full suite of 14 candidate variables), model performance decreased (Fig. S10). These decisions reflect practical constraints: BRTs are easy to fit, whereas the other models are computationally heavy given the need to fit many models to find the 'best' one. Even with our reduced predictor set, fitting all candidate DODMs *for a single species and training dataset* took between 2-14 days, compared to running BRTs *for all species and training datasets*, which ran in under five hours (see Appendix S3). Other studies fitting DODMs have instead considered parameters sequentially (Clement et al., 2016), adopted Bayesian approaches that avoid overfitting by using regularization or penalization (Rushing et al., 2019) or combined the structural rigor of hierarchical modeling with the algorithmic flexibility of machine learning methods (Joseph, 2020). Here, we deliberately chose to focus on a commonly adopted and readily available implementation of the maximum likelihood approach to test model pipelines that could be (more) readily applied by practioners. Yet, Bayesian methods show great promise for improving variable selection and model fitting in cases where the drivers and nature of relationships are poorly understood (Hooten & Hobbs, 2015). Similarly, non-parametric model implementations that are capable of capturing more complex non-linear responses could improve spatial accuracy of predictions (Rushing et al., 2019). The ability of DODMs to generate accurate spatial predictions is often not assessed using independent data (Briscoe et al., 2019), perhaps because data regarding the true presence or absence of a species are rarely available. In addition, studies employing these methods to date have often focused on quantifying current or past ranges or trends, using indirect co-variates such as elevation, year or site, which are not suitable for forecasting (Guisan &

Zimmermann, 2000). Our study highlights the need to focus on improving the spatial predictive performance of DODMs and how this can be critical to expanding their utility. A key result in our study was that none of the modelling methods considered performed well at sites that showed changes in occupancy. Better model selection could improve the ability of DODMs to predict where change is likely to occur, and therefore their value to inform spatial conservation decisions.

Metrics of temporal pattern prediction ability were generally poor. For species with stable population trends, this may simply reflect the fact that differences between years are just “noise” over a constant baseline, resulting in temporal AUC values close to 0.5. However, static methods also struggled with inferring temporal trends of species that showed either increases or declines over time. Dynamic occupancy models (both DODMs and DOMs) were much better than other methods in this context, when fitted with longer time series of data. This is an important result, particularly for users of standard SDMs who hope to model species trends. This result is likely driven by the fact that dynamic occupancy models directly model changes in occupancy via colonization and extinction and explicitly account for the status of the site in the previous timestep. The model structure is designed to describe how sites become occupied or unoccupied by species, and the model does not require the equilibrium assumption needed for static approaches, and therefore it is better suited to predict the fate of sites into the future (Yackulic et al., 2015; Clement et al., 2016). Output from static SDMs often indicated either no trend or rapid fluctuations between years, and were not, on average, useful for inferring population changes (mean temporal AUC for static models was 0.49). Static SDMs rely on the modelled relationship between the probability of occurrence and environmental covariates. Thus, changes in occupancy will only be predicted if these are driven by (or correlate with) changes in the environment, and because yearly predictions are completely independent, they can fluctuate widely. In contrast, in our dynamic occupancy models, the occupancy status of a site is dependent on its status in the previous timestep, often resulting in smoother predicted year-to-year changes. In addition, changes in occupancy occur via colonisation or extinction, and although these can vary through space and time as a function of environmental covariates, the environment does not need to change for changes in occupancy to occur. Perhaps surprisingly, colonisation and extinction rates were frequently modelled as a function of static land-use classes (Figs. S1.13-15).

While there are extensions to dynamic occupancy detection models that attempt to more directly capture population dynamics (MacKenzie et al., 2009, 2012), here we inferred temporal population trends using the change in the expected area of occupancy (AOO), as estimated by summing the occupancy probabilities in each timestep. This is consistent with approaches used to inform assessments of extinction risk (IUCN Standards & Petitions Subcommittee 2017) or impacts of environmental change (Clement et al., 2016; Hof et al., 2018). Several studies have highlighted that AOO may not always be a useful predictor of population trends, and even if there is a relationship, it may not be linear (Fordham et al., 2012). However, even when range changes were evident in our dataset (e.g. Common Kestrel, Red Kite), static SDMs often predicted no overall trend, which is a concern if these methods are being used to inform conservation risk assessments. Two important caveats are that: (i) we did not account for dispersal constraints or minimum habitat size when calculating AOO; and (ii) we focused on using yearly weather data to capture near-term trends (14 years in total), which differs from many correlative SDM studies that focus on predicting species responses over longer timespans as a function of long-term climate (Porfirio et al., 2014; Gardner et al., 2019; although see for example Runge et al., 2015). This shorter timeframe is relevant to many management decisions, for example, population declines are assessed over 3 generations under the IUCN guidelines, which is < 10 years for many bird species (Bird et al., 2020). However, the superior performance of DODMs for capturing population trends that we observed may not hold when predicting over longer timeframes and larger, directional changes if transitions in the training data do not adequately capture drivers of future changes and allow them to be modelled via extinction and colonisation. In DODMs and DOMs the transition from an occupied to an unoccupied site is dictated by the extinction probability, which is not informed by information from sites where the species never occurs.

In addition to generating predictions, distribution models are also commonly used for explanation (Araújo et al., 2019). One advantage of dynamic occupancy models relative to static models is that they can capture information about what drives both colonisation and extinction processes, helping us better understand and test the mechanisms that underlie range dynamics (Kalle et al., 2018). For example, we found that across the range of dynamic models fitted, extinction probability for Eurasian Skylarks (a declining species with strong avoidance of forests) was consistently higher in regions with more forest or wooded areas, and colonisation was very low. Such insight can be useful for guiding management (Hamer et al., 2016).

Accounting for imperfect detection did not consistenly improve performance in our dataset, where most species had high detectability. It did, however, affect estimates of the area of occupancy for the less detectable species, resulting in a predicted area of occupancy ~1.8 times larger for the species with the lowest detection probability (Eurasian Sparrowhawk, $p = 0.46$). The high levels of detection in our study are rare: a review by Kellner & Swihart (2014) showed that 70% of studies accounting for species detectability reported per-survey estimates of detection lower than 0.5. In general, accounting for imperfect detection will be particularly important if accurate estimates of AOO are needed; if detection (e.g. due to survey effort) varies through time and model outputs are being used to assess changes; or if detection varies with environmental co-variates and spatial outputs are being used to identify key habitat (Guillera-Arroita, 2017; Guillera-Arroita et al., 2015).

Data limitations are often cited as a barrier to the use of more process-explicit models of range dynamics (Urban et al., 2016), and our results also highlight this issue for more complex dynamic occupancy detection models. We found surprisingly little effect of the temporal extent of training data on model performance across most model classes, with performance more strongly driven by species attributes (prevalence, population trends) and modelling method. However, there was a trade-off between model complexity and the number of years of training data used to fit models, with the performance of the most complex models (DODMs) increasing when they were fitted with more years of data. This was most evident for their ability to capture temporal trends and suggests the utility of DODMs for better capturing species temporal trends will depend on the availability of relatively large datasets, with observations over multiple years (>5 years). This result however is linked to this relatively stable system and timeframe. Potentially, systems undergoing faster change could be characterized with data collected over shorter time periods. While our dataset included repeat visits to all sites every year, it is worth noting that this is not necessarily required, and that imperfect detection can be modelled when data for modelling the observation process are collected at a subset of sites and/or years (Guillera-Arroita 2017) or even with single visits when time to the first detection is recorded in each (Garrard et al., 2008).

Recent calls for wider use of process-explicit models to forecasts species' range shifts and population trends rely heavily on the assumption that models that explicitly capture the mechanisms known to underpin responses to environmental change, will provide more accurate forecasts (Urban et al., 2016; Zurell, 2017). Model evaluation studies, such as this one, that test which methods actually provide reliable forecasts under which circumstances are urgently needed. Our findings that DODMs were best at inferring past and future temporal trends in population size, while static SDMs (BRTs) were more adept at capturing general spatial distribution patterns, reinforces the notion that different modelling approaches often have different strengths, and so the selection of a modelling method and implementation should be informed by the intended application (Briscoe et al., 2019; Guisan et al., 2013). Our results showed that fitting DODMs is feasible for many species – though more computationally challenging than static approaches. It also highlighted that their ability to provide improved predictions of population trends depends on the level of population change and data availability, and that there is scope and a need for improving their spatial predictive performance. Overall, our study shows the potential utility of DODMs for better forecasting how environmental change will impact on species, while also highlighting areas of research investment – variable selection, model fitting and testing spatial predictive performance using independent data – that are likely to yield further improvements and make these methods practical in a wider range of situations.

## Acknowledgements

We thank the Swiss Ornithological Institute for providing access to species occurrence data and Markus Jenny for providing the Eurasian Skylark image. This research was funded by an Australia-Germany Joint Research Co-operation Scheme grant (57445394), and an ARC Discovery grant (DP180101852) awarded to JE and GGA. DZ, GF, CK and AKM acknowledge support from the German Science Foundation DFG (ZU 361/1-1). The Melbourne University Research Cloud & Research Computing Services (HPC) provided computational support.

## Supporting Information

Appendix S1: Supplementary figures and tables
Appendix S2: Comparative analyses of model performance
Appendix S3: ODMAP protocol (describing modelling steps and approach in more detail).

**Data availability**: Code and model evaluation data that support the findings of this study are available via gitlab (https://gitlab.unimelb.edu.au/nbriscoe/testing-dynamic-occupancy-models). Species occurrence and environmental data used for fitting models are available upon reasonable request (contact: mhb@vogelwarte.ch).

## References

Araújo, M. B., Anderson, R. P., Barbosa, A. M., 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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., Symes, A., Taylor, J., Şekercioğlu, Ç. H., & Butchart, S. H. M. (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, *34*(5), 1252–1261. https://doi.org/10.1111/cobi.13486

Briscoe, N. J., Elith, J., Salguero-Gómez, R., Lahoz-Monfort, J. J., Camac, J. S., Giljohann, K. M., Holden, M. H., Hradsky, B. A., Kearney, M. R., McMahon, S. M., Phillips, B. L., Regan, T. J., Rhodes, J. R., Vesk, P. A., Wintle, B. A., Yen, J. D. L., & Guillera-Arroita, G. (2019). Forecasting species range dynamics with process-explicit models: Matching methods to applications. *Ecology Letters*, *22*(11), 1940–1956. https://doi.org/10.1111/ele.13348

Briscoe, N. J., Kearney, M. R., Taylor, C. A., & Wintle, B. A. (2016). Unpacking the mechanisms captured by a correlative species distribution model to improve predictions of climate refugia. *Global Change Biology*, *22*(7), 2425–2439. https://doi.org/10.1111/gcb.13280

Chen, G., Kéry, M., Plattner, M., Ma, K., & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, *101*(1), 183–191. https://doi.org/10.1111/1365-2745.12021

Clement, M. J., Hines, J. E., Nichols, J. D., Pardieck, K. L., & Ziolkowski, D. J. (2016). Estimating indices of range shifts in birds using dynamic models when detection is imperfect. *Global Change Biology*, *22*(10), 3273–3285. https://doi.org/10.1111/gcb.13283

Cook, N. R. (2007). Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation*, *115*(7), 928–935. https://doi.org/10.1161/CIRCULATIONAHA.106.672402

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *METHODS IN ECOLOGY AND EVOLUTION*, *1*(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.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

Fordham, D. A., Akçakaya, H. R., Araújo, M. B., Elith, J., Keith, D. A., Pearson, R., Auld, T. D., Mellin, C., Morgan, J. W., Regan, T. J., Tozer, M., Watts, M. J., White, M., Wintle, B. A., Yates, C., & Brook, B. W. (2012). Plant extinction risk under climate change: Are forecast range shifts alone a good indicator of species vulnerability to global warming? *Global Change Biology*, *18*(4), 1357–1371. https://doi.org/10.1111/j.1365-2486.2011.02614.x

Fordham, D. A., Akçakaya, H. R., Brook, B. W., Rodríguez, A., Alves, P. C., Civantos, E., Triviño, M., Watts, M. J., & Araújo, M. B. (2013). Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nature Climate Change*, *3*(10), 899–903. https://doi.org/10.1038/nclimate1954

Fordham, D. A., Bertelsmeier, C., Brook, B. W., Early, R., Neto, D., Brown, S. C., Ollier, S., & Araújo, M. B. (2018). How complex should models be? Comparing correlative and mechanistic range dynamics models. *Global Change Biology*, *24*(3), 1357–1370. https://doi.org/10.1111/gcb.13935

Gardner, A. S., Maclean, I. M. D., & Gaston, K. J. (2019). Climatic predictors of species distributions neglect biophysiologically meaningful variables. *Diversity and Distributions*, *25*(8), 1318–1333. https://doi.org/10.1111/ddi.12939

Garrard, G. E., Bekessy, S. A., McCARTHY, M. A., & Wintle, B. A. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, *33*(8), 986–998. https://doi.org/10.1111/j.1442-9993.2008.01869.x

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, *40*(2), 281–295. https://doi.org/10.1111/ecog.02445

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276–292. https://doi.org/10.1111/geb.12268

Guillera-Arroita, G., Lahoz-Monfort, J. J., MacKenzie, D. I., Wintle, B. A., & McCarthy, M. A. (2014). Ignoring Imperfect Detection in Biological Surveys Is Dangerous: A Response to 'Fitting and Interpreting Occupancy Models'. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0099571

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. 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

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hamer, A. J., Heard, G. W., Urlus, J., Ricciardello, J., Schmidt, B., Quin, D., & Steele, W. K. (2016). Manipulating wetland hydroperiod to improve occupancy rates by an endangered amphibian: Modelling management scenarios. *Journal of Applied Ecology*, *53*(6), 1842–1851. https://doi.org/10.1111/1365-2664.12729

Harrell, F. E. (2019). *Hmisc: Harrell Miscellaneous* (R package version 4.2-0) [Computer software]. https://CRAN.R-project.org/package=Hmisc

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, *247*(18), 2543–2546.

Heard, G. W., Thomas, C. D., Hodgson, J. A., Scroggie, M. P., Ramsey, D. S. L., & Clemann, N. (2015). Refugia and connectivity sustain amphibian metapopulations afflicted by disease. *Ecology Letters*, *18*(8), 853–863. https://doi.org/10.1111/ele.12463

Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., & Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecology Letters*, *20*(5), 640–650. https://doi.org/10.1111/ele.12763

Hijmans, R. J., Phillips, S. J., Leathwick, J. R., & Elith, J. (2017). *dismo: Species Distribution Modeling* (R package version 1.1-4) [Computer software]. https://CRAN.R-project.org/package=dismo

Hof, C., Voskamp, A., Biber, M. F., Böhning-Gaese, K., Engelhardt, E. K., Niamir, A., Willis, S. G., & Hickler, T. (2018). Bioenergy cropland expansion may offset positive effects of climate change mitigation for global vertebrate diversity. *Proceedings of the National Academy of Sciences*, *115*(52), 13294–13299. https://doi.org/10.1073/pnas.1807745115

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28. https://doi.org/10.1890/14-0661.1

Joseph, M. B. (2020). Neural hierarchical models of ecological populations. *Ecology Letters*, *23*(4), 734–747. https://doi.org/10.1111/ele.13462

Kalle, R., Ramesh, T., & Downs, C. T. (2018). When and where to move: Dynamic occupancy models explain the range dynamics of a food nomadic bird under climate and land cover change. *Global Change Biology*, *24*(1), e27–e39. https://doi.org/10.1111/gcb.13861

Keith, D. A., Elith, J., & Simpson, C. C. (2014). Predicting distribution changes of a mire ecosystem under future climates. *Diversity and Distributions*, *20*(4), 440–454. https://doi.org/10.1111/ddi.12173

Kellner, K. F., & Swihart, R. K. (2014). Accounting for Imperfect Detection in Ecology: A Quantitative Review. *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0111436

Kelly, L. T., Haslem, A., Holland, G. J., Leonard, S. W. J., MacHunter, J., Bassett, M., Bennett, A. F., Bruce, M. J., Chia, E. K., Christie, F. J., Clarke, M. F., Stefano, J. D., Loyn, R., McCarthy, M. A., Pung, A., Robinson, N., Sitters, H., Swan, M., & York, A. (2017). Fire regimes and environmental gradients shape vertebrate and plant distributions in temperate eucalypt forests. *Ecosphere*, *8*(4), e01781. https://doi.org/10.1002/ecs2.1781

Kéry, M. (2011). Towards the modelling of true species distributions. *Journal of Biogeography*, *38*(4), 617–618. https://doi.org/10.1111/j.1365-2699.2011.02487.x

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

Kéry, M., & Royle, A. (2016). *Applied hierarchical modeling in ecology—Modeling distribution, abundance and species richness using R and BUGS. Volume 1: Prelude and Static Models.* Elsevier / Academic Press.

Kéry, M., & Royle, J. A. (2021). *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS. Volume 2: Dynamic and Advanced Models*. Elsevier Science Publishing Co Inc.

Kéry, M., & Schmid, H. (2006). Estimating species richness: Calibrating a large avian monitoring programme. *Journal of Applied Ecology*, *43*(1), 101–110. https://doi.org/10.1111/j.1365-2664.2005.01111.x

Knaus, P., Antoniazza, S., Wechsler, S., Guélat, J., Kéry, M., Strebel, N., & Sattler, T. (2018). *Swiss Breeding Bird Atlas 2013–2016. Distribution and population trends of birds in Switzerland and Liechtenstein.* Swiss Ornithological Institute.

Knaus, P., Schmid, H., Strebel, N., & Sattler, T. (2020). *The State of Birds in Switzerland 2020 online*. Swiss Ornithological Institute, Sempach. http://www.vogelwarte.ch/state

Kujala, H., Vepsäläinen, V., Zuckerberg, B., & Brommer, J. E. (2013). Range margin shifts of birds revisited – the role of spatiotemporally varying survey effort. *Global Change Biology*, *19*(2), 420–430. https://doi.org/10.1111/gcb.12042

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, *23*(4), 504–515. https://doi.org/10.1111/geb.12138

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

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, *83*(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

MacKenzie, D. I., Nichols, J. D., Seamans, M. E., & Gutiérrez, R. J. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, *90*(3), 823–835. https://doi.org/10.1890/08-0141.1

MacKenzie, D. I., Seamans, M. E., Gutiérrez, R. J., & Nichols, J. D. (2012). Investigating the population dynamics of California spotted owls without marked individuals. *Journal of Ornithology*, *152*(2), 597–604. https://doi.org/10.1007/s10336-010-0544-6

Maggini, R., Lehmann, A., Kéry, M., Schmid, H., Beniston, M., Jenni, L., & Zbinden, N. (2011). Are Swiss birds tracking climate change?: Detecting elevational shifts using response curve shapes. *Ecological Modelling*, *222*(1), 21–32. https://doi.org/10.1016/j.ecolmodel.2010.09.010

Merow, C., Bois, S. T., Allen, J. M., Xie, Y., & Silander, J. A. (2017). Climate change both facilitates and inhibits invasive plant ranges in New England. *Proceedings of the National Academy of Sciences*, *114*(16), E3276–E3284. https://doi.org/10.1073/pnas.1609633114

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, *37*(12), 1267–1281. https://doi.org/10.1111/ecog.00845

Pagel, J., & Schurr, F. M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, *21*(2), 293–304. https://doi.org/10.1111/j.1466-8238.2011.00663.x

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*(6918), 37–42. https://doi.org/10.1038/nature01286

Porfirio, L. L., Harris, R. M. B., Lefroy, E. C., Hugh, S., Gould, S. F., Lee, G., Bindoff, N. L., & Mackey, B. (2014). Improving the Use of Species Distribution Models in Conservation Planning and

Management under Climate Change. *PLOS ONE*, *9*(11), e113749. https://doi.org/10.1371/journal.pone.0113749

Rapacciuolo, G. (2019). Strengthening the contribution of macroecological models to conservation practice. *Global Ecology and Biogeography*, *28*(1), 54–60. https://doi.org/10.1111/geb.12848

Read, C. F., Duncan, D. H., Vesk, P. A., & Elith, J. (2011). Surprisingly fast recovery of biological soil crusts following livestock removal in southern Australia. *Journal of Vegetation Science*, *22*(5), 905–916. https://doi.org/10.1111/j.1654-1103.2011.01296.x

Runge, C. A., Tulloch, A., Hammill, E., Possingham, H. P., & Fuller, R. A. (2015). Geographic range size and extinction risk assessment in nomadic species. *Conservation Biology*, *29*(3), 865–876. https://doi.org/10.1111/cobi.12440

Rushing, C. S., Royle, J. A., Ziolkowski, D. J., & Pardieck, K. L. (2019). Modeling spatially and temporally complex range dynamics when detection is imperfect. *Scientific Reports*, *9*(1), 12805. https://doi.org/10.1038/s41598-019-48851-5

Santika, T., McAlpine, C. A., Lunney, D., Wilson, K. A., & Rhodes, J. R. (2014). Modelling species distributional shifts across broad spatial extents by linking dynamic occupancy models with public-based surveys. *Diversity and Distributions*, *20*(7), 786–796. https://doi.org/10.1111/ddi.12189

Schmid, H., Zbinden, N, & Keller, V. (2004). *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz.* Schweizerische Vogelwarte, Sempach.

Sequeira, A. M. M., Bouchet, P. J., Yates, K. L., Mengersen, K., & Caley, M. J. (2018). Transferring biodiversity models for conservation: Opportunities and challenges. *Methods in Ecology and Evolution*, *9*(5), 1250–1264. https://doi.org/10.1111/2041-210X.12998

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

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe'er, G., Singer, A., Bridle, J. R., Crozier, L. G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J. J., Holt, R. D., Huth, A., Johst, K., Krug, C. B., Leadley, P. W., Palmer, S. C. F., Pantel, J. H., … Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, *353*(6304), aad8466. https://doi.org/10.1126/science.aad8466

Yackulic, C. B., Nichols, J. D., Reid, J., & Der, R. (2015). To predict the niche, model colonization and extinction. *Ecology*, *96*(1), 16–23. https://doi.org/10.1890/14-1361.1

Yackulic, C. B., Reid, J., Davis, R., Hines, J. E., Nichols, J. D., & Forsman, E. (2012). Neighborhood and habitat effects on vital rates: Expansion of the Barred Owl in the Oregon Coast Ranges. *Ecology*, *93*(8), 1953–1966. JSTOR. http://www.jstor.org/stable/23225196

Zhang, J., Nielsen, S. E., Chen, Y., Georges, D., Qin, Y., Wang, S.-S., Svenning, J.-C., & Thuiller, W. (2017). Extinction risk of North American seed plants elevated by climate and land-use change. *Journal of Applied Ecology*, *54*(1), 303–312. https://doi.org/10.1111/1365-2664.12701

Zurell, D. (2017). Integrating demography, dispersal and interspecific interactions into bird distribution models. *Journal of Avian Biology*, *48*(12), 1505–1516. https://doi.org/10.1111/jav.01225

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., Thuiller, W., Pagel, J., Cabral, J. S., Münkemüller, T., Gravel, D., Dullinger, S., Normand, S., Schiffers, K. H., Moore, K. A., & Zimmermann, N. E. (2016). Benchmarking novel approaches for modelling species range dynamics. *Global Change Biology*, *22*(8), 2651–2664. https://doi.org/10.1111/gcb.13251

## Tables

**Table 1:** Modelling methods used. Methods differed in model structure, including whether they explicitly model detection and dynamics. For static occupancy models (i.e. correlative SDMs), we tested two commonly applied methods – Generalized Linear Mixed Models (GLMMs) and Boosted Regression Trees (BRTs). All models were fitted in R. Quotes are used for names of R packages and italics for functions within these packages. collDet = data aggregated across all survey visits, Det2 = data from second survey only. See Appendix 3 for additional details on model fitting.

| Model name | Detection | Dynamics | Notes on model fitting |
|---|---|---|---|
| Dynamic occupancy detection (DODM) | Y | Y | Fit using *colext* in 'unmarked' |
| Dynamic occupancy (DOM) | N | Y | Fit using *colext* in 'unmarked' using collDet data |
| Static occupancy detection (SOM) | Y | N | Fit using *occu* in 'unmarked' |
| Static occupancy – GLMMs | N | N | Fit using *glmer* in 'lme4' using collDet & Det2 data |
| Static occupancy – BRTs | N | N | Fit using *gbm.step* in 'dismo' using collDet & Det2 data |

## Figure legends

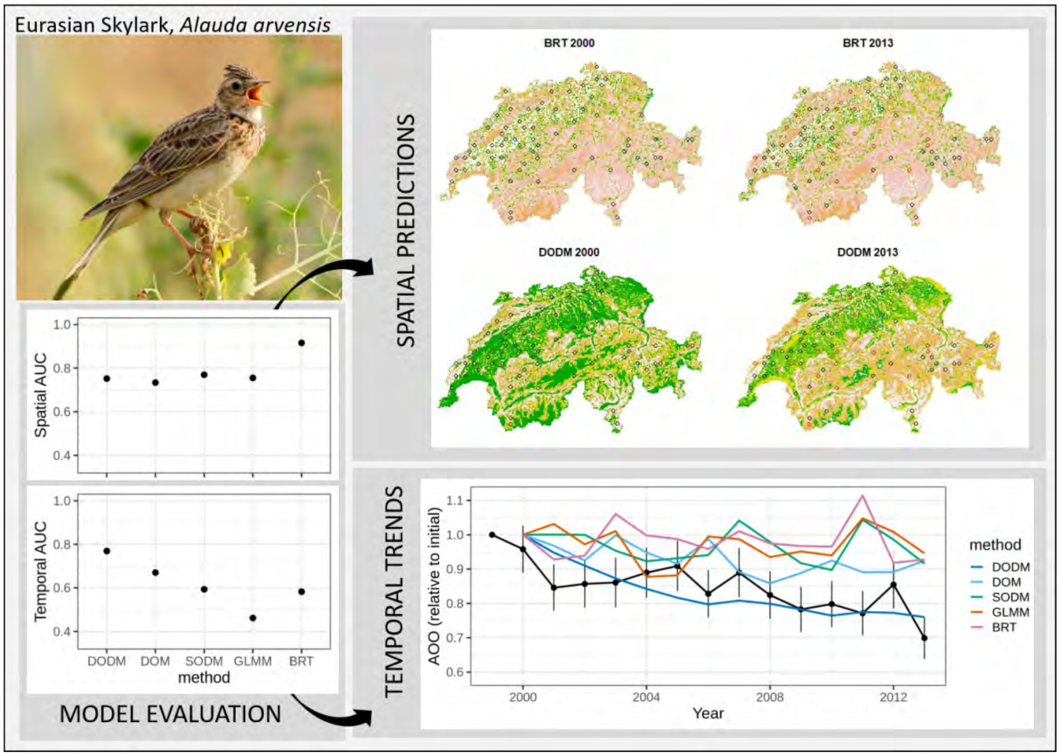
**Figure 1.** Example of approach and outputs used to evaluate predictive performance of model classes. Model evaluation metrics, spatial predictions and temporal trends are shown for the Eurasian Skylark (for simplicity only models fitted with 10 years training data are shown). Spatial AUC values were calculated using spatial predictions (top right panel) at survey sites in years subsequent to model fitting. Temporal AUC (C-Index) was calculated by summing spatial predictions in each year to get the change in predicted AOO (relative to year 1), and comparing this with data on relative population change (bottom right panel; black line = observed data). Black circles on the maps indicate sites where the species was observed in that year, and red circles indicate sites where the species was observed in 2000, but not in 2013. For the Eurasian Skylark, spatial AUC was highest for BRTs, which more accurately identified occupied habitat (top row of spatial predictions). However, temporal AUC (C-Index) was highest for DODMs, which captured overall pattern of population decline (bottom panel), but with lower spatial accuracy. Static SDMs were fitted with the collDet dataset. The same scale is used in the maps for both displayed methods. Image: Markus Jenny.

**Figure 2**. Box plots of spatial AUC and percent deviance explained across different types of distribution models fitted for 69 bird species in Switzerland and evaluated by projecting models beyond the training period (indicated in the top of each panel) and testing against data collected yearly up to 2013. Horizontal lines indicate median, with notches showing rough 95% confidence intervals.

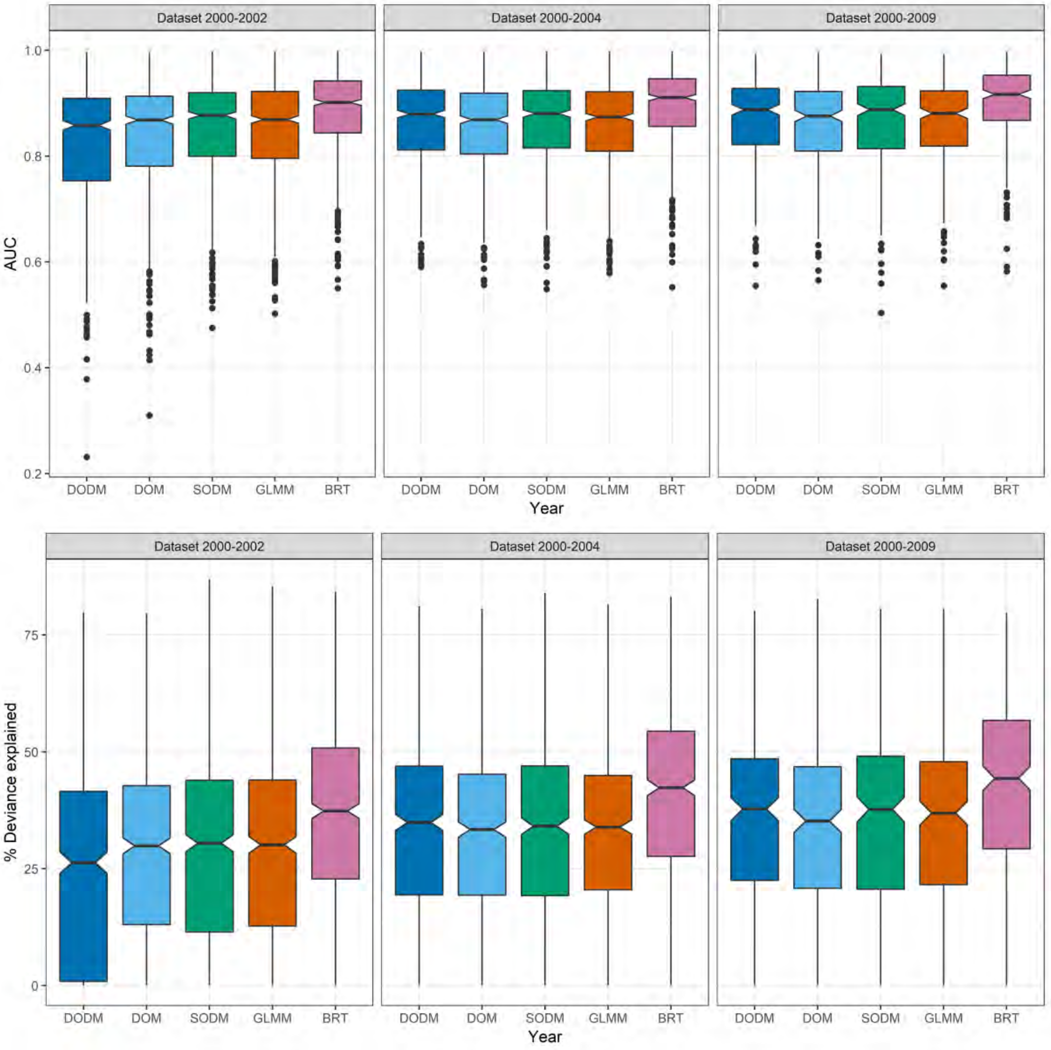
**Figure 3**. Temporal AUC of different types of distribution models fitted for 69 Swiss bird species using different training datasets (3, 5 or 10 years) and evaluated by testing model predictions of summed probability of occupancy in each year against yearly values of the Swiss Bird Index for that species. Panels show results for species classed based on population trends (2000-2013), n = the number of species in each group. Evaluated models: dynamic occupancy detection models (DODM),

dynamic occupancy models (DOM), static occupancy detection models(SODM), generalized linear mixed models (GLMM) and boosted regression trees (BRT).

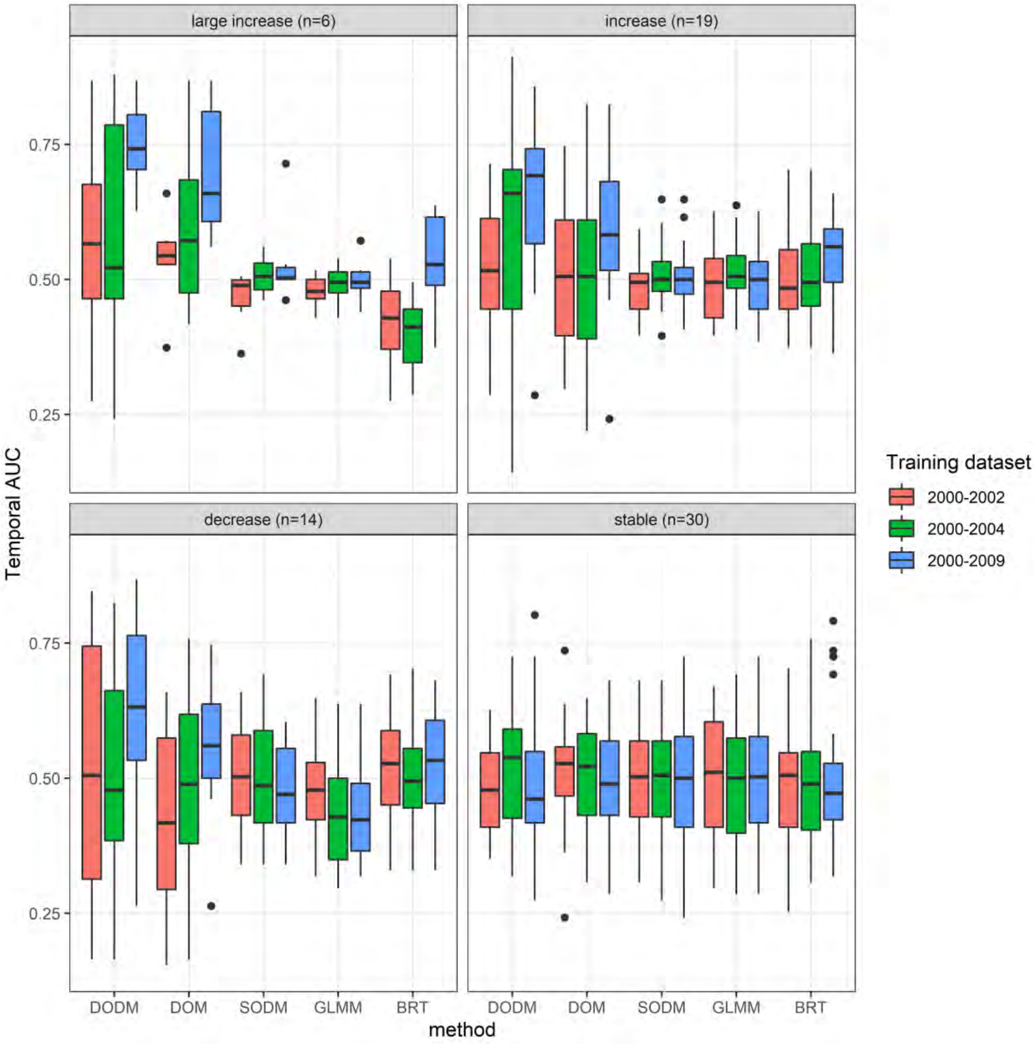
**Figure 4.** Box plots of spatial AUC, percent deviance explained and correlations between the predicted and observed change in occupancy from year 1 to year 14, across different types of distribution models fitted for 69 bird species in Switzerland using 10 years of data, and evaluated by projecting models beyond the training period (indicated in the top of each panel) and testing against data collected yearly up to 2013, focusing only on cells with some observed change. Horizontal lines indicate median, with notches showing rough 95% confidence intervals.



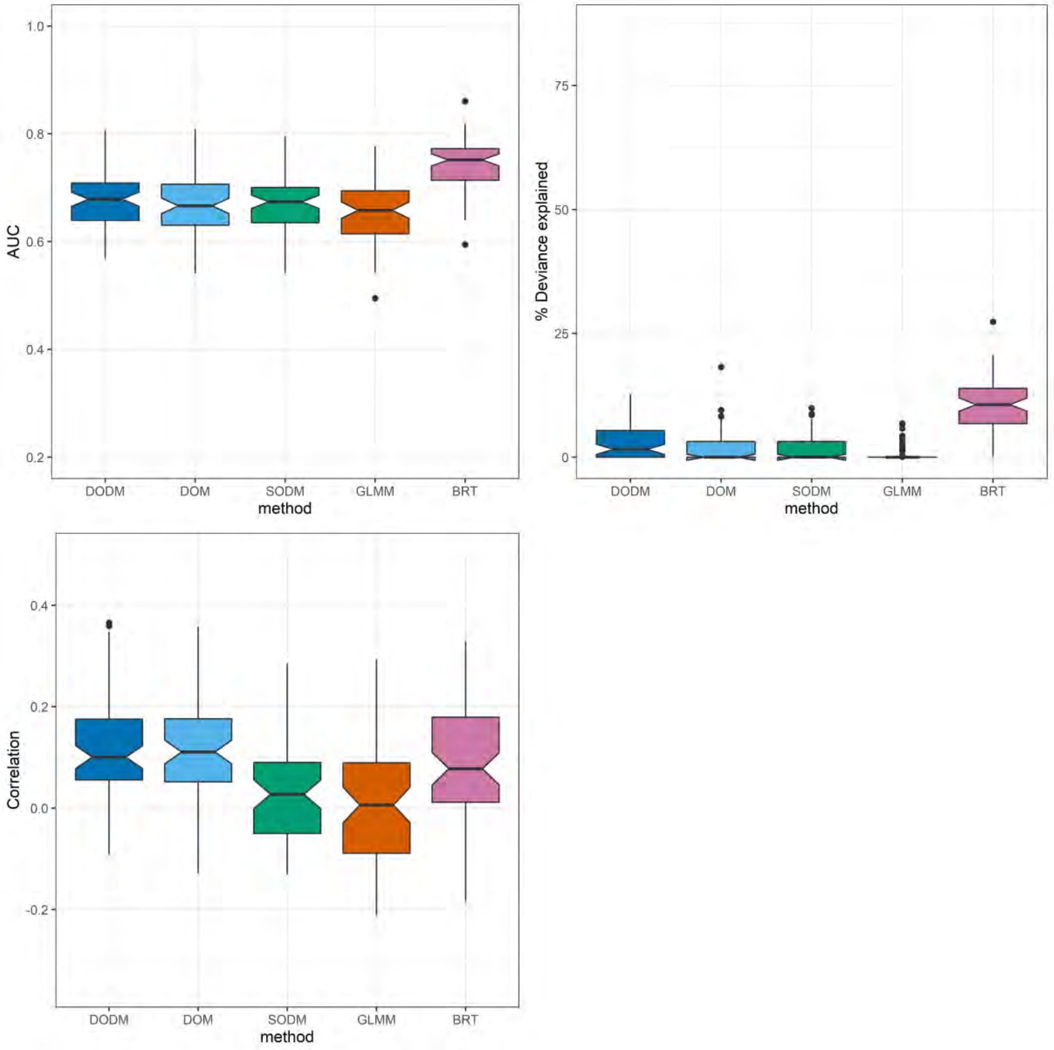
gcb_15723_f1.png



gcb_15723_f2.png



gcb_15723_f3.png



gcb_15723_f4.png