

# Comparing covariate selection strategies for predictive dynamic occupancy models

---

## Authors

Saoirse Kelleher<sup>1</sup>, Natalie Briscoe<sup>1</sup>, Jane Elith<sup>1</sup>, Marc Kéry<sup>2</sup>, Damaris Zurell<sup>3</sup>, Gurutzeta Guillera-Arroita<sup>4,1</sup>

<sup>1</sup>School of Agriculture, Food, and Ecosystem Sciences, University of Melbourne, Parkville, Victoria, Australia

<sup>2</sup>Swiss Ornithological Institute, 6204 Sempach, Switzerland

<sup>3</sup>University of Potsdam, Institute of Biochemistry and Biology, Potsdam, Germany

<sup>4</sup>Pyrenean Institute of Ecology, Spanish National Research Council (CSIC), Jaca, Spain

## Corresponding author

Saoirse Kelleher ([Saoirse.kelleher@unimelb.edu.au](mailto:Saoirse.kelleher@unimelb.edu.au))

## Abstract

Predicting how species will redistribute under global environmental change is a prerequisite for effective conservation. Dynamic occupancy models (DOMs) are increasingly popular tools for modelling range dynamics. DOMs estimate local colonisation and extinction probabilities while accounting for imperfect detection to capture the processes behind changing patterns of species occurrence. Covariate selection is particularly challenging for DOMs because four key parameters may be modelled as a function of covariates. Many techniques have been used for covariate selection with DOMs, but the relative merits of these for predictive performance have not been examined. We used long-term breeding bird survey data from Switzerland for 10 contrasting species to compare different covariate selection strategies in terms of the predictive performance of resulting models, computational efficiency, and consistency of estimated relationships. Using a shared set of candidate covariates, we attempted 39 alternative covariate selection workflows for each species, predicted site occupancy to hold-out years, and compared the outputs of each workflow. Covariate selection strategy affected predictive performance, although there was no overall best approach and differences tended to be relatively small. The largest drops in performance occurred when preliminary filtering stages of selection were used to reduce the size of the covariate pool. Selection strategies also resulted in different spatial predictions of site occupancy for some species, and methods were not consistent in which covariates were identified as important. In cases where DOMs are used to inform decision-making, covariate selection strategies should be applied in ways that capture the uncertainty around which model structure is best, and estimated ecological relationships should always be interpreted with caution. We encourage users to avoid unnecessarily filtering covariates prior to conducting full covariate selection to permit sufficiently complex models if supported by the available data.

### Data availability statement

Data and code used for this paper are available at <https://figshare.com/s/2a54afd1445027b6450d>. This includes the *sbmodels* R package containing functions for fitting and running models, and the *sbSelection* R project containing the runscripts, model outputs, and analysis code.

### Keywords

Covariate selection, model complexity, occupancy models, predictive modelling, process explicit model, species distribution, regularisation

## 1 Introduction

2 Dynamic occupancy models (DOMs) are increasingly valued tools for predicting patterns of species  
3 occurrence over space and time because they explicitly consider processes of local extinction and  
4 colonisation while still being readily fit to monitoring data (Kelleher et al. 2025). Introduced two  
5 decades ago by MacKenzie et al. (2003), DOMs use a hierarchical structure to jointly model the  
6 ecological processes describing how sites transition between occupied and unoccupied states, and  
7 the observation process determining whether a species is detected at occupied sites. DOMs have  
8 important advantages over other tools for modelling distributions – accounting for imperfect  
9 detection mitigates a common source of bias in field collected data (Chen et al. 2013, Kellner and  
10 Swihart 2014, Guillera-Arroita 2017), and explicitly estimating colonisation and local extinction  
11 processes improves predictions when distributions are not at equilibrium (Yackulic et al. 2015).  
12 They may help diagnose demographic problems in species of conservation concerns years before a  
13 decline becomes apparent (van Strien et al. 2011). However, incorporating these processes requires  
14 a model structure more complex than that of correlative species distribution models (SDMs, Briscoe  
15 et al. 2021). Typical DOMs include four main parameters which must be estimated: initial  
16 occupancy ( $\psi_1$ ), the probability that sites are occupied in the first time-step; local colonisation ( $\gamma$ )  
17 and local extinction ( $\epsilon$ ), the probabilities that sites become occupied or unoccupied between time-  
18 steps dependent on the prior occupancy state; and detectability ( $p$ ), the probability that a species is

observed during a survey at an occupied site (MacKenzie et al. 2003). Each of these processes can be expressed as a function of covariates to account for heterogeneity (MacKenzie et al. 2017).

Approaches to fitting any model must be guided by a study's objectives, whether that is to make accurate predictions or to make inference on specific relationships between parameters (Tredennick et al. 2021). In complex systems, the number of covariates needed to reasonably predict occupancy and its dynamics may exceed the number of parameters that can be reliably fit with available observation data (Araújo and Guisan 2006), and determining which covariates to include is a long-standing challenge (Burnham and Anderson 2002, Johnson and Omland 2004, Tredennick et al. 2021). Typical modelling workflows thus attempt to first identify candidate covariates with plausible effects on the response variable, then identify a final model (or final set of models). Covariate selection can be approached in many ways, from exhaustive comparisons of all possible combinations to stepwise approaches that directly test only a subset of plausible candidate models (Kelleher et al. 2025). This exercise can be particularly difficult for complex models like DOMs, where candidate covariates and relationships must be identified for four separate parameter types, typically resulting in a much larger pool of candidate models than for simpler models like SDMs. For a simple linear model, a pool of six candidate covariates can yield  $2^6 = 64$  possible combinations. For a DOM, a pool of six covariates for *each* of the four processes would result in 16 million possible combinations.

While many covariate selection strategies have been attempted for DOMs, little guidance exists on their relative merits. DOMs can be fit by both maximum likelihood estimation (MLE) and Bayesian methods (Kelleher et al. 2025). Most methods fit by MLE rely on comparisons of candidate models based on the relative fit of the model to training data by Akaike's information criterion (AIC, Akaike 1974). The particulars of these methods vary considerably: some compare a set of candidate models defined *a priori* (Mortelliti and Boitani 2007), while others apply sequential approaches to

determine an optimal structure one parameter type at a time (e.g., identifying the optimal structure for  $p$ , followed by  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ ; Duggan et al. 2011). A previous study comparing sequential covariate selection methods for DOMs using empirical data demonstrated that variation of the order in which parameter types were optimized lead to different final models, often failing to identify the lowest possible AIC (Morin et al. 2020). Research on covariate selection for Bayesian DOMs is scant, and methods include iterative approaches to remove non-informative covariates (Cook et al. 2022), comparisons using Bayesian information criteria (Rota et al. 2009) or leave-one-out cross-validation (Urban et al. 2023) and reversible jump Markov chain Monte-Carlo (Brown et al. 2014). While not strictly covariate selection, regularisation approaches can fulfill a similar role by shrinking the influence of non-informative covariates. These can be easily implemented in Bayesian settings as described in Hooten and Hobbs 2015, but to our knowledge have not yet been applied to (conventional) DOMs. Throughout this article we use the term ‘covariate selection’ loosely to refer to both conventional methods and regularisation.

Recent commentary has highlighted the importance of aligning covariate selection methods with a model’s intended use-case (Tredennick et al. 2021). However, study of covariate selection for DOMs has been minimal even as their usage has diversified. Early implementations of DOMs were mostly used to identify the drivers of occupancy dynamics or to make inference on a species’ relationship to environmental factors (Kelleher et al. 2025), however, since then several studies have highlighted the potential of DOMs to predict occupancy into the future or to unsurveyed locations (Kéry et al. 2013, Briscoe et al. 2021). These prediction-focused applications are increasingly popular (Kelleher et al. 2025), emphasising the need for covariate selection strategies that can reliably identify models with good predictive performance while remaining computationally tractable.

## 66 **Objectives**

67 Here, we compare covariate selection methods for dynamic occupancy models, with an emphasis on  
68 models used for prediction. Using empirical data from a long-term breeding bird monitoring  
69 program in Switzerland, we fit DOMs under a range of MLE and Bayesian covariate  
70 selection/regularisation workflows and compare:

- 71 i. computational cost;
- 72 ii. predictive performance at surveyed sites in hold-out years;
- 73 iii. variation in predictions of site occupancy dynamics at unsurveyed sites; and
- 74 iv. covariates selected and estimated coefficients.

## 75 **Methods**

76 Using detection data from a long-running breeding bird monitoring program in Switzerland, we  
77 built DOMs using 39 alternative covariate selection workflows for ten species chosen *a priori* for  
78 their contrasting occurrence patterns and characteristics (Figure 1). Each covariate selection  
79 workflow began with the same pool of candidate covariates and involved two stages to identify a  
80 final model, which was used to predict annual occupancy probabilities across Switzerland.  
81 Predictive performance was formally compared with mixed models regressing performance  
82 metrics against covariate selection approaches. We also assessed the degree of agreement in  
83 distribution maps, compared the regression coefficients of selected covariates, and benchmarked  
84 computational times (Figure 1).

## 85 **Dynamic occupancy models**

86 We used the original formulation of the DOM as defined in MacKenzie et al. (2003). In each time-  
87 step  $t$  (termed a ‘primary occasion’), each site  $i$  exists in either an occupied ( $z_{i,t} = 1$ ) or unoccupied

state ( $z_{i,t} = 0$ ). The occupancy state in the first primary occasion is determined by a Bernoulli trial with initial occupancy probability  $\psi_{i,t=1}$ . In all subsequent primary occasions, the occupancy state is determined by a Markovian process governed by a colonisation probability  $\gamma_{i,t}$  and a local extinction probability  $\epsilon_{i,t}$ :

$$\begin{aligned} z_{i,t=1} &\sim \text{Bern}(\psi_{i,t=1}), \\ z_{i,t>1} &\sim \text{Bern}(z_{i,t-1}(1 - \epsilon_{i,t-1}) + (1 - z_{i,t-1})\gamma_{i,t-1}). \end{aligned}$$

Within each primary occasion, multiple surveys  $j$ , termed ‘secondary occasions’, are conducted and in each, the species is either detected ( $y_{i,t,j} = 1$ ) or not ( $y_{i,t,j} = 0$ ) based on a conditional detection probability  $p_{i,t,j}$ .

$$y_{i,t,j} \sim \text{Bern}(z_{i,t} * p_{i,t,j}).$$

For brevity, hereafter we refer to the four parameter types as  $\psi$ ,  $\gamma$ ,  $\epsilon$ , and  $p$ . Each of these probabilities can be modelled by covariates, which conventionally are incorporated through a logit link function (MacKenzie et al. 2017). This model assumes that the occupancy state of a site cannot change within, but only between, primary occasions (closure assumption). It also assumes that there are no false positive detections (MacKenzie et al. 2017).

### *Swiss breeding bird data*

The Swiss Ornithological Institute’s breeding bird survey has a rich history of use as a case-study for hierarchical modelling methods (Schmid et al. 2004, Kéry and Royle 2016, Strebel et al. 2026). The monitoring program consists of 2-3 surveys per year conducted at established sites distributed around Switzerland (supplementary material section 1.1). Each site is a 1 square kilometre grid cell, and in each survey trained volunteers follow pre-designated transects throughout the site and record all observed breeding birds. In this study, we used species detection/non-detection data collected between 2000-2018, with data from the years 2000-2008 used for model fitting and data

from 2009-2018 used for model testing. There were 266 sites with data available for both time periods. To avoid violating the closure assumption, either by surveying prior to a migratory bird species' arrival or by counting transient individuals passing through sites during migration, we removed surveys before a species-specific date marking the end of spring migration. We chose ten species in this case study for their diversity in life history traits, known distributions within Switzerland, and trends in the number of yearly detections during the study period (Table 1).

**Table 1: Details of the 10 focal species included in our comparisons. Survey count details the total number of surveys included for each species between 2000 and 2018. Mean observed occupancy is the proportion of sites with surveys where the species was observed in each year averaged across all years, mean detectability was the survey detection probability estimated using a simple dynamic occupancy model for all years 2000-2018. Observed colonization events are the number of cases where a species was observed in one year but not the next; observed extinctions are the inverse – these are totalled for the training period (2000-2008) and testing period (2009-2018).**

| Species                                                 | Survey count | Mean observed occupancy | Mean detectability | Observed colonisations (training/testing) | Observed extinctions (training/testing) |
|---------------------------------------------------------|--------------|-------------------------|--------------------|-------------------------------------------|-----------------------------------------|
| Common kestrel (COKE, <i>Falco tinnunculus</i> )        | 14125        | 0.41                    | 0.44               | 288/414                                   | 246/390                                 |
| European green woodpecker (EGWO, <i>Picus viridis</i> ) | 14125        | 0.46                    | 0.57               | 220/310                                   | 209/286                                 |
| Eurasian skylark (EUSK, <i>Alauda arvensis</i> )        | 14125        | 0.22                    | 0.80               | 73/71                                     | 80/83                                   |
| Garden warbler (GAWA, <i>Sylvia borin</i> )             | 10313        | 0.40                    | 0.68               | 191/229                                   | 203/245                                 |
| Mallard (MALL, <i>Anas platyrhynchos</i> )              | 14125        | 0.20                    | 0.56               | 96/135                                    | 83/136                                  |
| Red crossbill (RECR, <i>Loxia curvirostra</i> )         | 14125        | 0.32                    | 0.51               | 238/291                                   | 239/308                                 |

|                                                       |       |      |      |         |         |
|-------------------------------------------------------|-------|------|------|---------|---------|
| Red kite (REKI, <i>Milvus milvus</i> )                | 14125 | 0.28 | 0.53 | 134/166 | 88/150  |
| Red-backed shrike (RBSH, <i>Lanius collurio</i> )     | 9377  | 0.21 | 0.63 | 131/175 | 135/192 |
| Whinchat (WHIN, <i>Saxicola rubetra</i> )             | 8362  | 0.12 | 0.77 | 51/58   | 52/161  |
| White-throated dipper (WTDI, <i>Cinclus cinclus</i> ) | 14125 | 0.13 | 0.39 | 90/130  | 87/120  |

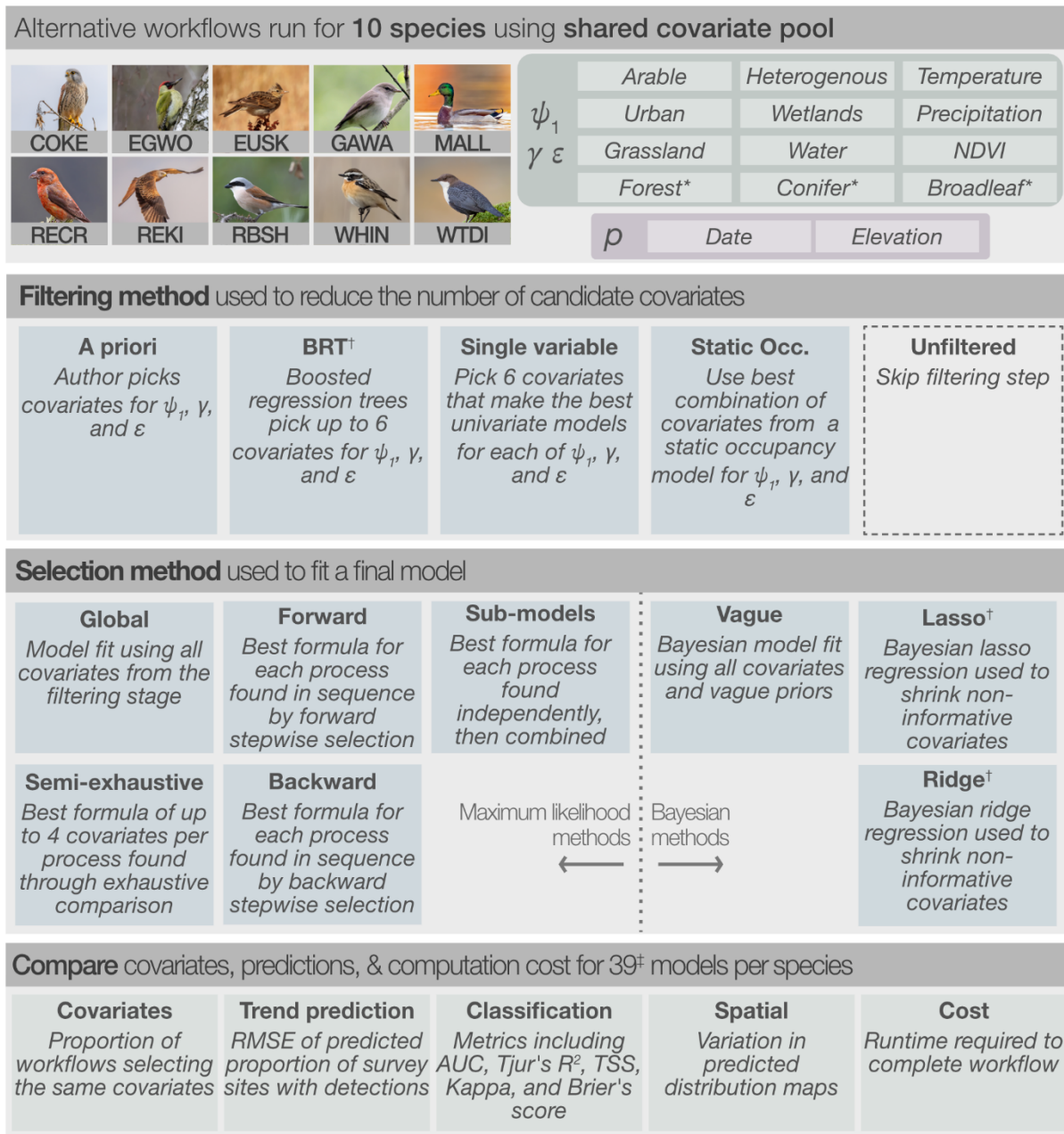


**Figure 1: Diagram of methods for the comparison of single-species dynamic occupancy model covariate selection workflows.** We ran 39 alternative covariate selection workflows for ten focal species, using a shared pool of candidate covariates. Each workflow consisted of one filtering method used to reduce the size of the initial candidate covariate pool, and one selection method used to identify a final model. The covariates selected, predictions, and computational cost were compared between each of the alternative workflows.

\*Conifer and Broadleaf covariates replaced Forest for one species, the Red Crossbill.

†Indicates methods not previously used for DOMs.

#5 filtering methods x 8 selection methods; Unfiltered+Semi-exhaustive workflow not attempted due to computational cost.

## 124 *Candidate covariates*

125 All modelling workflows began with a shared pool of candidate covariates representing common  
126 environmental drivers of avian occupancy. For the three occupancy-related parameter types ( $\psi_1$ ,  $\gamma$ ,  
127 and  $\varepsilon$ ), we used a pool of 12 covariates that represented land cover, vegetation, climate, and water  
128 availability, with the data summarized in a manner relevant to bird ecology and additional  
129 covariates discarded due to high collinearity (Pearson's coefficient  $<0.7$ ; Supplementary Material  
130 2). Covariates on the occupancy process parameters were always modelled with a linear and  
131 quadratic term as we expected that non-linear responses were plausible for all covariates; while  
132 this does limit flexibility, we considered that using only linear terms could be ecologically  
133 unrealistic and that considering linear and quadratic terms independently would make the  
134 covariate pool unworkably large.

135 Our primary objective was to compare the performance of different covariate selection approaches,  
136 focusing on the occupancy-related parameters. Hence, we considered only a simplified set of  
137 candidate covariates for  $p$  used in previous analyses of these data (Kéry et al. 2013, Briscoe et al.  
138 2021). We considered five alternative structures for  $p$ : intercept only,  $\sim \text{DATE}$ ,  $\sim \text{DATE} + \text{DATE}^2$ ,  $\sim$   
139  $\text{DATE} + \text{ELEV} + \text{DATE} * \text{ELEV}$ , and  $\sim \text{DATE} + \text{DATE}^2 + \text{ELEV} + \text{DATE} * \text{ELEV}$ . The covariate ELEV was  
140 only included in a model if it was selected as part of an interaction with DATE. All covariates were  
141 z-transformed using the mean and standard deviation of the training data.

142

## Modelling workflows

Our covariate selection workflows comprised two stages: an initial ‘filtering’ stage where the initial pool of candidate covariates was reduced, and a second ‘selection’ stage where a final model was identified. This two-stage approach allowed flexible combinations of methods and resembled common practice in the literature (Kelleher et al. 2025). We chose methods seen in the existing literature on DOMs plus some approaches not yet used for DOMs. Unless otherwise stated, MLE approaches were fit in the *unmarked* R package (v.1.5.0.9002, Kellner et al. 2023) and Bayesian models were fit in JAGS (v.4.2.0, Plummer et al. 2016) using the *autojags* function from package *jagsUI* (v.1.6.2, Kellner 2015), with steps of 10,000 MCMC samples. Brief details are outlined below, with additional information and code in the supplementary materials (Section 3).

*Filtering stage methods:* We applied one of five methods to the initial covariate pools for  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ . Aside from a ‘no filtering’ option, all filtering methods were permitted to retain a maximum of six covariates per process to keep comparisons as fair as possible. We did not filter the covariate pool for  $p$  as it was much smaller than that for the other processes.

- ***A priori:*** Users of DOMs often simply choose a model *a priori* (Kelleher et al. 2025, e.g. Scott and Rissler (2015)). To replicate this approach, the author most familiar with the study system (redacted for peer review) chose the covariates they believed would produce the best model.
- **Boosted regression trees (BRTs):** A machine learning approach with strong performance in variable selection (Elith et al. 2008), BRTs have been used to fit static occupancy models directly (Hutchinson et al. 2011) although we have not seen prior applications using them to select covariates for these models. Using binarised outcomes of observed initial occupancy states, possible colonisation events, and possible

extinctions as response variables, we fit BRTs using the *dismo* R package (v1.3-16, Hijmans et al. 2017) to identify up to six covariates per process rate based on variable importance.

- **Single variable:** For each covariate, we fit models with only a single covariate on one of  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ , and the global formula for  $p$  (Studholme et al. 2023). We then compared the AIC for each parameter/covariate combination and retained the six covariates with the lowest AIC per parameter.
- **Static occupancy:** Static occupancy models were used to identify relevant covariates (Smith et al. 2012). Using detection histories from the first primary occasion in 2000, we fit models with all combinations of up to six covariates on the occupancy process and with the global formula for  $p$ . We retained the covariates used for the occupancy parameter in the lowest AIC model for all occupancy-related parameter types ( $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ ).
- **Unfiltered:** The full covariate pool was retained for the ‘selection’ stage.

*Selection stage methods:* Eight methods from MLE and Bayesian frameworks were used to fit and select a single final DOM considering the covariates retained from the filtering stage. We allowed methods to select from the five possible formulae for  $p$  in addition to selecting for  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ .

- **Global (MLE):** A model was fit with all covariates retained after filtering for  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ , and the global formula for  $p$ .
- **Forward sequential (MLE):** We first fit an intercept-only model. A stepwise process was then used for  $p$ , where models were fit adding one covariate at a time and retaining the covariate that most reduced AIC. This was repeated until further additions no longer improved AIC; the structure for  $p$  was then fixed. The same process was then repeated for

$\psi_1$ ,  $\gamma$ , and  $\varepsilon$ , in that order. To confirm that this parameter fixing order was competitive with other possibilities we tested other possible orders (Supplement 5.1), but here report only on  $p/\psi_1/\gamma/\varepsilon$ .

- **Backward sequential (MLE):** We first fit a global model with all retained covariates and the global formula for  $p$ , then used a stepwise process to test removal of one covariate at a time for  $p$  until further removals no longer reduced the AIC. The structure for  $p$  was then fixed and the process repeated was for  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$ , in that order.
- **Submodels (MLE):** An alternative approach described in Bromaghin et al. (2013) and Morin et al. (2020) is to identify the optimal structure for each process in isolation, rather than in sequence. Hence, we fit all possible combinations of up to 6 covariates for each of  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$  while using null structures on the other parameter types. For  $p$ , all five candidate formulae were tested in the same manner. The formulae that led to the lowest-AIC model for each parameter type were combined in a final model.
- **Semi-exhaustive (MLE):** While fully exhaustive comparisons of all possible combinations of covariates remain infeasible even with the reduced pool of covariates, semi-exhaustive comparisons of all models with a constrained number of covariates are an alternative (Briscoe et al. 2021). We fit all possible DOMs that included up to four covariates on each of  $\psi_1$ ,  $\gamma$ , and  $\varepsilon$  and any of the five candidate formulae for  $p$ . We then chose the final model based on AIC.
- **Vague (Bayes):** We fit a global Bayesian model with vague priors; effectively numerically identical to a global model fit by MLE.
- **Lasso (Bayes):** Lasso regularisation is used to shrink the regression coefficients of non-informative covariates towards zero, effectively applying a penalty term that controls the

bias-variance trade-off. While not strictly covariate selection, regularisation fulfills a similar role by reducing the influence of non-informative covariates. A Bayesian implementation of the Lasso has been proposed using a shared double-exponential prior with precision set to the penalty parameter for each regression coefficient (Park and Casella 2008); rather than tuning the penalty parameter by cross-validation, we tuned it within the model by incorporating it in a hyper-prior as described by van Erp et al. (2019). To confirm that this formulation allowed for the optimal hyper-prior, we also tested alternatives, (Supplement 5.2).

- **Ridge (Bayes):** Ridge regularisation is another common technique to reduce the influence of uninformative covariates (Hoerl and Kennard 1970) that does not pull covariates as strongly towards zero compared to the Lasso. A similar Bayesian equivalent has been proposed for ridge regularisation using normal rather than double-exponential priors; we implemented this in the same manner as the lasso regression with the penalty parameter included as a hyper-prior in the model (van Erp et al. 2019).

We attempted all combinations of the five filtering methods and eight selection methods aside from the prohibitively computationally expensive combination of ‘no filtering’ and ‘semi-exhaustive,’ resulting in 39 covariate selection workflows per species. Most workflows were run in parallel on identical virtual machines (details in Supplementary Material section 3.2), thus computation time should be regarded comparatively rather than as representative of actual computing time on a standard laptop or desktop computer. The exception, semi-exhaustive models, required a high-performance computing platform with a sample re-run on the virtual machines to allow comparisons of run-times.

## 234 *Predictions*

235 After a final model was identified for each workflow, we predicted occupancy probabilities for all  
236 1x1 square kilometre grid cells in Switzerland for each year from 2000-2018. As point estimates,  
237 we used either the MLE point estimates or Bayesian posterior medians.

238 For the 266 survey sites, we also calculated the probability of observing the species at each site  
239 during each primary occasion from 2000-2018, given the number of surveys that were conducted at  
240 that site, in that primary occasion:

$$241 \quad pr(Ob_{s_{i,t}}) = \psi_{i,t} \prod_{j=1}^J (1 - p_{i,t,j})$$

242 This probability was used as the basis for assessing the performance of the predictions because the  
243 actual variable of interest – site occupancy – is a latent state that is imperfectly observed.

## 244 *Model comparison and performance evaluation*

245 We compared the performance of final models in predicting country-wide trends in observed  
246 occupancy as well as site-level predictions. For the former, we calculated the normalised root-  
247 mean-squared-error (RMSE) for the predicted number of sites with at least one detection per  
248 primary occasion against the actual number of sites with detections (normalising by the number of  
249 sites with surveys); for the latter, we calculated five metrics (AUC, Tjur's  $R^2$ , True Skill Statistic,  
250 Cohen's Kappa, and Brier score; Abrego and Ovaskainen 2023) comparing the predicted probability  
251 of recording a species at a site in a given primary occasion against the outcome of the surveys. In  
252 both cases metrics were calculated for both the set of years used to fit the model and those reserved  
253 for testing. For each metric, we then fit a mixed model in *brms* (v 2.230; Bürkner 2017) regressing  
254 that metric (on the logit scale, where applicable) against random intercepts for each species and

fixed effects for each of the filtering and selection methods (e.g.,  $\text{logit}(\text{AUC}) \sim (1|\text{Species}) + \text{Filter} + \text{Select}$ ).

While we focused on future prediction and did not reserve data for spatial cross-validation, we did compare maps of occupancy predictions calculated for all sites from 2000-2018 to determine whether different workflows produced materially different predicted distributions. We also compared estimates of the regression coefficients among different workflows for the same species. We considered there was evidence of covariate effects if they were included in a model and their confidence or credible intervals did not include zero.

## Results

### *Workflow success rate and computation speed*

Of the 390 possible workflow-species combinations, we were successful in fitting final models for 296 (Supplementary material Table S5-1); for the remaining cases, models either did not converge after 20 numerical optimization attempts (MLE) or retained  $\hat{R}$  values greater than 1.1 for estimated parameters after 500,000 MCMC iterations (Bayesian). Failures generally occurred using methods that required fitting complex models with many covariates to species with limited observations. The *Unfiltered+Global* and *Unfiltered+Backward* workflows failed for all species — each of these required fitting a fully global model with dozens of terms, unsurprisingly resulting in failures to converge. Of the MLE selection methods, only *Forward* was successful for all species, largely because it begins with a null model and does not strictly require greater complexity. All frequentist workflows except *SemiExhaustive* had median runtimes below 100 minutes, whereas all Bayesian workflows had median runtimes between 116 and 1000 minutes. While the *SemiExhaustive* method was run using a high-performance computing system, based on samples using the same virtual

277 machine as the other workflows, projected runtimes were 6-60 days per species. In this study we  
278 used an updated and substantially faster version of function *colex* (revised in v.1.5.0.9002).

### 279 *Predictive performance*

280 Our first comparison of predictive performance examined the models' ability to estimate the  
281 proportion of surveyed sites with detections in each year (Figure 2). The range in these predictions  
282 across workflows varied between species, although nearly all workflows predicted the overall  
283 trajectory during the testing period similarly well (2009-2018). Nonetheless, in some cases the  
284 most and least optimistic predictions for a species were separated by large margins by the end of  
285 the testing period — for example, estimates for the number of sites with Eurasian Skylark  
286 observations in 2018 ranged from 42 to 63.

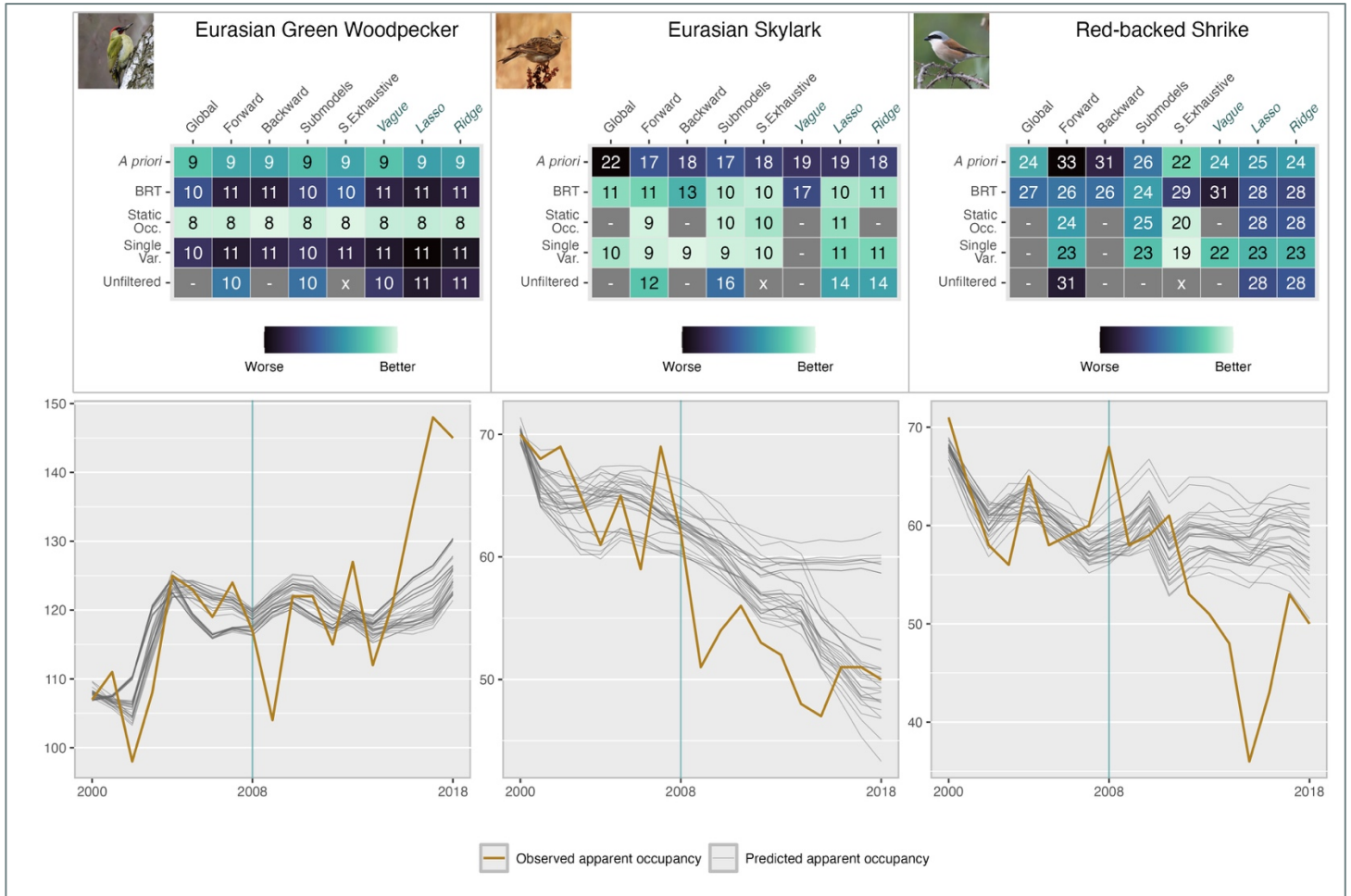


**Figure 2: Comparisons of predictions for three of the ten focal species. The top row gives the RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018), normalised by the number of surveys in each year; i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% of the surveyed sites in each year. ‘-’ indicates workflows that could not be successfully completed, and ‘x’ represents workflows that were not attempted. The bottom row compares the number of the 266 1km<sup>2</sup> sites where a species was detected in each year (gold lines) against the number of observed sites predicted by the final model from each workflow (gray lines). Dynamic occupancy models were fit to data from the Swiss common breeding bird survey from 2000 through 2008, indicated by the blue vertical line. Note that the y-axes are species-specific due to variation in prevalence. Remaining species presented in supplementary materials Figures S5-1 to S5-9.**

Comparing the normalised RMSE of these predictions during the testing period did not yield any clear patterns in predictive performance across workflows (Figure 2, remaining species in supplementary material figures S5-1 to S5-12). Differences in performance were relatively marginal for Eurasian Green Woodpecker (RMSE from 8-11%), but considerably larger for Eurasian Skylark (9-22%) or Red-backed Shrike (19-33%). No workflow offered uniformly superior or inferior performance in trend prediction across species (Figure 3).

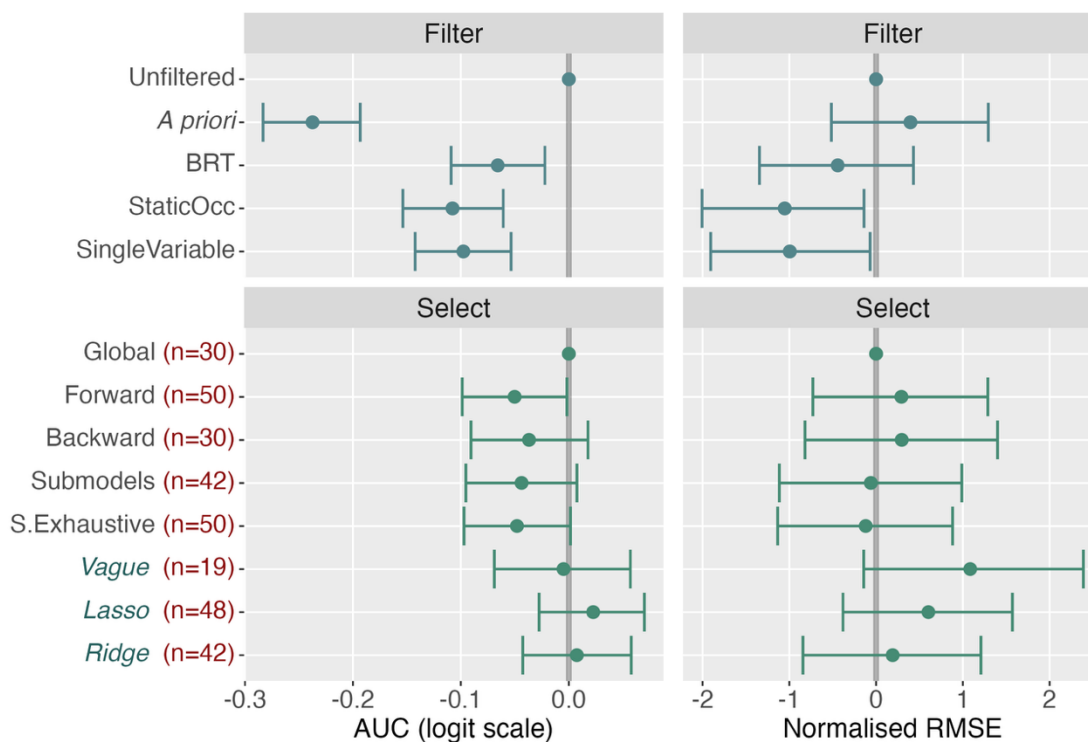


**Figure 3: Comparative predictive performance of dynamic occupancy model covariate selection workflows across 10 common breeding birds in Switzerland. Coefficients relating each filtering & selection method to AUC (on the logit scale) and normalised RMSE, based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.**

Conversely, differences between methods in terms of site-level predictive performance were more consistent (Figure 3). GLMs fit for the five metrics of site-level predictive performance gave similar results, so we present only those for AUC and RMSE in the main text; the other four are available in Supplementary materials. Differences in AUC were more strongly associated with differences in the filtering method rather than the selection method, with all filtering approaches resulting in lower

AUC compared to *Unfiltered* (Figure 3). All within-group differences in AUC due to covariate selection were dwarfed by between-group variation in AUC between species. These differences were not clearly associated with either the apparent prevalence or apparent number of transitions for a species and are presumably better explained by other facets of the species' biology and the pool of covariates available.

Values of all metrics declined between the training and testing period for many species (Supplement figure S6-1), suggesting that models may have missed or mis-estimated key dynamics occurring during the testing period. This is particularly noticeable for the White-throated dipper, Garden Warbler, and Common Kestrel, which had consistently low-performing models that each exhibited a large drop in testing versus training AUC (Supplement figure S6-1). The best and worst performing workflows by AUC were relatively consistent across all species and metrics. This contrasts with the differences in RMSE for predicted trends in occupancy, where the same workflows did not consistently achieve the best performance for each species. For site-level predictions, the most consistently successful workflows combined the *Unfiltered* method in the first stage with the Lasso and Ridge Bayesian approaches (Figure 4). The *a priori* filtering methods had comparatively lower predictive performance for most species. The remaining filtering methods fell between these, with none of the BRT, StaticOcc, or SingleVariable filtering methods consistently outperforming the others. However, the BRT method was frequently better than the other two.

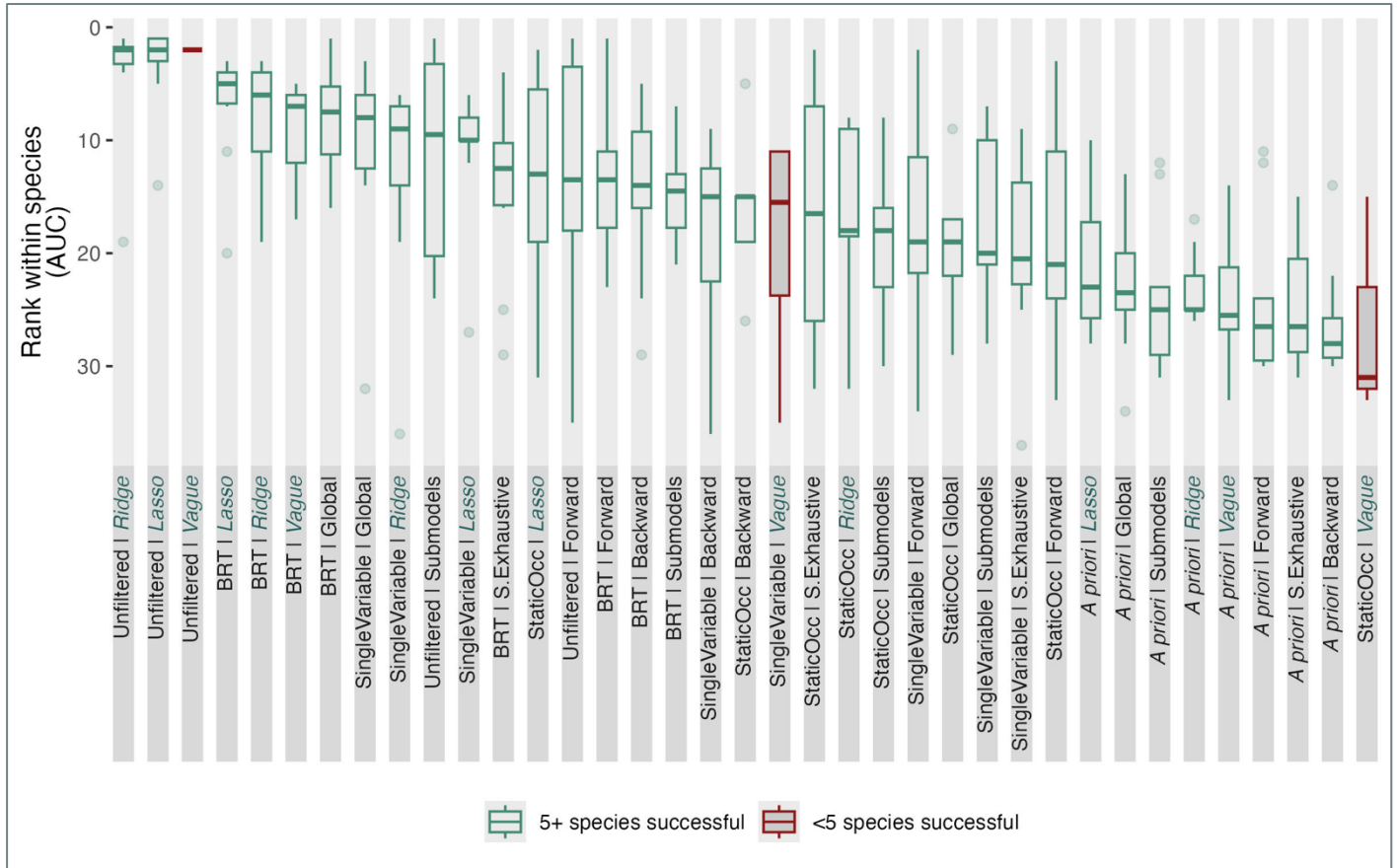


Figure 4: Dynamic occupancy modelling workflows ordered by their median rank in predictive performance as measured by AUC across 5 10 species. Each data-point represents a workflow’s rank within a species on one metric; e.g., ‘Unfiltered+Vague’ had the 1st ranked AUC for the Eurasian Green Woodpecker. Workflows that could not be completed for at least 5 species are highlighted in red.

### 322 Spatial predictions

323 Predictions of site occupancy across Switzerland between 2000-2018 were not spatially consistent

324 for some species (Figure 5). Differences were largest for the Garden Warbler and Mallard, each of

325 which have sizeable areas where different models predicted very high or low occupancy

326 probabilities.

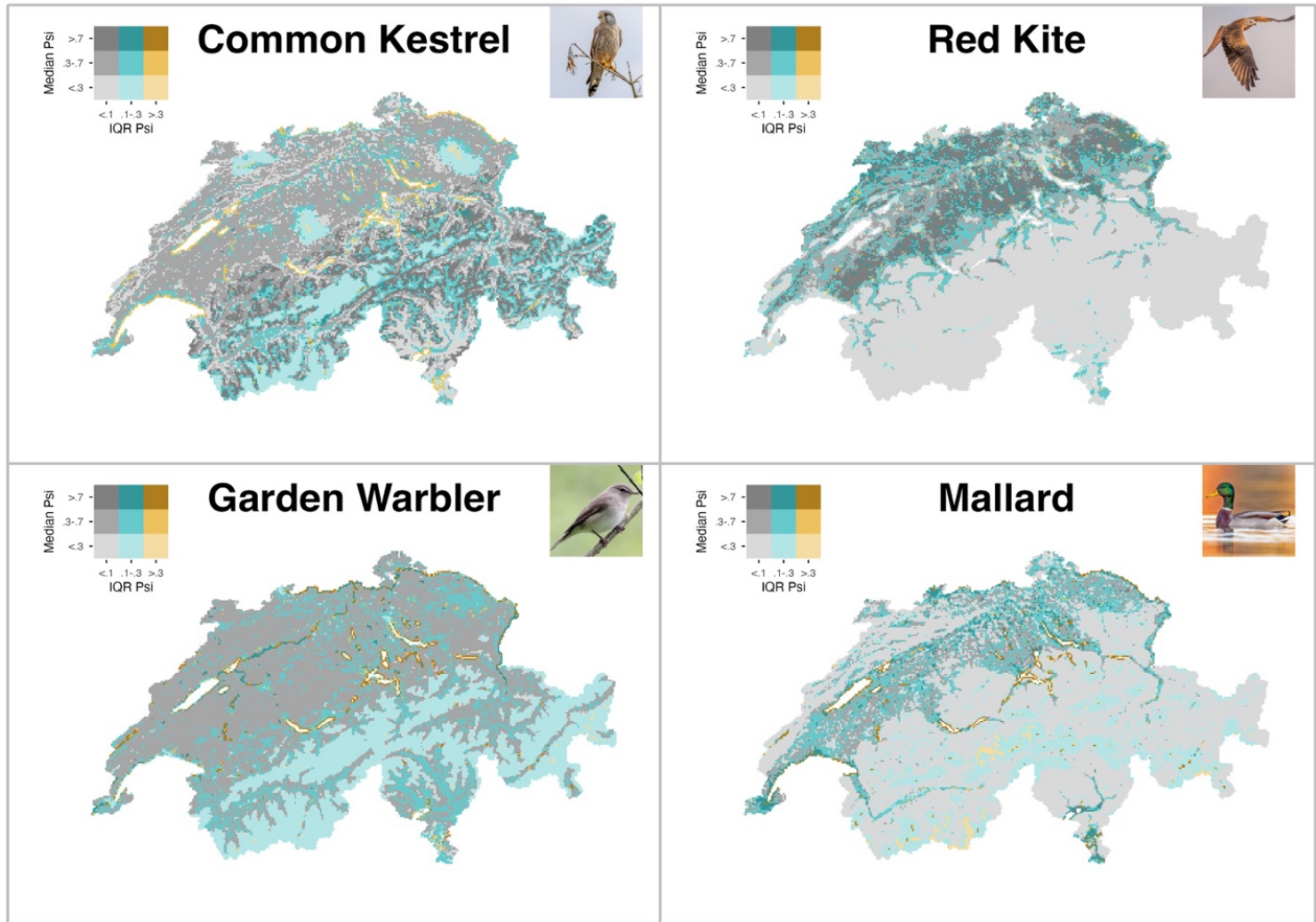


Figure 5: Consistency of occupancy estimates across workflows for a selection of species. Each map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities. Remaining species are available in the supplementary materials (Figures S5-11 to S5-20).

### 327 *Covariate inclusion*

328 For all species, models varied considerably in which covariates showed evidence of covariate  
 329 effects for the occupancy parameter types (Figure 6), even for the species where workflows had

330 similar predicted trends, distributions, and performance. Only six covariate-parameter  
331 relationships showed consistent evidence of covariate effects across all workflows for a species.

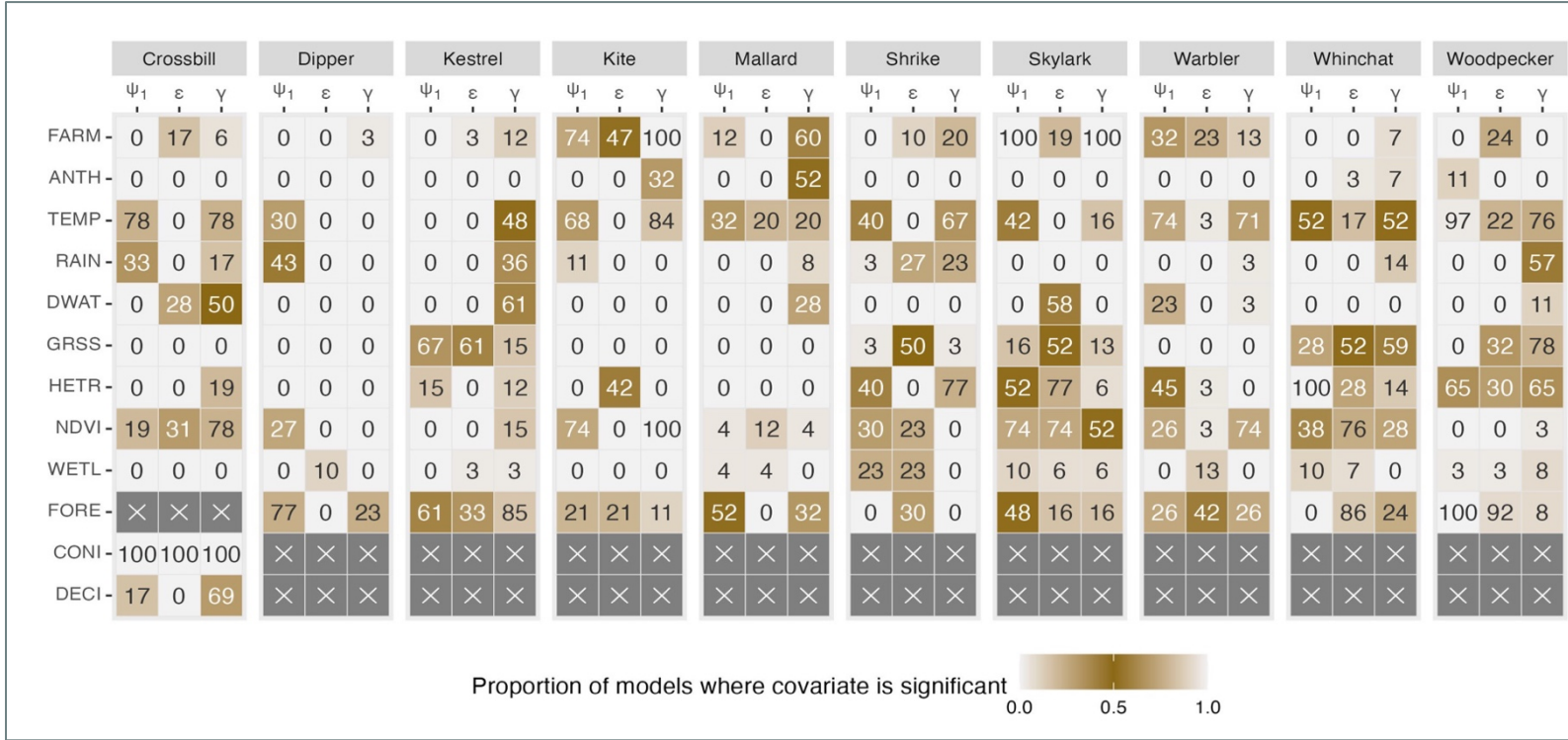


**Figure 6: Proportion of workflows where evidence of a covariate effect was found for each coefficient.**

Evidence of a covariate effect was assumed if the covariate was included in the model and the 95% confidence or credible interval for either the linear or quadratic component did not overlap zero. Cells with values of 0 or 100 indicate complete agreement between models on whether a covariate has an effect, hence the colour scale shows both as paler. 'X' represents covariates that were not included in the candidate covariate pool for a species.

## Discussion

Clear guidance for modelling workflows can ease model application for practitioners and improve outcomes. We tested a wide range of methods for selecting covariates in dynamic occupancy models (DOMs; MacKenzie et al. 2003) used for prediction. Our results suggest that no covariate selection method is unambiguously superior, and the models produced across workflows differ in terms of the covariates used and spatial and temporal predictions. Modellers need to be aware of this uncertainty, particularly where DOMs are used to guide conservation action, where distribution maps are being produced, and where inference is made on covariates as drivers of occupancy dynamics. While our results do not point towards a single best workflow, they do provide some guidance on approaches that are more or less suitable for certain applications.

Our models produced mixed predictions of the number of sites where each species was detected (Figure 2). Encouragingly, all workflows predicted the same general trends. Mean yearly prediction error was less than 30% for most workflows, with small differences (<20%) between the best and worst predictions for most species. However, models were unable to capture highly stochastic dynamics like abrupt increases or decreases in occupancy and erroneously predicted stable trajectories where these abrupt changes occurred. This may be due to environmental effects not captured by our covariates, such as conservation interventions (Roilo et al. 2024) or masting events (Summers et al. 2024). This underscores the importance of a carefully considered pool of candidate covariates informed by subject-matter knowledge and emphasize that robust covariate selection strategies might identify a 'best', but not necessarily also a 'good' model (MacKenzie and Bailey 2004).

We did not find an optimal method across all species in terms of performance at predicting apparent prevalence across sites (Figure 2). In some cases, groups of workflows performed best for individual species with variation largely driven by the filtering method. This suggests that the

inclusion or exclusion of certain covariates at the filtering stage played a key role in determining predictive performance. Our assessment of variation in site-level discrimination skill between workflows produced more obvious patterns. Methods that used the *a priori* filtering method were most often the lowest ranked, and methods that bypassed the filtering stage (*Unfiltered*) often the best (Figure 4). Our implementation of the *a priori* method was limited as it was intended to simulate execution by a lone researcher, rather than by larger-scale elicitation. Nonetheless, we believe that it can be difficult for any individual, however well-informed, to consistently identify informative covariates – particularly in multispecies studies. This method also typically retained the fewest covariates of any filtering method, which is characteristic of current practice for published DOMs (Kelleher et al. 2025). Restricting the complexity of the final model through strong filtering in this manner may reduce predictive performance, particularly where there is sufficient information to accommodate more complex models. Filtering does often remain necessary (as illustrated by our failure to fit many ‘unfiltered’ workflows), however, it could be limited to reducing the covariate pool to the largest practicable number rather than adopting a highly restrictive approach (Brun et al. 2020).

Among the filtering methods that we tested, the *BRT* approach consistently had reasonable performance and was the method for which the most selection methods could be completed across species. Much of this success can likely be attributed to two factors not related to the *BRTs* themselves: firstly, unlike in the *Single variable* method, fewer covariates per parameter type could be passed to the next stage if fewer covariates were identified as informative, possibly avoiding inclusion of spurious terms. Second, unlike the *Static occupancy* (i.e. single season) method, covariates were chosen for colonisation and extinction independently rather than being shared across all occupancy-related parameter types. While these characteristics could be achieved by other means, *BRTs* offer other advantages — they are efficient and relatively easy to implement,

and extensions to our study could permit identifying interactions between covariates, which is straightforward with BRTs (Elith et al. 2008).

Some methods attempted for the selection stage (including the *Global* and *Backwards* MLE approaches and the vague Bayesian approach) could not always be completed because convergence was not achieved for overly-complex models; conversely, the *Forward* selection method was always achievable but was rarely best-ranked. Morin et al. (2020) have highlighted how sequential selection methods can be inconsistent because the candidate models tested are determined by the order in which parameters are optimized. The *Submodels* method attempted to avoid this by identifying the optimal structure for each process rate independently, but this method could not always be fit and offered only moderate performance. In contrast to the MLE methods, the Lasso and Ridge Bayesian approaches we attempted could be fit for most species and were often among the top-ranked techniques. Beyond simply improving model fit and hedging against over-fitting, another advantage of Bayesian regularisation may simply be that models can be fit without reducing the covariate set and potentially excluding important predictors. While these models are markedly more time-consuming to fit (at least with the tools and convergence requirements we used) and may be more difficult to implement, the performance benefits may be worth the additional effort.

While not the main focus of our study, we found strong variation in the covariates that were identified as exerting considerable effects on the modelled response when using different workflows for the same species, re-emphasising concerns about the reliability of these strategies when attribution of occupancy dynamics to specific environmental factors is desired. For all species, evidence of covariate effects was found in only a portion of workflows; as only one workflow is typically conducted in a study, this inconsistency is unlikely to be noticed or reported. This behaviour is inherent in methods based on models' relative fit using metrics like AIC, where a

well-predicting model does not necessarily mean that variation is correctly apportioned to different covariates (Tredennick et al. 2021). This is also seen in prior studies on occupancy and related capture-recapture models (Doherty et al. 2012, Morin et al. 2020) and can potentially be explained by collinearity and shared drivers for many variables in the candidate set (Stewart et al. 2023). Despite these concerns, covariate selection is still routinely applied in cases where covariate inference is the primary objective (Kelleher et al. 2025). While some form of covariate selection may still be useful for exploratory analyses to screen drivers of occupancy dynamics from a large pool of candidates, statistical evidence of covariate effects should be interpreted with caution.

When comparing spatial predictions across workflows we found that even when metrics of performance assessed at survey sites were similar, distribution maps often showed local differences. When maps are used to support conservation decisions, this variation could determine whether a site is considered within a species' distribution and potentially impact management decisions. To hedge against this uncertainty, modellers could consider applying multiple variants of covariate selection workflows to a given dataset or consider multiple models with comparable support and where possible reserve hold-out data or apply cross-validation to better understand spatial predictive performance (Valavi et al. 2019).

Using the comprehensive Swiss breeding bird dataset, we were able to test models across contrasting species using long-term and high-quality data. While assessing performance of hidden-state models with field data can be challenging, the relatively high cumulative detection probabilities across species make these data particularly suitable for model evaluation and comparison. However, as with any case study results partially reflect the focal taxa and local environment. Our results are also influenced by the methods we considered and how they were implemented – given the wide array of methods used for covariate selection in DOMs (Morin et al. 2020; Kelleher et al., 2026), our aim was to compare a range of contrasting approaches rather than

to optimise individual approaches. However, sensitivity analyses exploring forward-fitting and regularisation methods (where we expected alternative implementations might alter predictive performance) indicate that our implementations performed satisfactorily relative to other possibilities (Supplement figures S5-1 and S5-2). Simulation studies that provide known truth and can explore a broad range of scenarios and levels of data availability could help to optimise individual approaches and further test the generality of our findings.

Absent a clearly optimal method to select covariates for predictive performance, modellers should be aware of potential pitfalls and consider the following recommendations when predictive success is the main objective:

- While it is important that all covariates are ecologically justifiable, filtering candidate covariates too strictly may reduce predictive performance. Filtering should be limited to the maximum computationally practicable number of covariates.
- Regularisation methods offer strong performance, and Bayesian approaches are increasingly accessible. We advise considering these options where the necessary computation resources and expertise are available.
- When filtering is necessary, methods like our BRT-based approach that consider covariates for each parameter independently (rather than screening for initial occupancy, colonisation, and extinction together) may be effective.

Regardless of the method used, robust modelling must be grounded in careful application of ecological theory and be undertaken with sufficient awareness of uncertainties in the covariate selection process.

## References

- 450 Abrego, N. and Ovaskainen, O. 2023. Evaluating the predictive performance of presence–absence  
451 models: Why can the same model appear excellent or poor? - *Ecology and Evolution* 13:  
452 e10784.
- 453 Akaike, H. 1974. A new look at the statistical model identification. - *IEEE Transactions on Automatic*  
454 *Control* 19: 716–723.
- 455 Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. -  
456 *Journal of Biogeography* 33: 1677–1688.
- 457 Arif, S. and Massey, M. D. B. 2023. Reducing bias in experimental ecology through directed acyclic  
458 graphs. - *Ecology and Evolution* 13: e9947.
- 459 Briscoe, N. J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A., Kéry, M., Schmid, H. and Guillerá-  
460 Arroita, G. 2021. Can dynamic occupancy models improve predictions of species' range  
461 dynamics? A test using Swiss birds. - *Glob Change Biol* 27: 4269–4282.
- 462 Bromaghin, J. F., McDonald, T. L. and Amstrup, S. C. 2013. Plausible combinations: An improved  
463 method to evaluate the covariate structure of Cormack-Jolly-Seber mark-recapture models.  
464 - *Open Journal of Ecology* 3: 11–22.
- 465 Brown, J. L., Collopy, M. W. and Smallwood, J. A. 2014. Habitat fragmentation reduces occupancy of  
466 nest boxes by an open-country raptor. - *Bird Conservation International* 24: 364–378.
- 467 Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z. and Zimmermann, N. E. 2020.  
468 Model complexity affects species distribution projections under climate change. - *Journal of*  
469 *Biogeography* 47: 130–142.
- 470 Bürkner, P.-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. - *Journal of*  
471 *Statistical Software* 80: 1–28.
- 472 Burnham, K. P. and Anderson, D. R. 2002. *Model Selection and Multimodel Inference*. - Springer.
- 473 Chen, G., Kéry, M., Plattner, M., Ma, K. and Gardner, B. 2013. Imperfect detection is the rule rather  
474 than the exception in plant distribution studies. - *Journal of Ecology* 101: 183–191.
- 475 Cook, J. D., Williams, D. M., Porter, W. F. and Christensen, S. A. 2022. Improved predictions and  
476 forecasts of chronic wasting disease occurrence using multiple mechanism dynamic  
477 occupancy modeling. - *The Journal of Wildlife Management* 86: e22296.
- 478 Doherty, P. F., White, G. C. and Burnham, K. P. 2012. Comparison of model building and selection  
479 strategies. - *J Ornithol* 152: 317–323.
- 480 Duggan, J. M., Schooley, R. L. and Heske, E. J. 2011. Modeling occupancy dynamics of a rare species,  
481 Franklin's ground squirrel, with limited data: are simple connectivity metrics adequate? -  
482 *Landscape Ecol* 26: 1477–1490.
- 483 Elith, J., Leathwick, J. R. and Hastie, T. 2008. A working guide to boosted regression trees. - *Journal*  
484 *of Animal Ecology* 77: 802–813.

485 Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities  
 486 under imperfect detection: advances, challenges and opportunities. - *Ecography* 40: 281–  
 487 295.

488 Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J. and Hijmans, M. R. J. 2017. Package ‘dismo.’ - *Circles*  
 489 9: 1–68.

490 Hoerl, A. E. and Kennard, R. W. 1970. Ridge Regression: Applications to Nonorthogonal Problems. -  
 491 *Technometrics* 12: 69–82.

492 Hooten, M. B. and Hobbs, N. T. 2015. A guide to Bayesian model selection for ecologists. - *Ecological*  
 493 *Monographs* 85: 3–28.

494 Johnson, J. B. and Omland, K. S. 2004. Model selection in ecology and evolution. - *Trends in Ecology*  
 495 *& Evolution* 19: 101–108.

496 Joseph, M. B., Preston, D. L. and Johnson, P. T. J. 2016. Integrating occupancy models and structural  
 497 equation models to understand species occurrence. - *Ecology* 97: 765–775.

498 Kelleher, S., Guillera-Arroita, G., Elith, J. and Briscoe, N. J. 2025. Twenty years of dynamic occupancy  
 499 models: A review of applications and look to the future. - *Ecography* in press.

500 Kellner, K. 2015. jagsUI: a wrapper around rjags to streamline JAGS analyses. - R package version in  
 501 press.

502 Kellner, K. F. and Swihart, R. K. 2014. Accounting for Imperfect Detection in Ecology: A Quantitative  
 503 Review. - *PLOS ONE* 9: e111436.

504 Kellner, K. F., Smith, A. D., Royle, J. A., Kéry, M., Belant, J. L. and Chandler, R. B. 2023. The unmarked  
 505 R package: Twelve years of advances in occurrence and abundance modelling in ecology. -  
 506 *Methods in Ecology and Evolution* 14: 1408–1415.

507 Kéry, M. and Royle, J. A. 2016. *Applied hierarchical modelling in ecology*. - Academic Press.

508 Kéry, M., Guillera-Arroita, G. and Lahoz-Monfort, J. J. 2013. Analysing and mapping species range  
 509 dynamics using occupancy models (M Patten, Ed.). - *J. Biogeogr.* 40: 1463–1474.

510 MacKenzie, D. I. and Bailey, L. L. 2004. Assessing the fit of site-occupancy models. - *JABES* 9: 300–  
 511 318.

512 MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G. and Franklin, A. B. 2003. Estimating site  
 513 occupancy, colonization, and local extinction when a species is detected imperfectly. -  
 514 *Ecology* 84: 2200–2207.

515 MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. and Hines, J. E. 2017. *Occupancy*  
 516 *Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. - Elsevier  
 517 *Science & Technology*.

518 Morin, D. J., Yackulic, C. B., Diffendorfer, J. E., Lesmeister, D. B., Nielsen, C. K., Reid, J. and Schaubert, E.  
 519 M. 2020. Is your ad hoc model selection strategy affecting your multimodel inference? -  
 520 *Ecosphere* 11: e02997.

521 Mortelliti, A. and Boitani, L. 2007. Estimating species' absence, colonization and local extinction in  
522 patchy landscapes: an application of occupancy models with rodents. - *Journal of Zoology*  
523 273: 244–248.

524 Nichols, J. D. and Cooch, E. G. 2024. Predictive Models are Indeed Useful for Causal Inference. -  
525 *Ecology* in press.

526 Park, T. and Casella, G. 2008. The Bayesian Lasso. - *Journal of the American Statistical Association*  
527 103: 681–686.

528 Plummer, M., Stukalov, A. and Denwood, M. 2016. rjags: Bayesian graphical models using MCMC. - R  
529 package version in press.

530 Roilo, S., Spake, R., Bullock, J. M. and Cord, A. F. 2024. A cross-regional analysis of red-backed shrike  
531 responses to agri-environmental schemes in Europe. - *Environ. Res. Lett.* 19: 034004.

532 Rota, C. T., Fletcher Jr, R. J., Dorazio, R. M. and Betts, M. G. 2009. Occupancy estimation and the  
533 closure assumption. - *Journal of Applied Ecology* 46: 1173–1181.

534 Schmid, H., Zbinden, N. and Keller, V. 2004. Überwachung der Bestandsentwicklung häufiger  
535 Brutvögel in der Schweiz.

536 Scott, P. A. and Rissler, L. J. 2015. Integrating dynamic occupancy modeling and genetics to infer the  
537 status of the imperiled flattened musk turtle. - *Biological Conservation* 192: 294–303.

538 Smith, D. H. V., Jones, B., Randall, L. and Prescott, D. R. C. 2012. DIFFERENCE IN DETECTION AND  
539 OCCUPANCY BETWEEN TWO ANURANS: THE IMPORTANCE OF SPECIES-SPECIFIC  
540 MONITORING. - *Herpetological Conservation and Biology* in press.

541 Stewart, P. S., Stephens, P. A., Hill, R. A., Whittingham, M. J. and Dawson, W. 2023. Model selection in  
542 occupancy models: Inference versus prediction. - *Ecology* 104: e3942.

543 Strebel, N., Wechsler, S., Bühler, R., Häfliger, G., Keller, V., Kéry, M., Rogenmoser, C., Spiess, M., Varga,  
544 K., Volet, B., Zbinden, N. and Schmid, H. 2026. Data of the Swiss Common Breeding Bird  
545 Monitoring Program. - *Ecology* in press.

546 Studholme, K. R., Fiorino, G. E., Grabas, G. P. and Tozer, D. C. 2023. Influence of surrounding land  
547 cover on marsh-breeding birds: Implications for wetland restoration and conservation  
548 planning. - *Journal of Great Lakes Research* 49: 318–331.

549 Summers, R. W., Swallow, B., Fridman, J., Hokkanen, T., Newton, I. and Buckland, S. T. 2024.  
550 Irruptions of crossbills *Loxia* spp. in northern Europe – patterns and correlations with seed  
551 production by key and non-key conifers. - *Ibis* 166: 1172–1183.

552 Tredennick, A. T., Hooker, G., Ellner, S. P. and Adler, P. B. 2021. A practical guide to selecting models  
553 for exploration, inference, and prediction in ecology. - *Ecology* 102: e03336.

554 Urban, M. C., Nadeau, C. P. and Giery, S. T. 2023. Using mechanistic insights to predict the climate-  
555 induced expansion of a key aquatic predator. - *Ecological Monographs* 93: e1575.

556 Valavi, R., Elith, J., Lahoz-Monfort, J. J. and Guillerá-Arroita, G. 2019. blockCV: An r package for  
 557 generating spatially or environmentally separated folds for k-fold cross-validation of  
 558 species distribution models. - *Methods in Ecology and Evolution* 10: 225–232.

559 van Erp, S., Oberski, D. L. and Mulder, J. 2019. Shrinkage priors for Bayesian penalized regression. -  
 560 *Journal of Mathematical Psychology* 89: 31–50.

561 van Strien, A. J., van Swaay, C. A. M. and Kéry, M. 2011. Metapopulation dynamics in the butterfly  
 562 changed decades before occupancy declined in The Netherlands. - *Ecological Applications*  
 563 21: 2510–2520.

564 Yackulic, C. B., Nichols, J. D., Reid, J. and Der, R. 2015. To predict the niche, model colonization and  
 565 extinction. - *Ecology* 96: 16–23.

566

## Supplemental Information for:

# Comparing covariate selection strategies for predictive dynamic occupancy models

### Table of Contents:

|                                |         |
|--------------------------------|---------|
| <b>1. Swiss bird dataset</b>   | Page 2  |
| <b>2. Covariate data</b>       | Page 4  |
| <b>3. Modelling workflows</b>  | Page 7  |
| <b>4. Predictions</b>          | Page 11 |
| <b>5. Sensitivity Analyses</b> | Page 14 |
| <b>6. Additional figures</b>   | Page 15 |

## 1. Swiss Bird dataset

The data used for this study was provided by the Swiss Ornithological Institute (Vogelwarte Sempach), who conduct the Swiss common breeding bird monitoring program (MHB). The program is described in greater detail in Schmid et al. (2004) (in German, French, and Italian) and in Strebel et al. (2026). Here, we describe the essential details for the data used in the current study.

### 1.1. MHB Surveys

The MHB is conducted at 267 sites distributed across Switzerland, with the same sites surveyed in each year. Sites are defined as a 1 square kilometre grid cell. In each year, 2-3 surveys are conducted at each site, with 3 surveys at site at lower elevations and 2 surveys at higher elevation sites near the treeline at around 2000 m a.s.l. However, some surveys are missed due to inclement weather or accessibility issues, so not all sites receive their allotted surveys in each year. Surveys are conducted by trained volunteers who traverse the entire site on a pre-defined transect route of typical length 4 – 6 kms with coverage of major biomes within the grid-cell. During each survey, counts are made of all birds breeding within Switzerland.

For the purposes of this study, these counts were converted into binary presence/absence records. We used all records between 2000 and 2018, with 2000-2008 designated as training data and 2009-2018 reserved for testing model performance. One of the 267 sites in the

dataset was not surveyed between 2000 and 2008 and was thus removed, leaving 266 sites for analysis.

## 1.2. Cut-off dates and remaining data

To avoid violations of the closure assumption caused by surveying prior to a migratory species' arrival in Switzerland or by detecting transient individuals passing through sites on their migration, we removed any surveys that occurred prior to a species-specific date representing the end of the migratory period. The table below includes these dates and the total number of surveys included between 2000 and 2018 after applying the cut-off.

| Species                   | Cut-off  | Total surveys |
|---------------------------|----------|---------------|
| Common Kestrel            | April 13 | 14125         |
| Eurasian Green Woodpecker | April 13 | 14125         |
| Eurasian Skylark          | April 13 | 14125         |
| Garden Warbler            | May 8    | 10313         |
| Mallard                   | April 13 | 14125         |
| Red Crossbill             | April 13 | 14125         |
| Red Kite                  | April 13 | 14125         |
| Red-backed Shrike         | May 13   | 9377          |
| Whinchat                  | May 18   | 8362          |
| White-throated dipper     | April 13 | 14125         |

*Table S1-1: Cut-off dates for each of ten focal species. Surveys prior to the cut-off dates were censored. Total surveys is the number of valid surveys remaining following censoring.*

## 1.3. Available data

We have made the data used for the modelling in this survey as part of the `sbmodels` package uploaded to figshare here: <https://figshare.com/s/2a54afd1445027b6450d>. The spatial locations of each site have been anonymised, and records of unmodelled species have been removed.

## 2. Covariate data

We considered 10 candidate covariates (11 for the Red Crossbill) for the occupancy related-parameters  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ , and 2 covariates for the detection parameter  $p$ . All covariates for all parameters were scaled by subtracting the mean and dividing the result by the standard deviation of the training data — that is, only values at the 266 sites surveyed between 2000 and 2008.

### 2.1. CORINE land cover data

Seven of our covariates were derived from the Copernicus Land Monitoring Service's CORINE (Condition of Information on the Environment) Land Cover dataset (Copernicus Land Monitoring Service 2018). CORINE divides Europe into 44 land cover categories at 100 metre resolution, with updates published every six years — we used the 2000, 2006, 2012, and 2018 editions. The 44 classes were consolidated into 7 broader groups that capture landscape categories relevant to birds:

| Grouping           | Subclasses                                                                                                                                                                                                                                                       |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Artificial         | Continuous urban fabric, discontinuous urban fabric, industrial or commercial units, road and rail networks and associated land, port areas, airports, mineral extraction sites, dump sites, construction sites, green urban areas, sport and leisure facilities |
| Arable             | Non-irrigated arable land, permanently irrigated land, rice fields                                                                                                                                                                                               |
| Heterogenous       | Vineyards, fruit trees and berry plantations, olive groves, pastures, annual crops associated with permanent crops, complex cultivation patterns, land principally occupied by agriculture with significant areas of natural vegetation, agroforestry areas      |
| Forest             | Broad-leaved forest, coniferous forest, mixed forest                                                                                                                                                                                                             |
| Grassland          | Natural grassland, moors and heathland, sclerophyllous vegetation, transitional woodland/shrub                                                                                                                                                                   |
| Bare (not used)    | Beaches dunes sands, bare rock, sparsely vegetated areas, burnt areas, glaciers and perpetual snow                                                                                                                                                               |
| Water and wetlands | Inland marshes, peatbogs, salt marshes, salines, intertidal flats, water courses, water bodies, coastal lagoons, estuaries, sea and ocean                                                                                                                        |

Table S2-1: Groupings of CORINE land-cover subclasses consolidated into each of seven categories for modelling.

## 2.2. Forest cover

We created two additional land cover covariates using the Swiss Department of the Environment's forest mixture rate dataset, which gives the fraction of coniferous and deciduous trees in each forested cell at 10 metre resolution (Swiss Federal Office of the Environment 2025). We aggregated this data to produce the covariates CONI, the proportion of each cell covered in conifer trees, and DECI, the proportion covered in deciduous trees. In contrast to FORE, both CONI and DECI were static covariates. We found that FORE was highly collinear (Pearson's correlation greater than 0.7) with both covariates and chose to remove CONI and DECI from the candidate covariate pool for 9 of our 10 focal species. The exception was the Red Crossbill (*Loxia curvirostra*), which is well-known for its strong association with conifer trees (Génard and Lescourret 1987). For this species, we instead removed FORE from the covariate pool and retained CONI and DECI.

## 2.3. Climate covariates

We produced two climate covariates using data from CHELSA (Climatologies at high resolution for the earth's land surface areas, Karger et al. 2017, 2021). Using their 1 kilometre resolution monthly mean air surface temperature data, we calculated the mean temperature for each cell during the warmest 3-month period of each year. This is meant to capture summer conditions and is roughly equivalent to the 'bio10' variable from the popular BioClim set. We averaged this value between the focal year and the prior year to create our TEMP covariate, doing so to account for a possible lagged effect. We also used CHELSA's monthly precipitation records at 1 kilometre resolution to calculate total annual precipitation, equivalent to the 'bio12' covariate. This was also averaged between the focal and previous year to create the RAIN covariate.

Our final environmental covariate was derived from the Copernicus Land Monitoring Service's global normalised derived vegetation index (NDVI) dataset, which includes 10-daily rasters at 1 kilometre resolution back to 1999 (Copernicus Land Monitoring Service 2021). This dataset is pre-processed, with NA values for cells covered in clouds or snow. We averaged the non-missing values collected between April and August of each year (corresponding with the bird survey window). Based on manual inspection, the only cells with no non-missing values in a year were high-alpine cells fully covered by glaciers or snowpack. The NDVI value of these cells was set to 0. As with TEMP and RAIN, for our NDVI covariate we averaged the values for the focal and prior year to account for lagged effects.

## 2.4. Detection covariates

Our primary objective in this case study was to compare the performance of different covariate selection approaches, focusing on the occupancy-related parameters. Therefore, we elected to consider only a simplified set of candidate covariates for  $p$  based on prior investigations using this dataset. These covariates included DATE, the calendar date of each survey, and ELEV, the average elevation within each cell, because the influence of the survey date was expected to vary with elevation. ELEV was derived from the Swiss 25 metre Digital Height Map (Federal Office of Topography swisstopo 2004). We considered five alternative structures for  $p$ :

intercept only, ~ DATE, ~DATE + DATE2, ~ DATE + ELEV + DATE\*ELEV, and ~DATE + DATE2 + ELEV + DATE\*ELEV. ELEV was only included in a model if it was selected as part of an interaction with DATE.

## 2.5. Covariate correlations

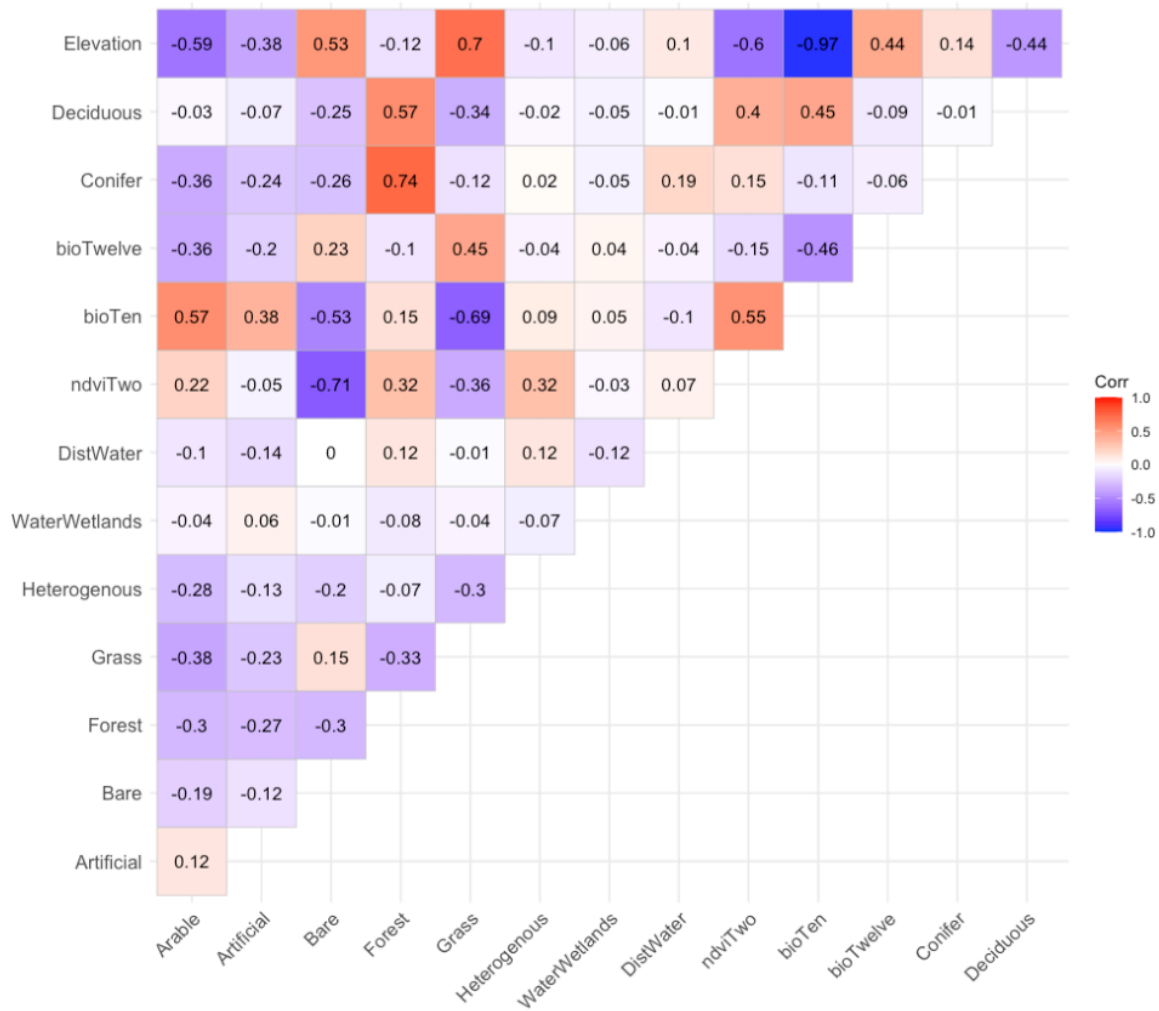


Table S2-2: Pearson's correlation coefficients between covariates considered for modelling.

### 3. Modelling workflows

Our covariate selection workflows consisted of two stages: an initial ‘filtering’ stage where the original pool of candidate covariates was reduced in size, and a second ‘selection’ stage where a final model was identified and fit. This two-stage approach reflected practices seen in the DOM literature (Kelleher et al. 2025).

#### 3.1. Filtering stage methods

We applied one of five methods to reduce the size of the original covariate pools for  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ . Aside from a ‘no filtering’ option, all filtering methods were permitted to retain a maximum of six covariates per process to keep comparisons as fair as possible. Because the covariate pool for  $p$  was much smaller than the pool for the other processes, we did not remove any covariates for  $p$  during the filtering stage.

- **Unfiltered:** The full covariate pool was retained for the ‘selection’ stage.
- **A priori:** The author most familiar with the study system (*redacted for review*) was asked to choose the covariates that they believed would produce the best model, allowing up to six covariates for each of  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ .
- **Boosted regression trees (BRTs):** We fit three sets of BRTs to identify covariates useful for explaining variation in  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ . For each process we converted observed yearly detections at each site into a binary response variable. For  $\psi_1$ , this meant converting the observed occupancy states for each site in 2000 to 1s (detected in at least one survey) to 0s (undetected in all surveys). For  $\gamma$  we used observed colonisation events; for all sites & occasions where a species was not detected during the prior year, sites which had detections became 1s (observed colonisations) and sites which did not became 0s. The inverse was used for  $\epsilon$  to model observed local extinction events. For each parameter, all candidate covariates were used as predictor variables. We fit binary BRTs using the *dismo* R with a bag fraction of 0.75 and a tree complexity of 1. Learning rate was initially set to 0.003, and was halved iteratively until the output fit greater than 1000 trees following guidance from Elith et al. (2008). For each process, the six covariates with the highest variable importance were retained. If fewer than six covariates for a process had variable importance values greater than 0, only those covariates were retained.
- **Single variable:** For each covariate, we fit models with only a single covariate on one of  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ , and the global formula for  $p$ . For example, to test the *FARM* covariate for  $\psi_1$  we fit the model  $\psi_1(\sim FARM + FARM^2)$   $\gamma(\sim 1)$   $\epsilon(\sim 1)$   $p(\sim DATE + DATE^2 + ELEV + DATE * ELEV)$ . We then compared the AIC for each parameter/covariate combination,

and retained the six covariates that produced the lowest AIC for each parameter. These single-variable models were fit in the *unmarked* R package using the *colext* function (Kellner et al. 2023).

- **Static occupancy:** Using detection histories from the first primary occasion in 2000, we fit single-season occupancy models with all combinations of up to six covariates on the occupancy parameter. For the detection parameter, we used the global formula for  $p$ . We retained all covariates used for the occupancy parameter in the model with the lowest AIC, keeping the same covariates for  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ . Static occupancy models were fit in *unmarked* using the *occu* function.

### 3.2. Selection stage methods

In the second stage of our workflow, we used one of eight methods covering MLE and Bayesian frameworks to select and fit a single final DOM using the covariates retained from the filtering stage. When including any of the covariates in models for occupancy-related parameters, we always used both the linear and quadratic forms to account for any non-linearity in their relationships with parameters. For example, we only considered  $\sim \text{FARM} + \text{FARM2}$  rather than  $\sim \text{FARM}$  or  $\sim \text{FARM2}$  alone. During this stage, we allowed methods to select from the five possible formulae for  $p$  in addition to selecting for  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ . All MLE methods were fit using the *colext* function from a development version of the *unmarked* R package. To avoid local minima, models were fit with at least six sets of starting values and up to 20 sets if required for convergence. Unless otherwise stated, default settings in *unmarked*'s *colext* function were used.

All Bayesian models were fit in JAGS using the *autojags* function from *jagsui* (Kellner 2015), procedurally adding iterations in steps of 10,000 until all r-hat values for coefficients fell below 1.1 or an upper limit of 500,000 iterations was reached. Models were run on 4 chains after a burn-in of 10,000 iterations with a thinning value of 5 (Plummer et al. 2016). R-hat values were checked for the parameter intercepts and regression coefficients of each model to ensure that all were below a limit of 1.1. We assumed the minimal burn-in of 10,000 was sufficient due to the use of the auto-updating *jagsUI::autojags* function – under default settings, only the first round of 10000 iterations where all parameters achieve an R-hat of  $<1.1$  are included in the posterior. A thinning constant of 5 was used only to reduce file size given the large number of models; as the final batch of iterations for each model was 10000 on each of four chains, the posterior was represented by 8000 total iterations.

- **Global:** A model was fit in *unmarked* with the full set of covariates retained after filtering for  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ , and the global formula for  $p$ .
- **Forward sequential:** We first fit a fully null intercept-only model. A stepwise process was then used for  $p$ , where models were fit adding one covariate at a time and retaining the covariate that most reduced the AIC. This was repeated until further additions no

longer improved AIC, at which point the structure for  $p$  was fixed. The same process was then repeated to fix  $\psi_1$ ,  $\gamma$ , and  $\epsilon$  in that order.

- **Backward sequential:** We first fit a global model with all retained covariates for  $\psi_1$ ,  $\gamma$ , and  $\epsilon$ , using the global formula for  $p$ . A stepwise process was then used for  $p$ , testing removal of one covariate at a time until further removals no longer reduced the AIC and the structure for  $p$  was fixed. The same process was then repeated to fix  $\psi_1$ ,  $\gamma$ , and  $\epsilon$  in that order.
- **Submodels:** We fit all possible combinations of up to 6 covariates for each of  $\psi_1$ ,  $\gamma$ , and  $\epsilon$  while using null structures on the other parameters. For  $p$ , all five candidate formulae were tested in the same manner. The formulae that fit the lowest-AIC model for each parameter were then combined to fit a final model.
- **Semi-exhaustive:** We fit all possible DOMs that included up to four covariates on each of  $\psi$ ,  $\gamma$ , and  $\epsilon$  and any of the five candidate formulae for  $p$ . We then chose the final model based on AIC.
- **Vague:** To implement vague priors, we provided intercepts for four process using uniform priors from 0.001 to 0.999 on the probability scale. Regression coefficients are on the logit scale and were provided uniform priors from -10 to 10, which covers the range of reasonable coefficients.
- **Lasso:** Following Biswas and Lin (2012)'s example of the lasso in a logistic regression context, we chose a Gamma hyperprior for  $\lambda$  with a single parameter  $\alpha$  for both shape and rate.

$$\begin{aligned}\beta_i &\sim \text{DoubleExponential}(0, \lambda) \\ \lambda &\sim \text{Gamma}(\alpha, \alpha)\end{aligned}$$

For the Gamma hyperprior we set  $\alpha = 3.59$  to produce a prior for the regression coefficients with a standard deviation of 2.5 on the logit scale. While all regression coefficients on any given parameter had the same double exponential prior we used a different  $\lambda$  for each of  $\psi_1$ ,  $\gamma$ ,  $\epsilon$ , and  $p$  to allow covariates on each parameter to be penalised independently. Parameter intercepts were not regularised and had uniform priors from 0.001 to 0.999 on the probability scale.

- **Ridge:** As for Lasso, a Bayesian equivalent has been proposed for ridge regularisation in linear (normal) models, using in this case a Gaussian prior for the regression coefficients rather than a double-exponential (van Erp et al. 2019). As the variance of the Gaussian distribution is  $\tau^{-1}$  compared to  $2\tau^{-2}$  for the double-exponential, we set the precision  $\tau$  to  $\lambda^2/2$  to achieve the same standard deviation at a given value of  $\lambda$ .

$$\beta_i \sim \text{Gaussian}(0, \tau)$$

$$\tau = \lambda^2 / 2$$

$$\lambda \sim \text{Gamma}(\alpha, \alpha)$$

We used the same value of  $\alpha = 3.59$  for the hyperprior on  $\lambda$  to achieve a standard deviation of 2.5 on the prior for the regression coefficients. As with the lasso, we used different  $\lambda$ 's for each parameter and gave each intercept a uniform prior from 0.001 to 0.999.

We attempted all combinations of the five filtering methods and eight selection methods, aside from the prohibitively computationally expensive combination of 'no filtering' and 'semi-exhaustive.' In total, 39 covariate selection workflows were attempted per species (390 total workflows). With the exception of the 'semi-exhaustive' selection method, all workflows were run in parallel on identical virtual machines (running on 12 VCPU cores with 48gb of RAM on Ubuntu 22.04). Semi-exhaustive models were run on a high-performance computing platform due to the computational demands, with a sample re-run on the virtual machines to allow comparisons of run-times.

## 4. Predictions

After a final model was identified for each workflow we predicted occupancy probabilities for all 1x1 kilometre grid cells in Switzerland for each year from 2000-2018. Predictions were initiated with estimated  $\psi_1$  in 2000 and the estimated probability of occupancy in subsequent years was derived from the  $\gamma$  and  $\epsilon$  estimates. Point estimates were used to make predictions for the MLE models, and the median value from the posterior distribution used for the Bayesian models.

For the 266 survey sites, we also calculated the probability of observing the species at each site during each primary occasion from 2000-2018, given the number of surveys that were conducted at that site, in that primary occasion:

$$pr(Ob_{i,t}) = \psi_{i,t} \prod_{j=1}^{J_{i,t}} (1 - p_{i,t,j})$$

This probability was used as the basis for assessing the performance of the predictions because the actual variable of interest – site occupancy – is a latent state that is imperfectly observed.

## 5. Sensitivity analyses

### 5.1. Forward selection

The forward selection approach we used fixed the four DOM parameters one-by-one. The order of fixing these parameters we used  $(p, \psi_1, \gamma, \epsilon)$  was chosen *a priori* based on one of the more approaches tested in Morin et al. 2020. To confirm that this choice did not influence our results, we also tested the other possible options *post hoc*. Our method ranked 8<sup>th</sup> of 24 alternatives, comparably to the parameter orders above it.

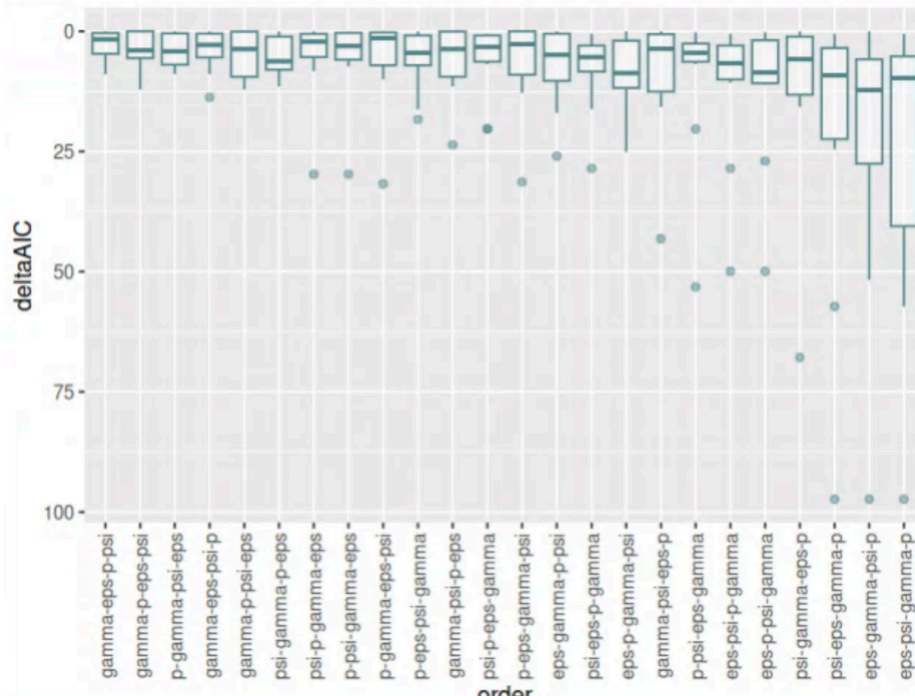


Figure S5-1: Boxplots show the distribution of deltaAICs for models fit for 10 species using forward selection with a given parameter order. deltaAICs are between each final model and the best model fit for each species.

### 5.2. Bayesian regularisation

The priors on the hyperpriors used for the Lasso and Ridge regularisation methods were specified in Supplement 3.2 as a  $\text{gamma}(3.59, 3.59)$  distribution. These permitted a reasonably large range of penalization strengths, but to test the sensitivity of these methods to other regularisation strengths we also tested two alternative priors for three species:  $\text{gamma}(.359, .359)$  and  $\text{gamma}(35.9, 35.9)$ . We found that the differences between these prior sets in terms of predictive performance of the resulting models were minimal ( $<0.01$  points of AUC).

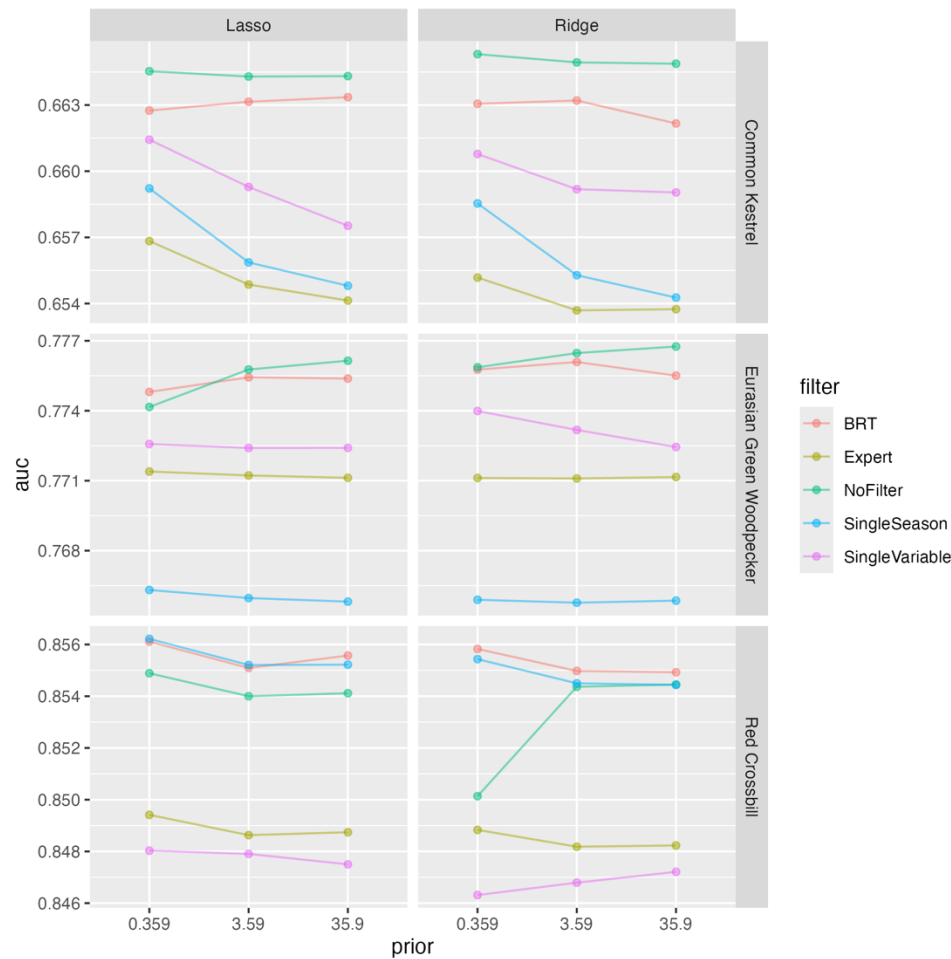


Figure S5-2: Comparison of AUC scores predicting to testing data for models fit with each of three priors for regularisation penalty hyperpriors.

### 5.3. Frequentist AIC comparison

We also compared the lowest AIC achieved for each species across each of the workflows attempted. These results indicate that the differences between the top models in each workflow were varied, generally beyond the 2 deltaAIC threshold often used to consider models ‘equivalent.’

| Filter          | Select       | COKE | EUGW | EUSK | GAWA | MALL       | RECR  | REKI | RBSH | WHIN | WTDI |
|-----------------|--------------|------|------|------|------|------------|-------|------|------|------|------|
| BRT             | Backward     | 21.1 | 9.2  | 24.5 | 1.7  | 12.9       | 11.7  | 3.6  | 16.2 | 22.9 | 5.6  |
| BRT             | S.Exhaustive | 17.7 | 10.0 | 30.3 | 4.0  | 25.1       | 20.9  | 2.3  | 28.9 | 26.4 | 2.1  |
| BRT             | Forward      | 10.3 | 12.0 | 24.6 | 0.0  | 3.9        | 11.7  | 12.6 | 14.5 | 20.4 | 2.1  |
| BRT             | Global       | 27.8 | 18.9 | 39.3 | 16.4 | 22.8       | 23.1  | 12.4 | 27.9 | 34.7 | 36.1 |
| BRT             | Submodels    | 11.7 | 12.3 | 25.2 | 1.4  | <b>0.0</b> | 17.8  | 5.4  | 18.2 | 21.0 | 2.1  |
| <i>A priori</i> | Backward     | 44.1 | 17.8 | 59.7 | 46.1 | NA         | 105.0 | NA   | 19.2 | 32.8 | 26.7 |
| <i>A priori</i> | S.Exhaustive | 40.3 | 14.5 | 59.6 | 47.5 | 63.5       | 105.0 | 3.7  | 23.2 | 41.8 | 20.4 |

|                 |              |            |            |            |            |       |            |            |            |            |            |
|-----------------|--------------|------------|------------|------------|------------|-------|------------|------------|------------|------------|------------|
| <i>A priori</i> | Forward      | 41.6       | 19.9       | 62.8       | 46.1       | 63.3  | 105.0      | 14.3       | 32.1       | 38.6       | 20.4       |
| <i>A priori</i> | Global       | 61.2       | 32.9       | 74.1       | 59.2       | NA    | 107.0      | NA         | 34.0       | 44.6       | 50.4       |
| <i>A priori</i> | Submodels    | 42.7       | 18.4       | 60.1       | 46.6       | 65.0  | 105.0      | NA         | 24.8       | 33.9       | 20.4       |
| Unfiltered      | Backward     | NA         | NA         | NA         | NA         | NA    | NA         | NA         | NA         | NA         | NA         |
| Unfiltered      | Forward      | 7.3        | 3.9        | 25.2       | <b>0.0</b> | 2.1   | <b>0.0</b> | <b>0.0</b> | <b>0.0</b> | <b>0.0</b> | <b>0.0</b> |
| Unfiltered      | Global       | NA         | NA         | NA         | NA         | NA    | NA         | NA         | NA         | NA         | NA         |
| Unfiltered      | Submodels    | <b>0.0</b> | <b>0.0</b> | <b>0.0</b> | 1.3        | NA    | 14.3       | NA         | NA         | NA         | 0.3        |
| NullOccFullDet  | Global       | 157.8      | 188.5      | 187.0      | 118.9      | 114.3 | 348.7      | 169.1      | 117.6      | 155.6      | 21.8       |
| NullOccNullDet  | Global       | 170.1      | 249.6      | 184.8      | 290.3      | 257.5 | 533.7      | 325.3      | 132.8      | 159.4      | 32.7       |
| StaticOcc       | Backward     | NA         | 29.4       | NA         | NA         | NA    | 18.9       | 1.6        | NA         | 26.4       | 11.2       |
| StaticOcc       | S.Exhaustive | 47.7       | 29.4       | 33.0       | 8.3        | 5.3   | 21.8       | 1.6        | 26.6       | 39.3       | 4.7        |
| StaticOcc       | Forward      | 65.5       | 32.0       | 36.6       | 3.4        | 19.4  | 19.0       | 12.9       | 18.1       | 26.4       | 4.7        |
| StaticOcc       | Global       | NA         | 32.0       | NA         | NA         | NA    | 26.5       | 7.0        | NA         | 40.6       | 38.1       |
| StaticOcc       | Submodels    | NA         | 32.0       | 33.8       | 6.8        | NA    | 27.3       | 7.0        | 19.8       | 27.4       | 4.8        |
| SingleVariable  | Backward     | 30.7       | 6.1        | 27.7       | 4.0        | 15.3  | 28.6       | NA         | NA         | 20.7       | NA         |
| SingleVariable  | S.Exhaustive | 21.4       | 5.0        | 45.6       | 1.3        | 2.8   | 24.9       | 6.7        | 19.0       | 31.1       | 8.2        |
| SingleVariable  | Forward      | 23.2       | 6.9        | 28.7       | 0.0        | 9.0   | 24.9       | 8.0        | 9.8        | 19.2       | 10.8       |
| SingleVariable  | Global       | 34.0       | 20.0       | 40.2       | 21.0       | 33.5  | 41.5       | NA         | NA         | 33.9       | NA         |
| SingleVariable  | Submodels    | 23.2       | 8.0        | 28.7       | 3.4        | 4.2   | 27.5       | 5.8        | 14.6       | NA         | 8.2        |

Table S5-1: DeltaAIC between the best overall model for a species and the best model achieved under each workflow.

## 6. Supplemental figures

### 6.1. Computation and runtime

| Filter method          | Selection method |           |          |            |                 |              |              |              |
|------------------------|------------------|-----------|----------|------------|-----------------|--------------|--------------|--------------|
|                        | Global           | Forward   | Backward | Sub-models | Semi-exhaustive | <i>Vague</i> | <i>Lasso</i> | <i>Ridge</i> |
| <b><i>A priori</i></b> | 1<br>80          | 3<br>100  | 4<br>80  | 3<br>90    | 9k<br>100       | 116<br>60    | 151<br>100   | 189<br>90    |
| <b>BRT</b>             | 1<br>100         | 4<br>100  | 5<br>100 | 6<br>100   | 68k<br>100      | 197<br>50    | 148<br>100   | 186<br>90    |
| <b>Static Occ.</b>     | 22<br>50         | 22<br>100 | 26<br>50 | 23<br>80   | 15k<br>100      | 272<br>30    | 229<br>100   | 206<br>70    |
| <b>Single Var.</b>     | 2<br>70          | 6<br>100  | 8<br>70  | 8<br>90    | 81k<br>100      | 536<br>40    | 236<br>90    | 182<br>90    |
| <b>Unfiltered</b>      | —<br>0           | 10<br>100 | —<br>0   | 86<br>60   | Not<br>run<br>— | 2k<br>10     | 581<br>90    | 1k<br>80     |

Table S6-1: Computational results of 39 covariate selection workflows for the ten focal species. Each cell gives the median runtime for a workflow in minutes (top row in cell), and the percentage of species for which the workflow could be completed (bottom row in cell). Computational runtime (in minutes on a virtual machine) includes total processing time of the filtering and selection stages.

## 6.2. Performance metrics

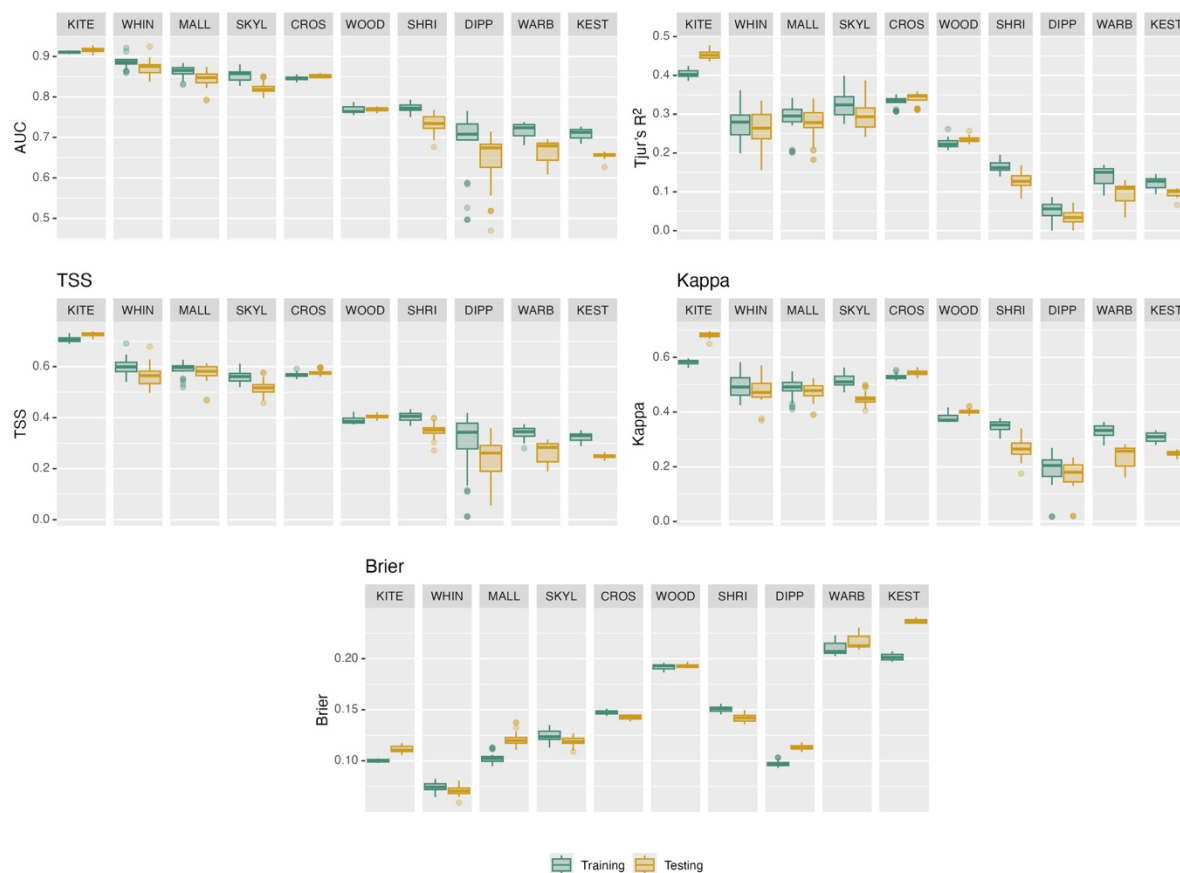


Figure S6-1: Distribution of performance metrics for models fit for each species, calculated on the training and testing period.

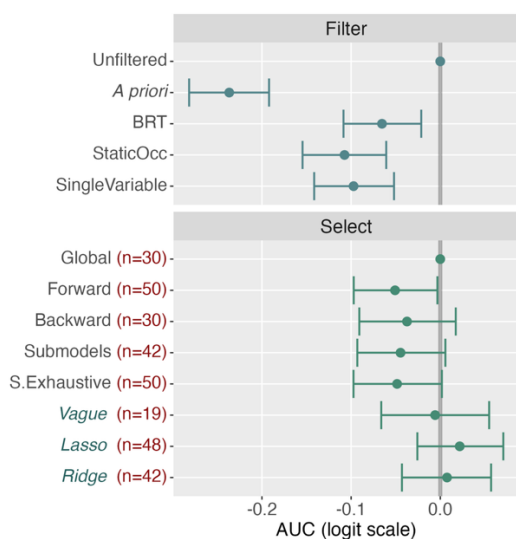


Figure S6-2: Coefficients relating each filtering & selection method to AUC (on the logit scale), based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

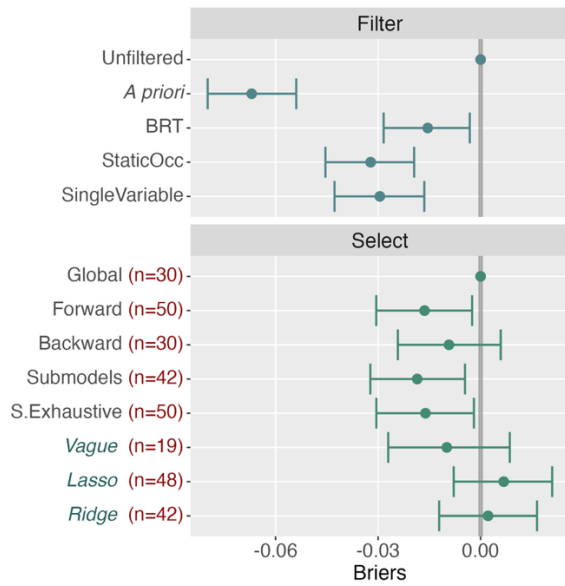


Figure S6-3: Coefficients relating each filtering & selection method to Brier's score, based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

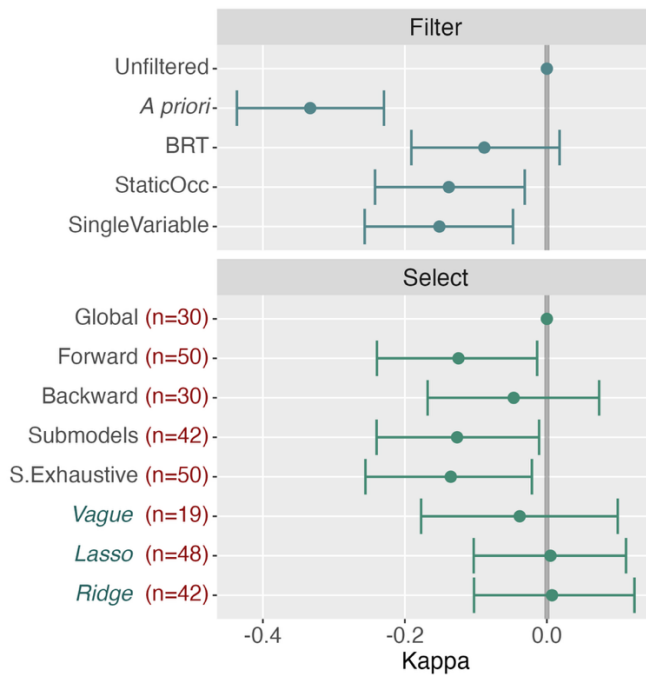


Figure S6-4: Coefficients relating each filtering & selection method to Kappa, based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

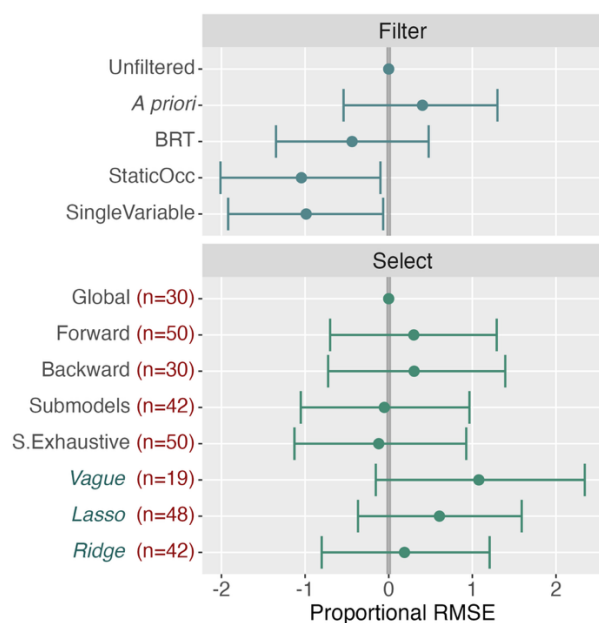


Figure S6-5: Coefficients relating each filtering & selection method to proportional RMSE, based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

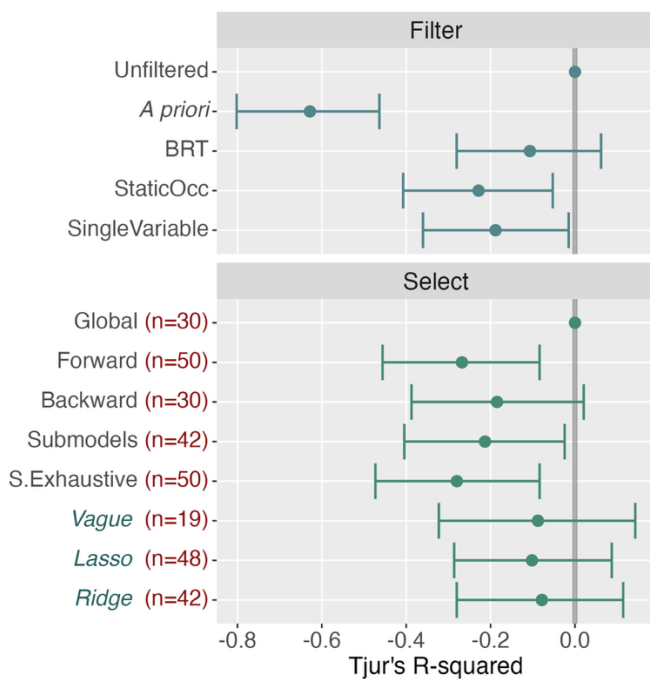


Figure S6-6: Coefficients relating each filtering & selection method to proportional Tjur's  $R^2$ , based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

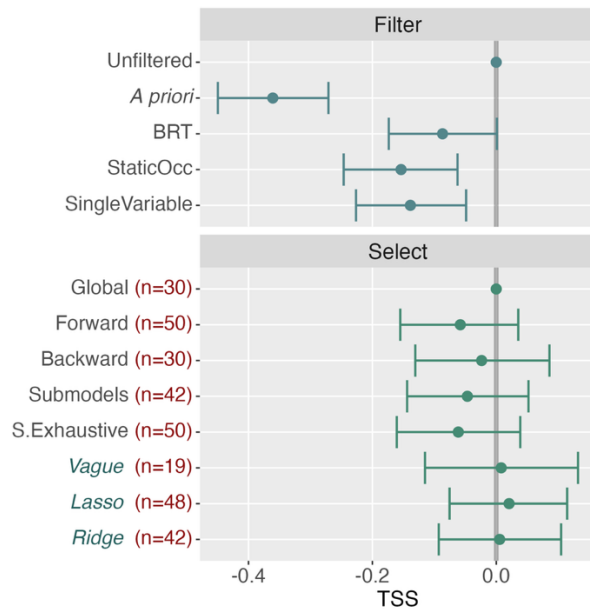


Figure S6-7: Coefficients relating each filtering & selection method to proportional True Skill Statistic, based on models including a species-specific random intercept, filtering method, and selection method. The reference class is the 'Unfiltered' and 'Global' filtering and selection methods respectively.

### 6.3. Predicted trends

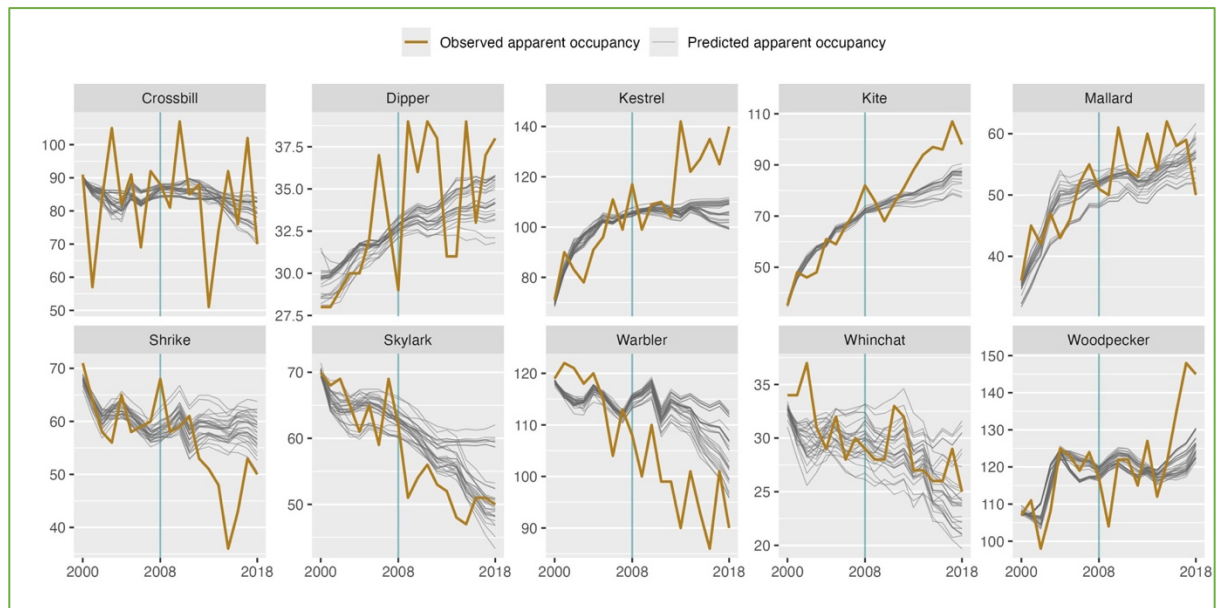


Figure S6-8: Compares the number of the 266 1km<sup>2</sup> sites where a species was detected in each year (gold lines) against the number of observed sites predicted by the final model from each workflow (gray lines). Models were fit to data from 2000 through 2008, indicated by the blue vertical line. Note that the y-axes are species-specific due to variation in prevalence.

#### 6.4. Normalised RMSE performance

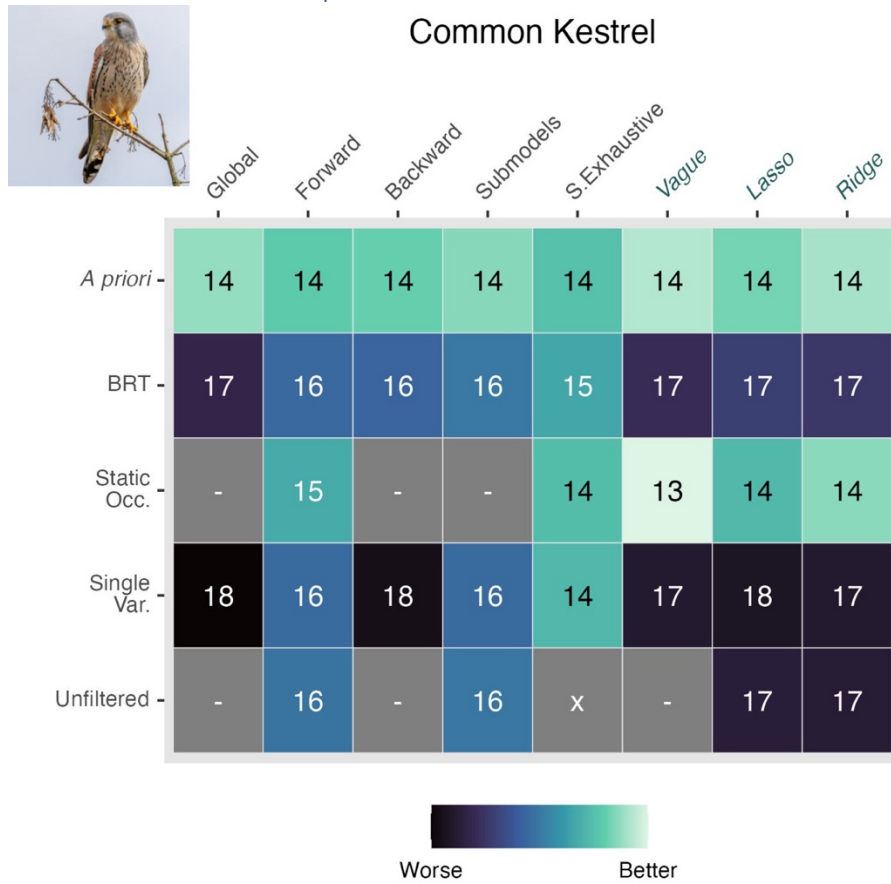
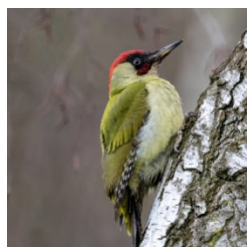


Figure S6-9: Common kestrel normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Eurasian Green Woodpecker

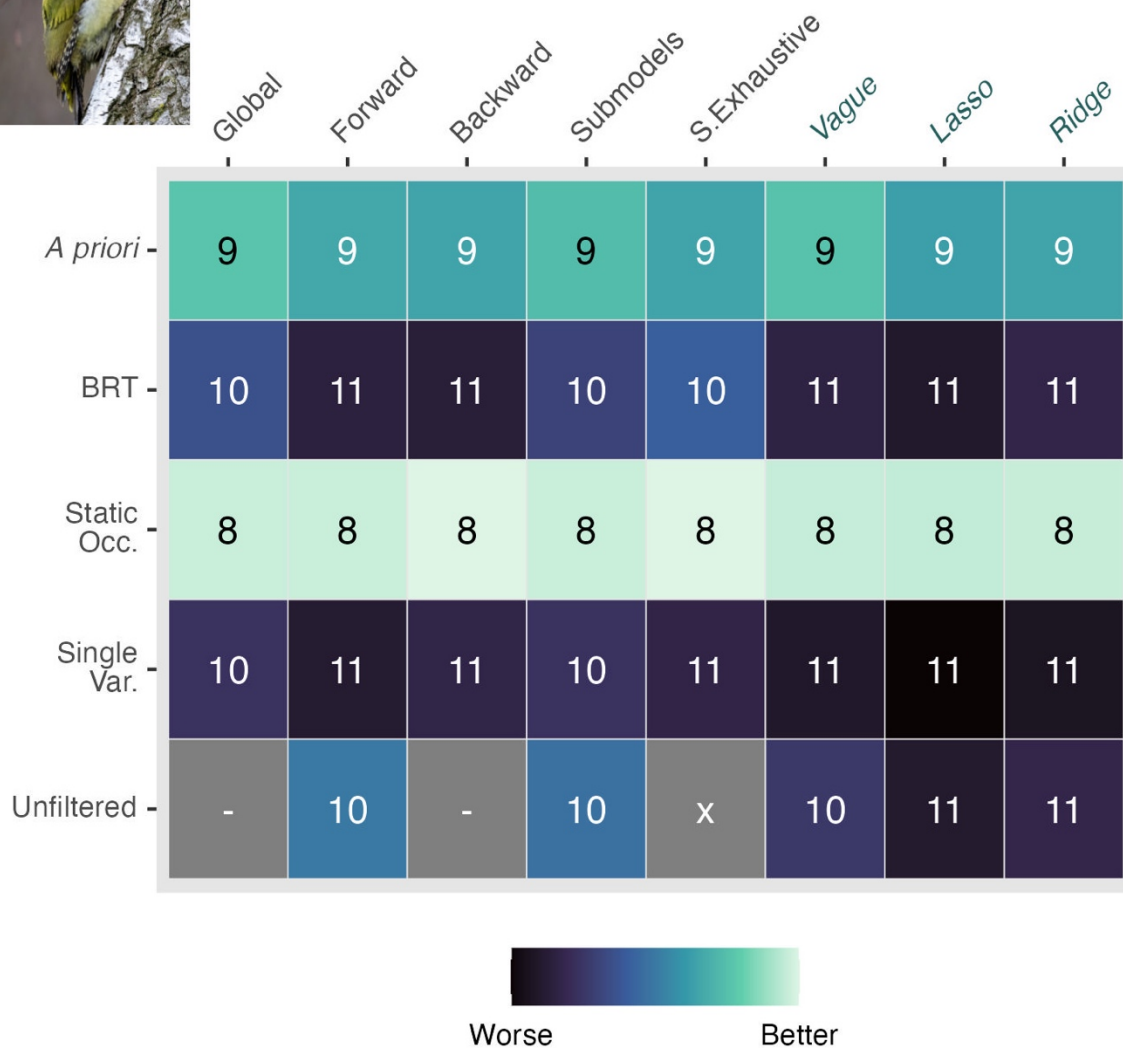


Figure S6-10: Eurasian Green Woodpecker normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Eurasian Skylark

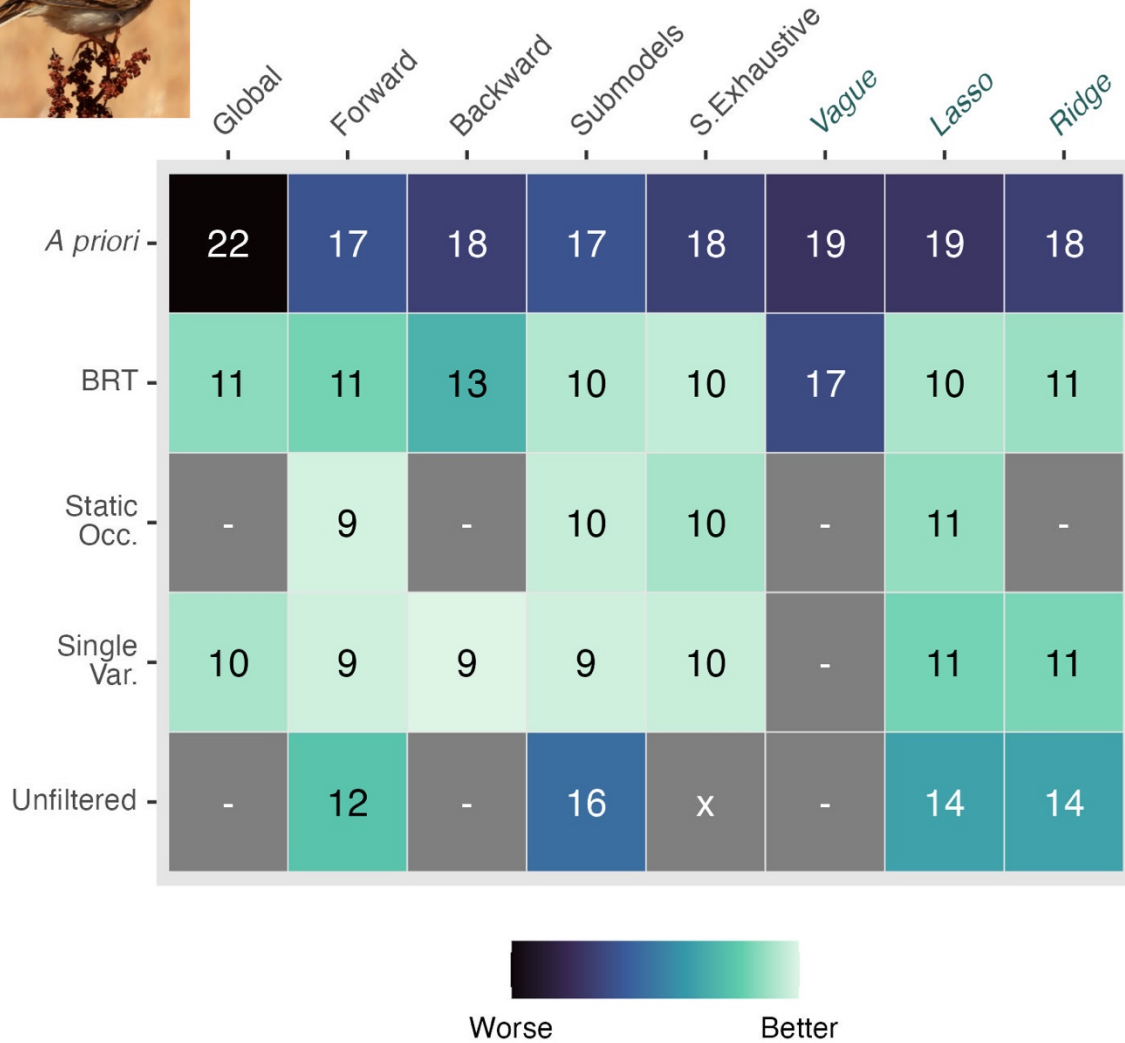


Figure S6-11: Eurasian Skylark normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Garden Warbler

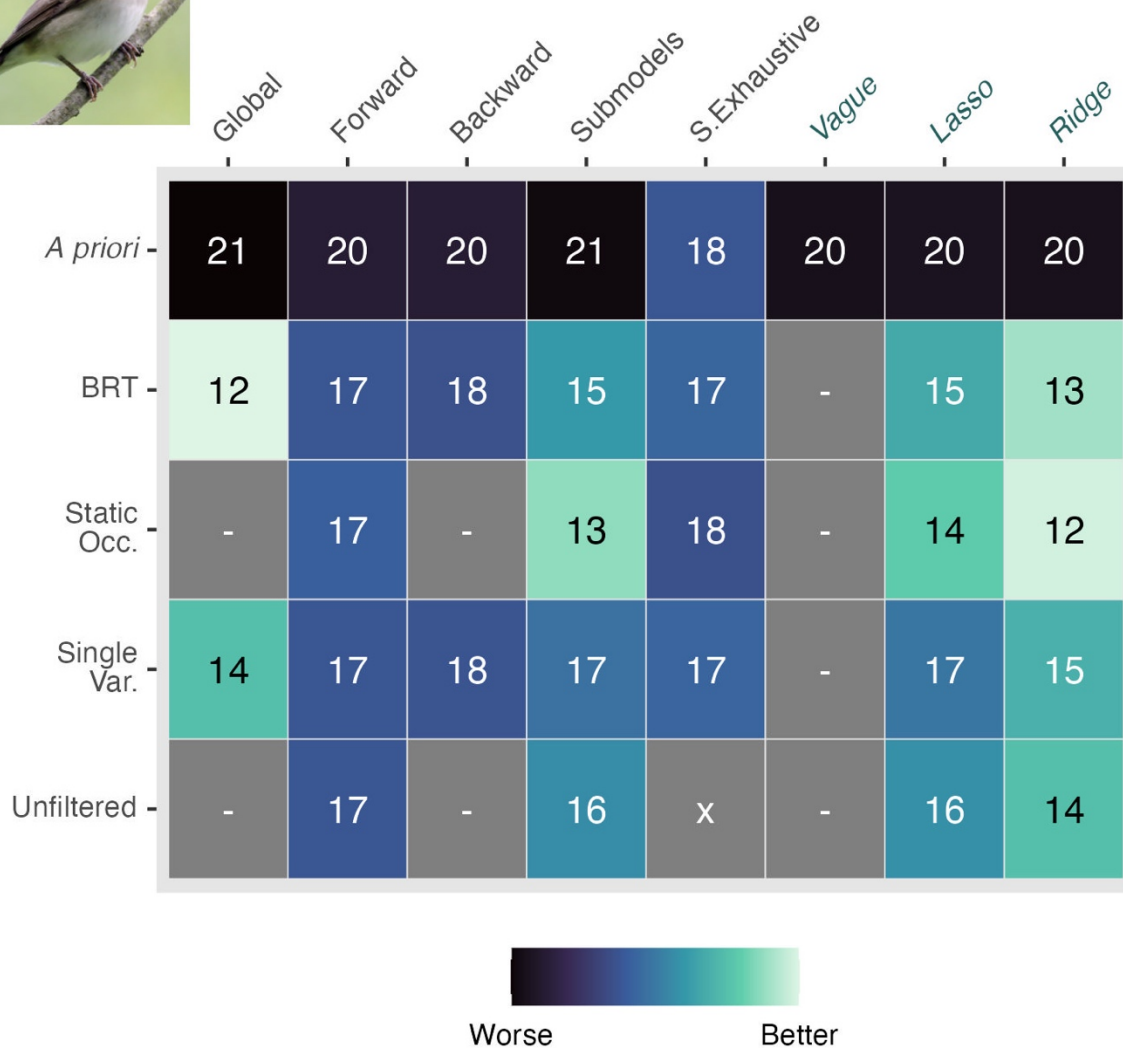
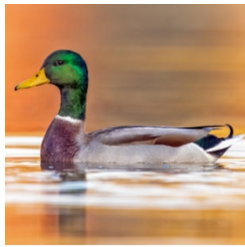


Figure S6-12: Garden Warbler normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Mallard

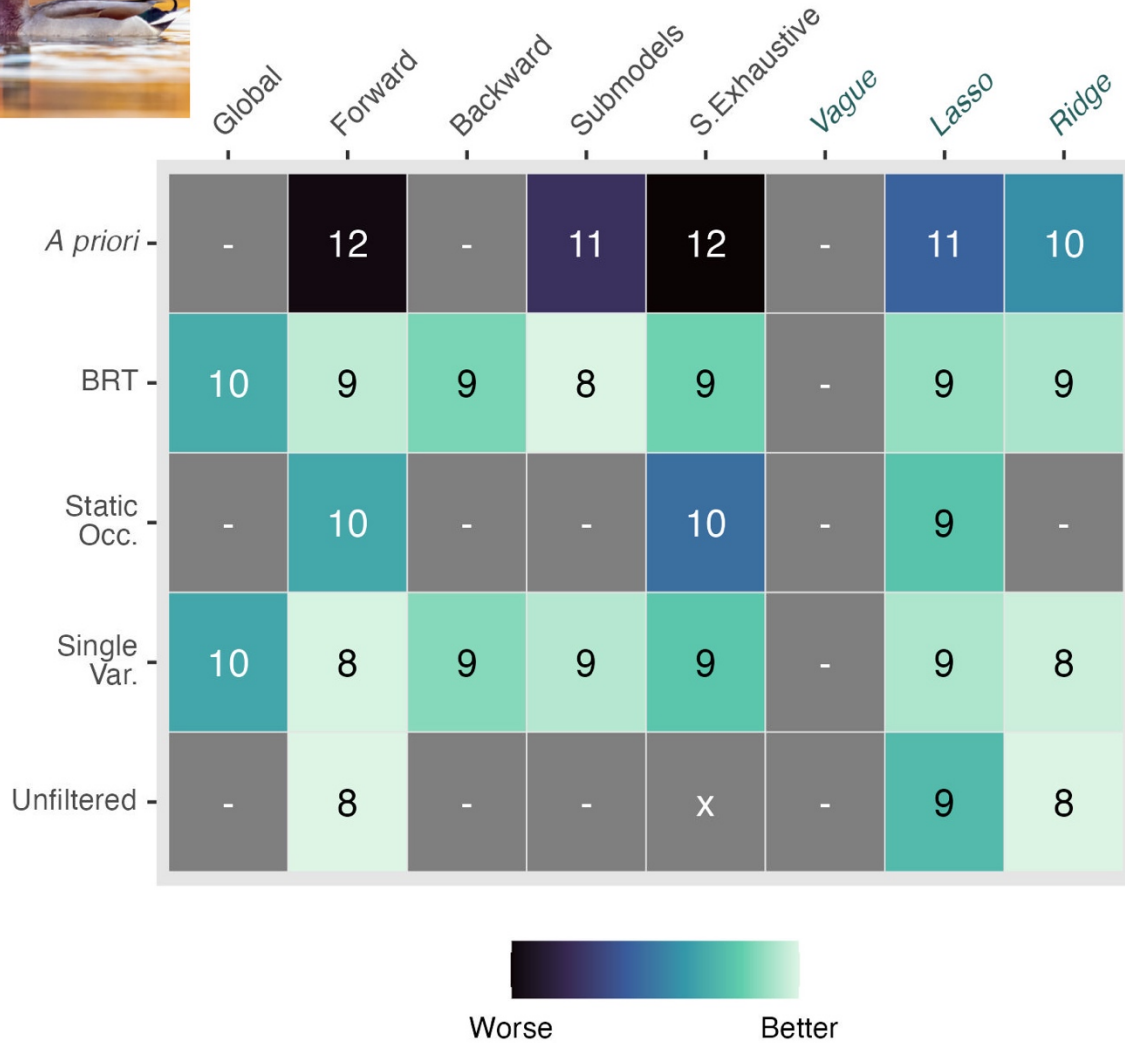


Figure S6-13: Mallard normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Red Crossbill

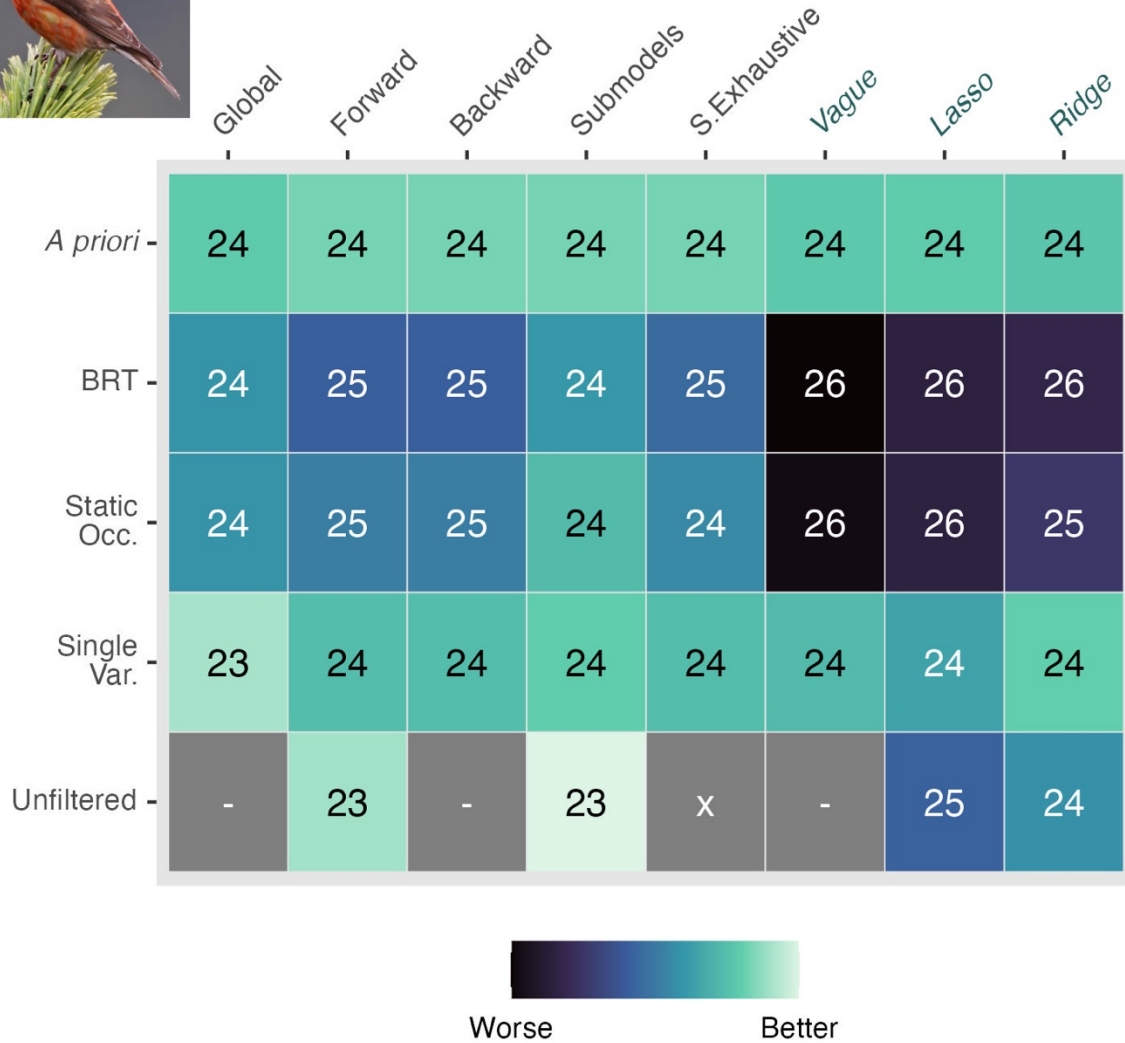
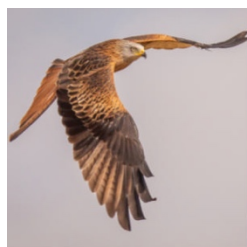


Figure S6-14: Red Crossbill normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Red Kite

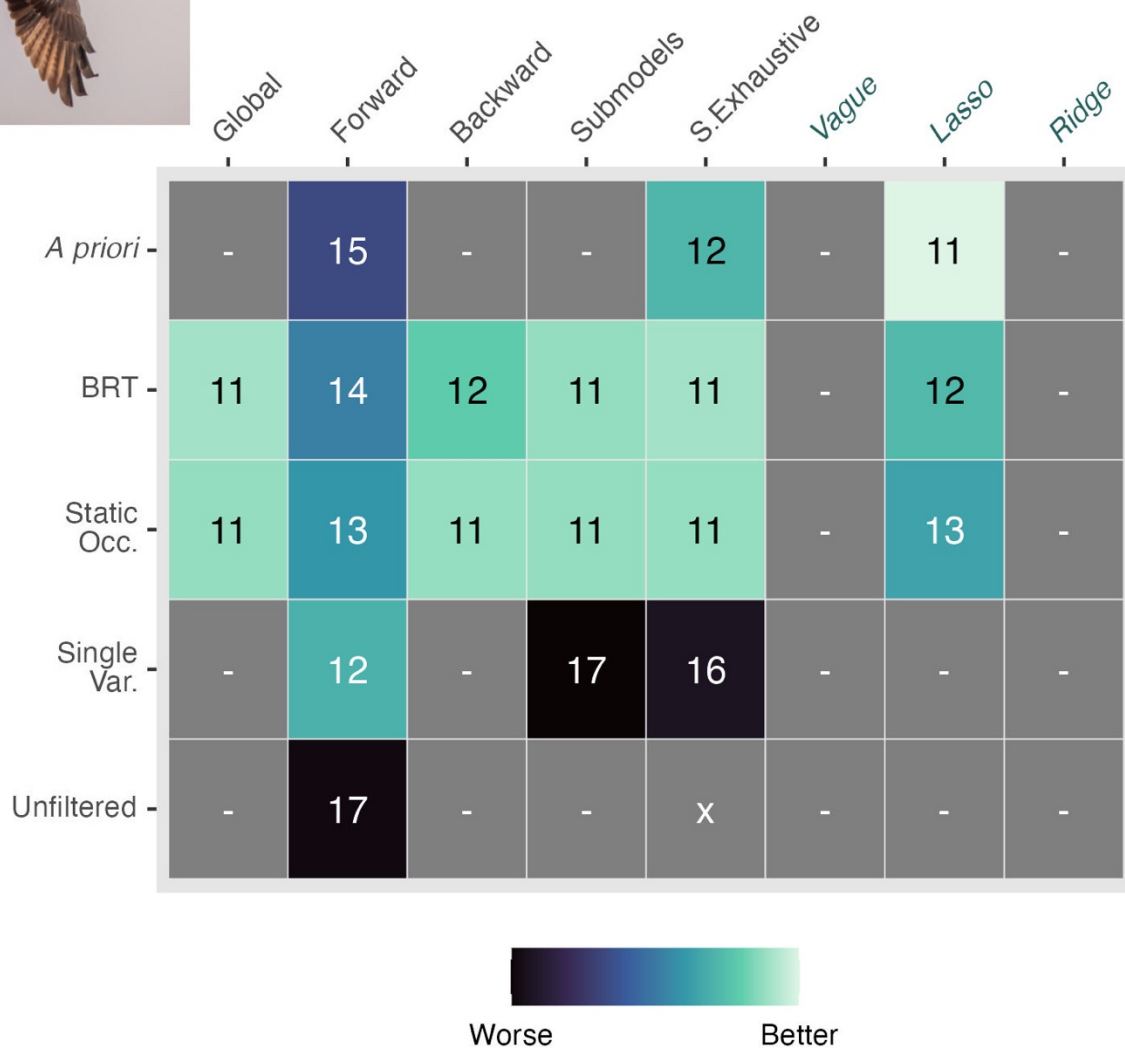


Figure S6-15: Red Kite normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Red-backed Shrike

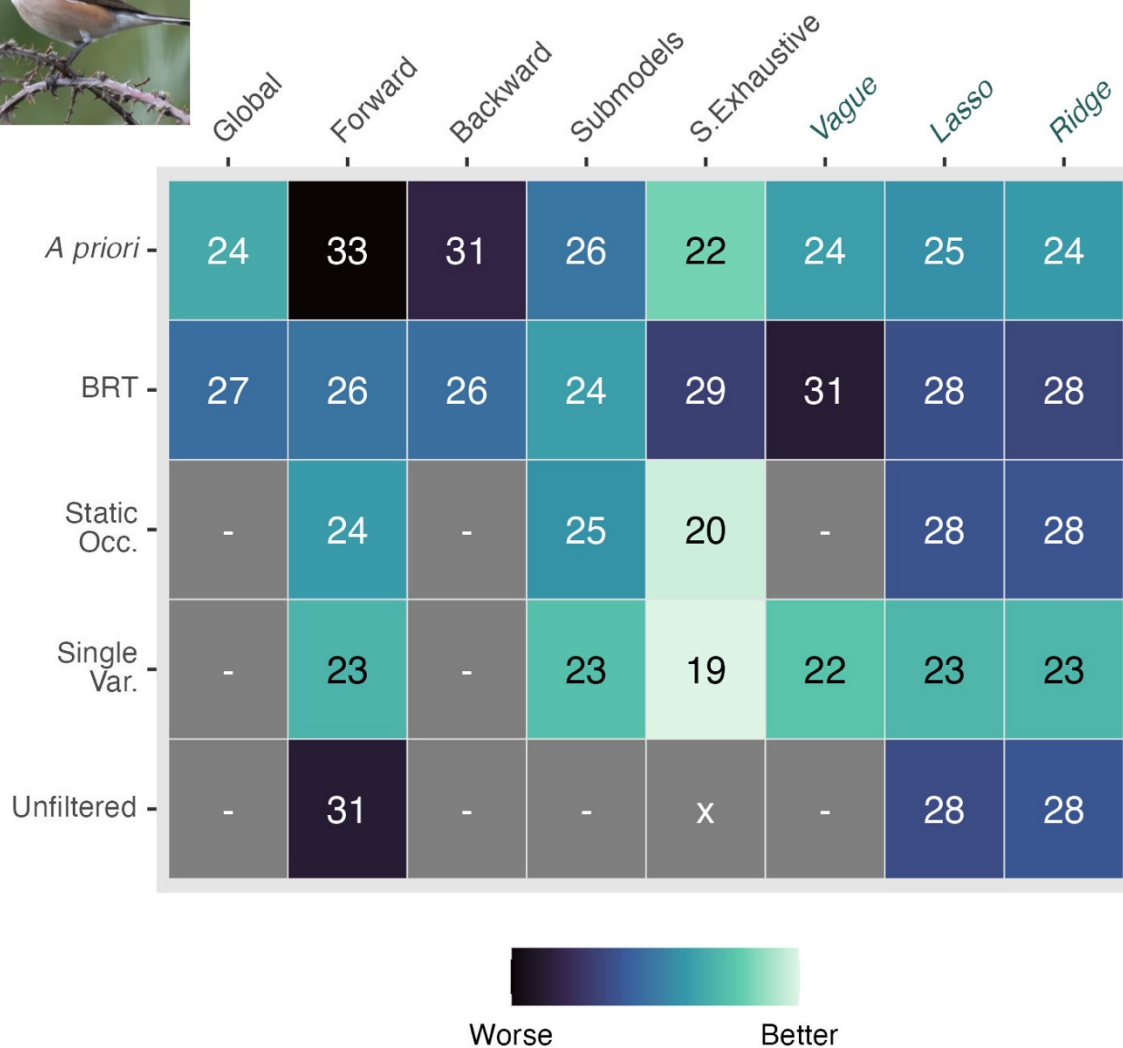


Figure S6-16: Red-backed Shrike normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## Whinchat

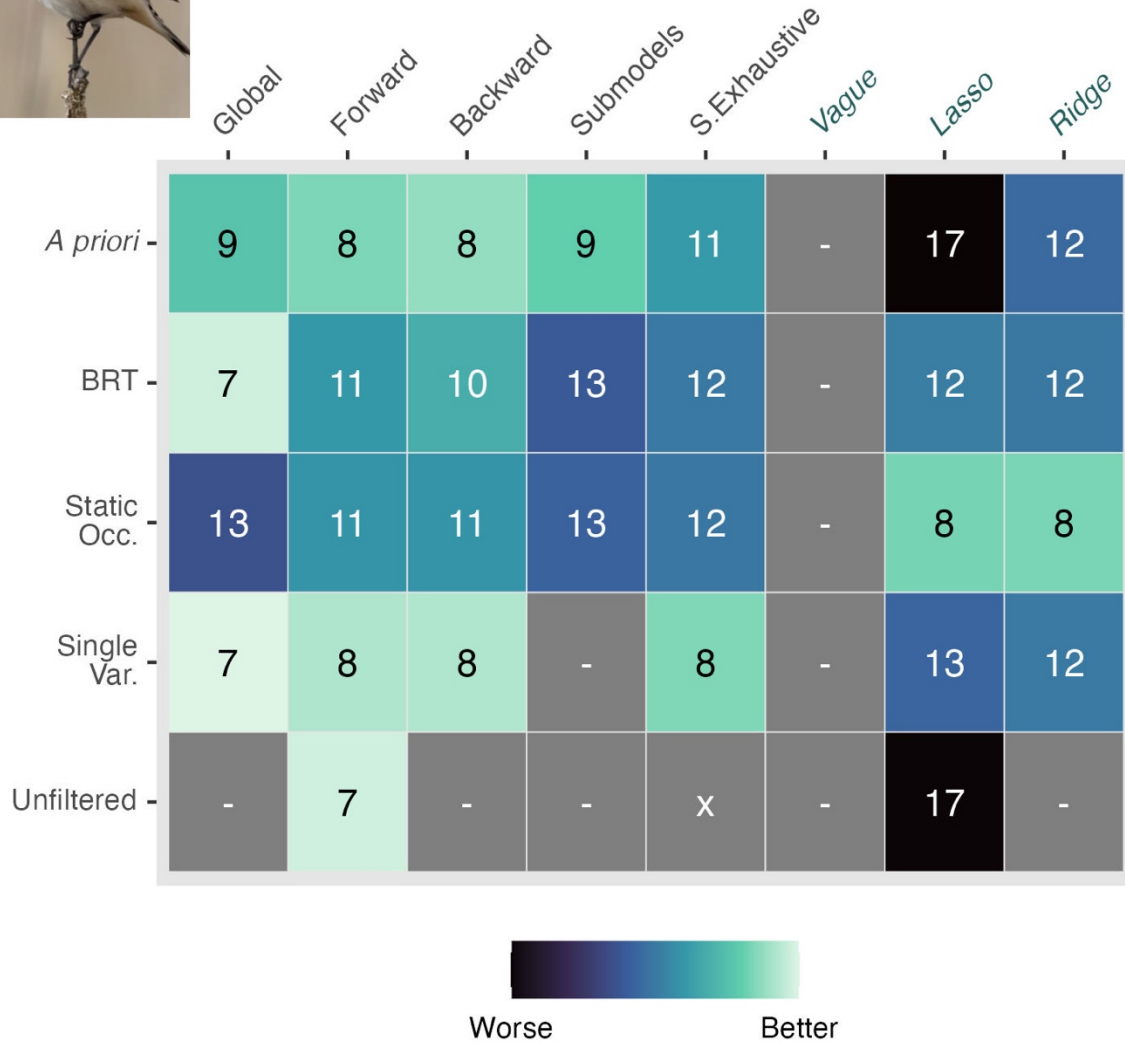
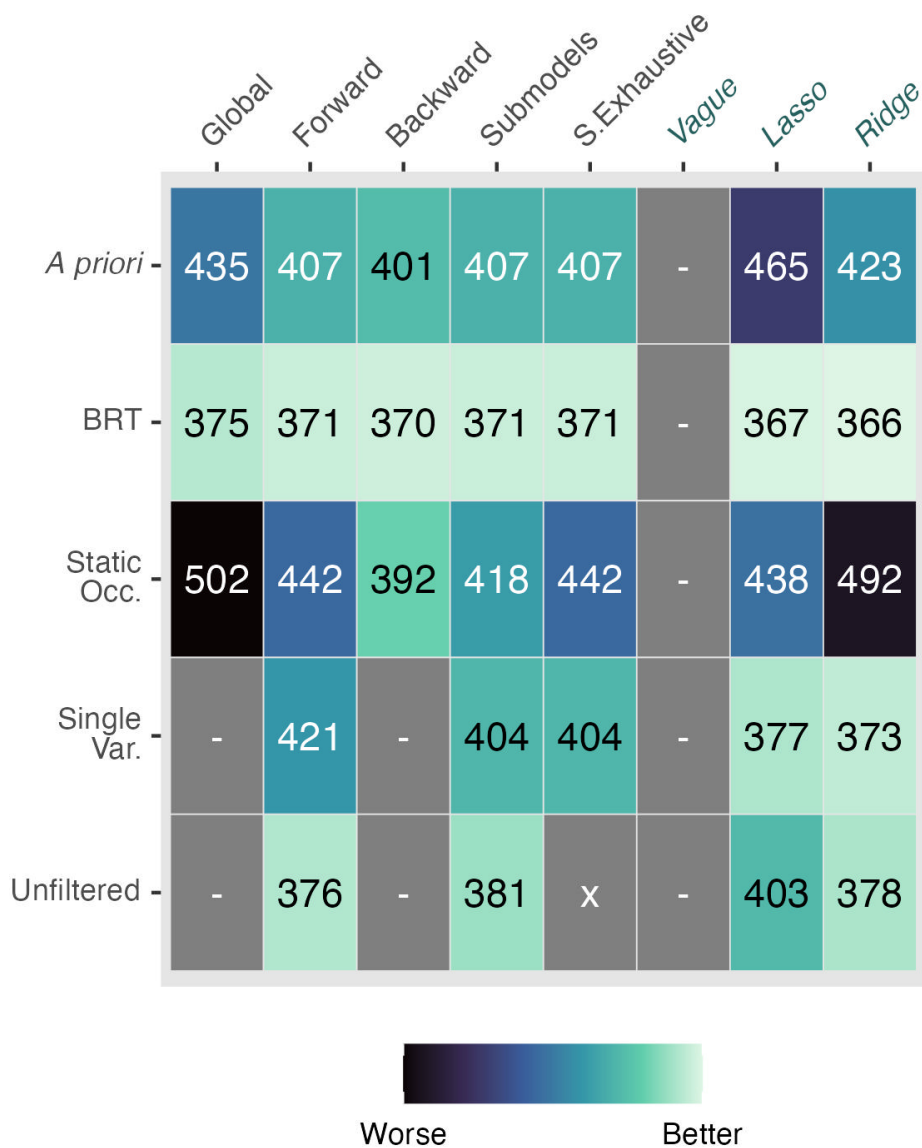
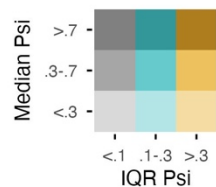


Figure 6-17: Whinchat normalised RMSE for the predicted number of sites with observations across the 10 testing years (2009-2018); i.e., a value of 20 states that the predicted number of sites with observations was off by an average of 20% in each year. '-' indicates workflows that could not be successfully completed, and 'x' represents workflows that were not attempted.



## 6.5. Spatial predictions



# Common Kestrel

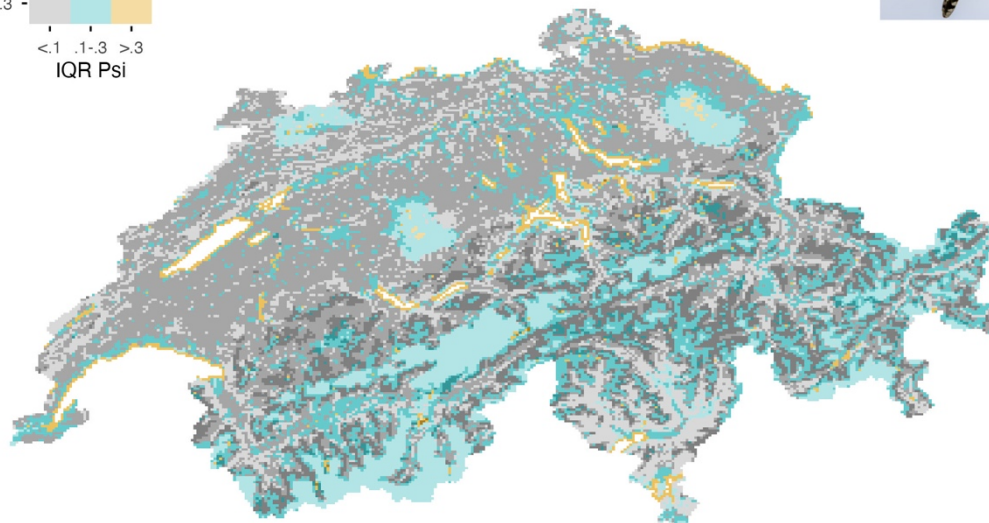
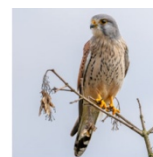


Figure S6-19: Consistency of occupancy estimates across workflows for Common Kestrel. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.

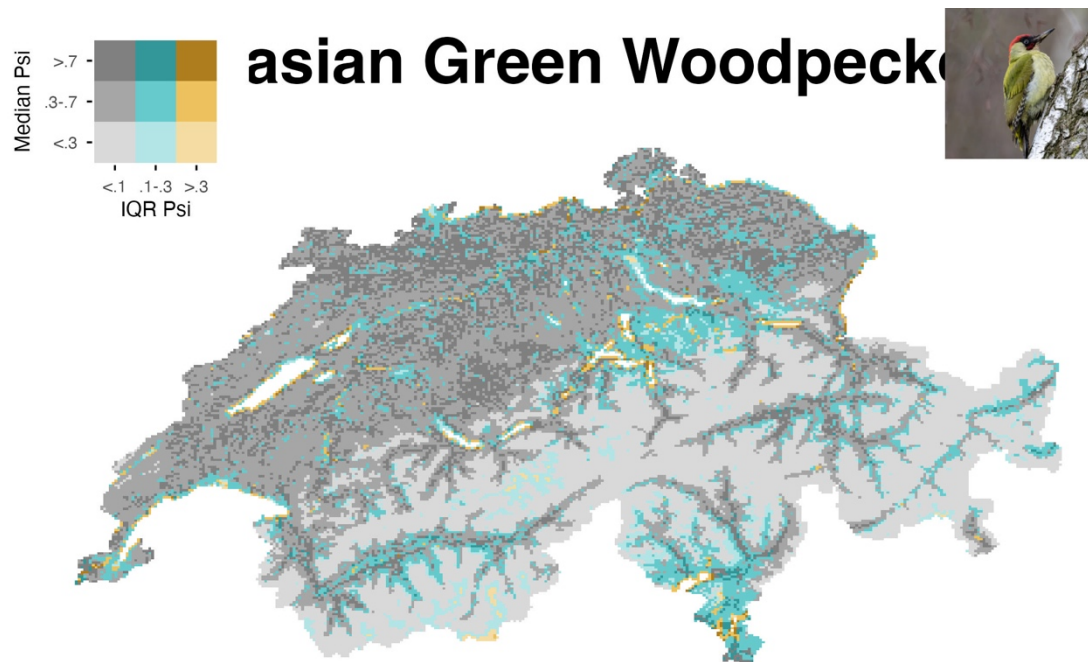
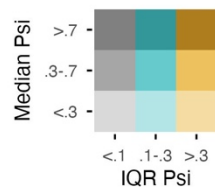


Figure S6-20: Consistency of occupancy estimates across workflows for Eurasian Green Woodpecker. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.



# Eurasian Skylark

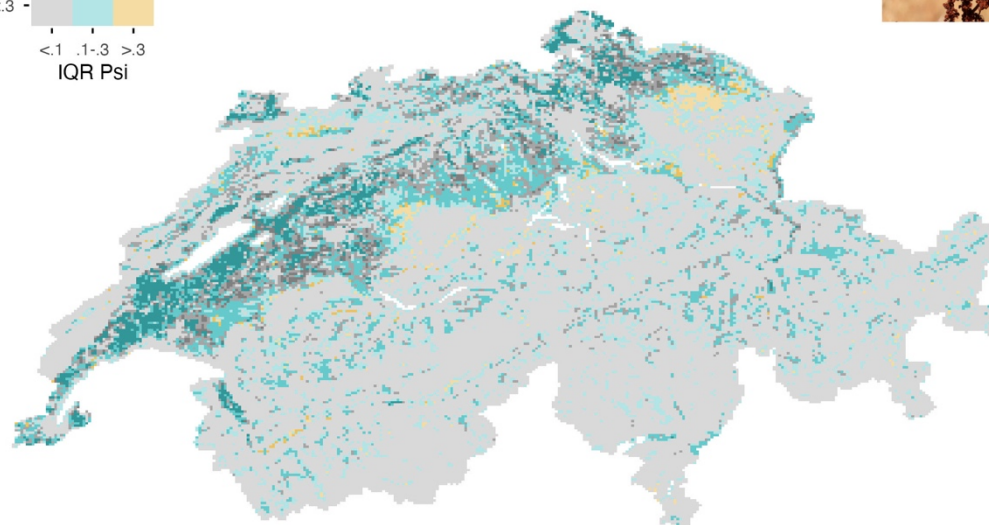
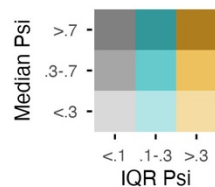


Figure S6-21: Consistency of occupancy estimates across workflows for Eurasian Skylark. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.



# Garden Warbler

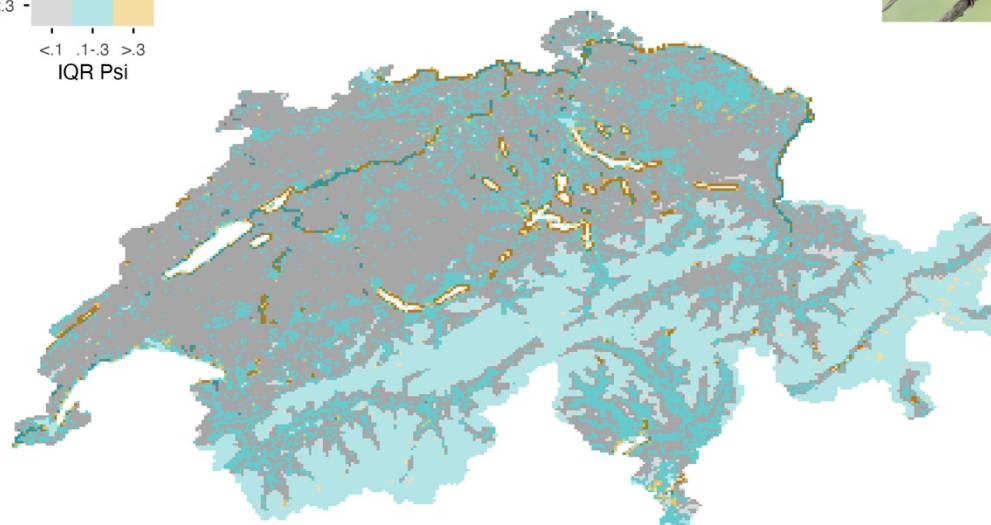


Figure S6-21: Consistency of occupancy estimates across workflows for Garden Warbler. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.

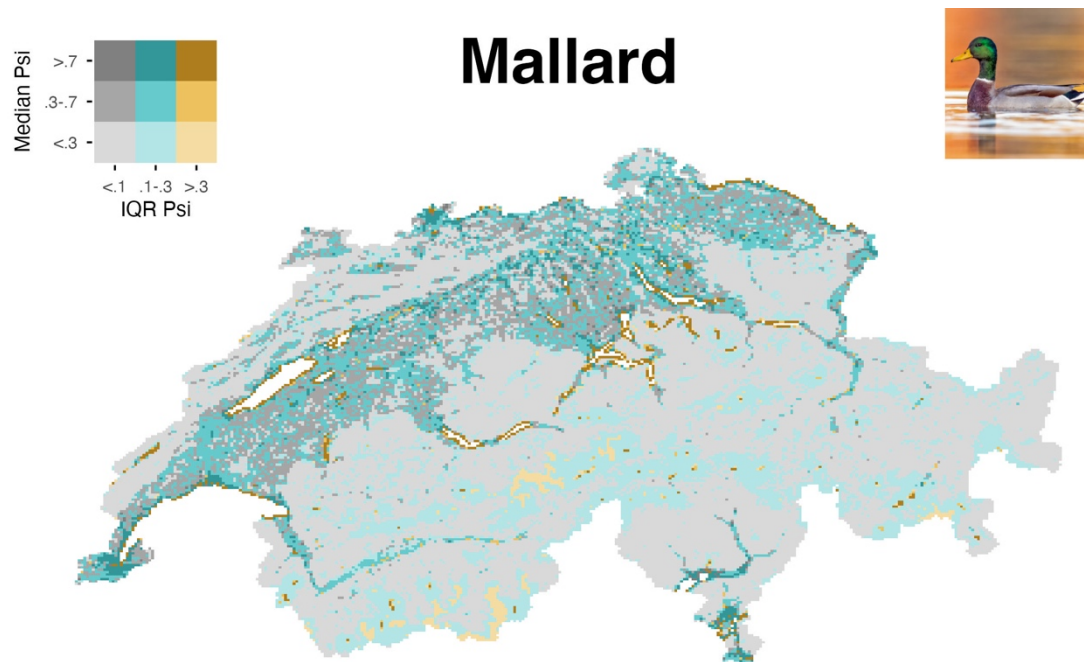
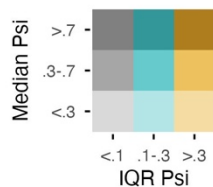


Figure S6-22: Consistency of occupancy estimates across workflows for Mallard. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.



# Red Crossbill

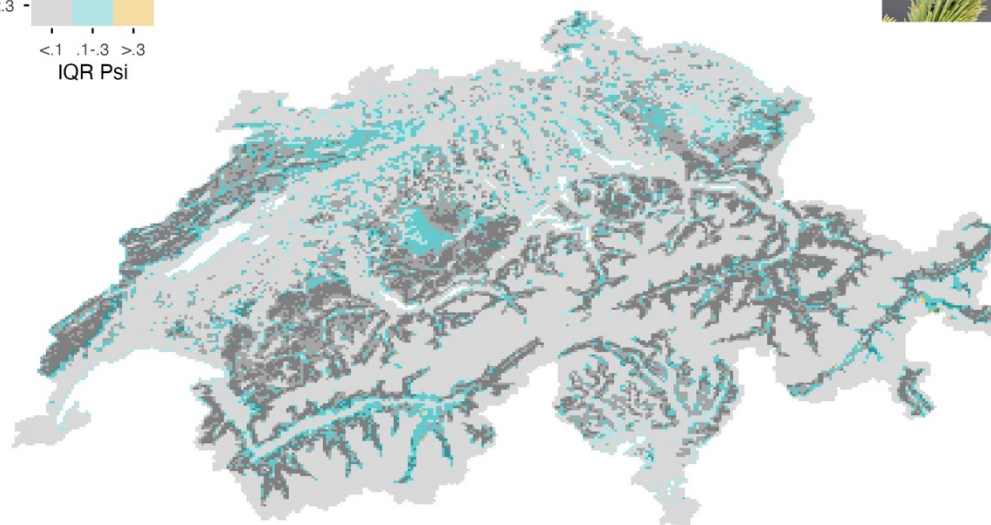


Figure 6-23: Consistency of occupancy estimates across workflows for Red Crossbill. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.

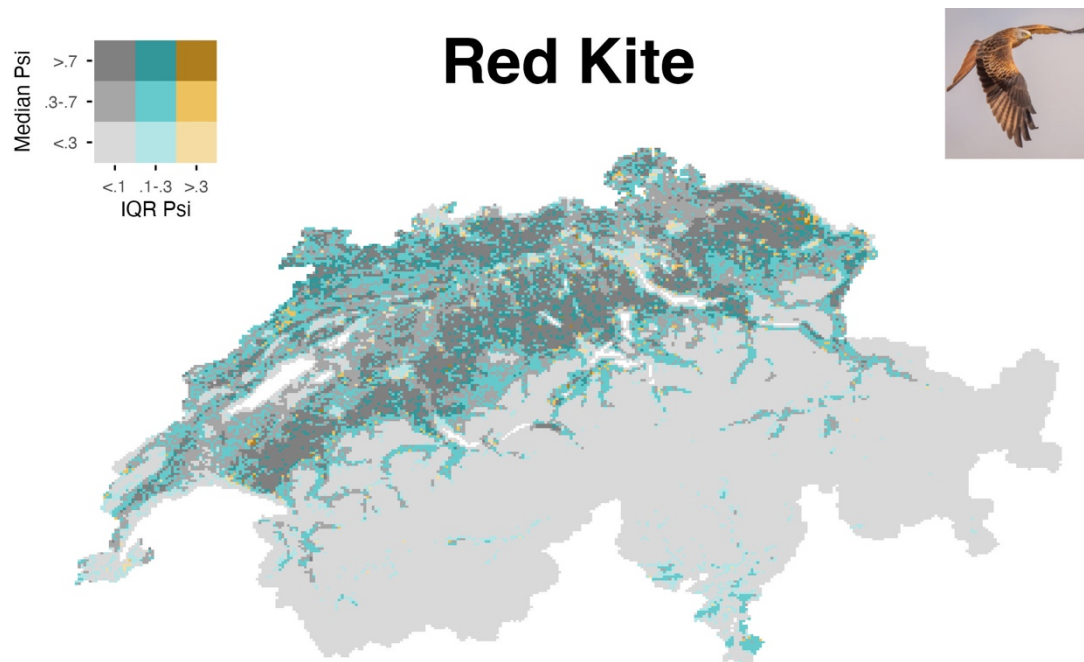
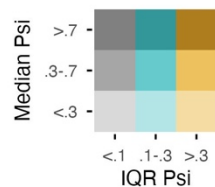


Figure S6-24: Consistency of occupancy estimates across workflows for Red Kite. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.



## Red-backed Shrike

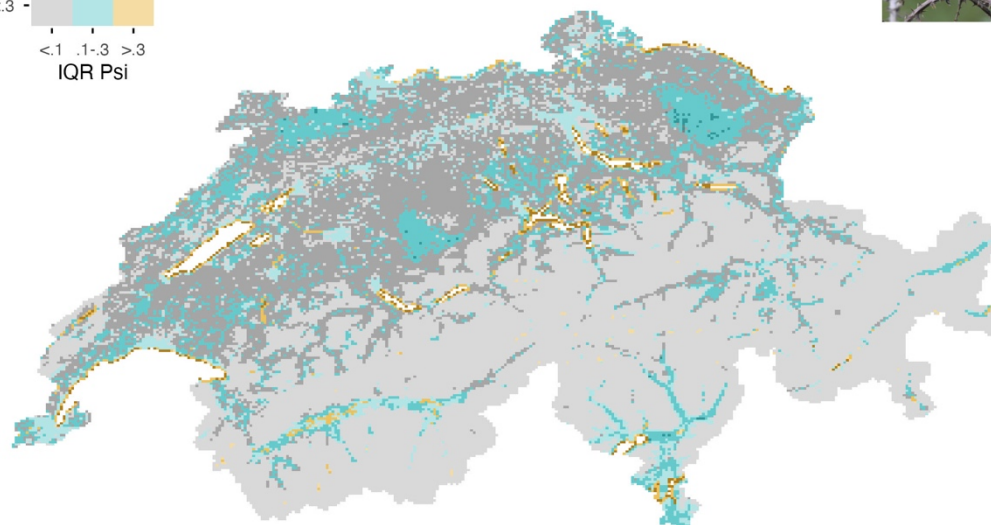
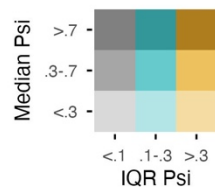


Figure S6-25: Consistency of occupancy estimates across workflows for Red-backed Shrike. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.



# Whinchat

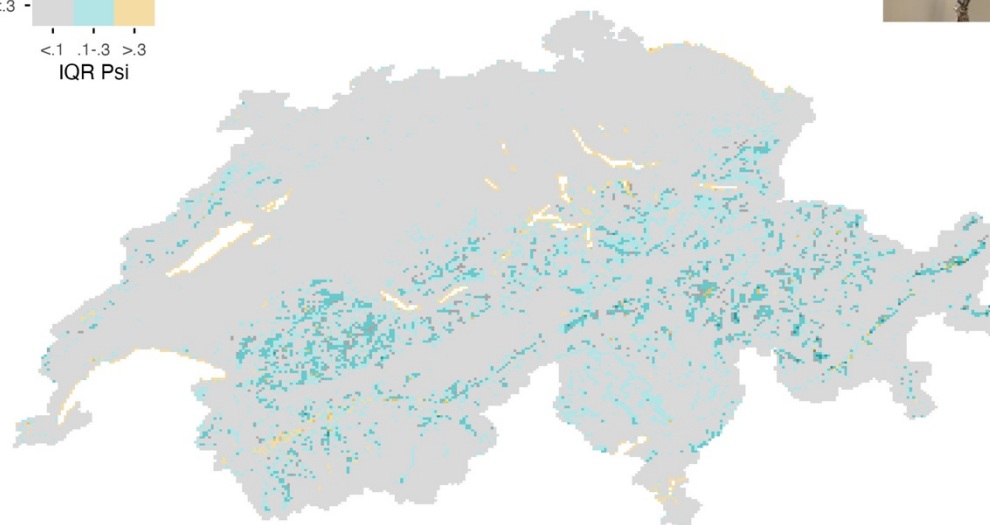
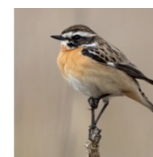


Figure S6-1: Consistency of occupancy estimates across workflows for Whinchat. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.

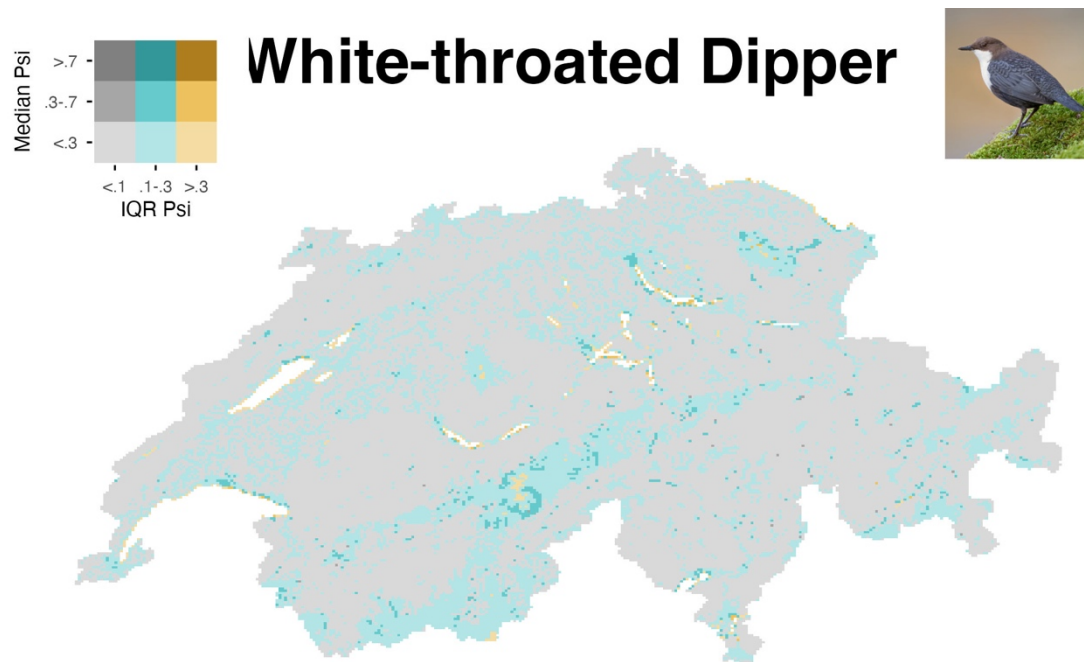


Figure S6-27: Consistency of occupancy estimates across workflows for White-throated dipper. Map represents the median estimated probability of occupancy (2000-2018) for a species across all workflows, and the inter-quartile range of the same occupancy estimates. Cells are grey where estimates are consistent, and cyan where they are more divergent. Darker cells have higher estimated occupancy probabilities.

## References

- Biswas, S. and Lin, S. 2012. Logistic Bayesian LASSO for identifying association with rare haplotypes and application to age-related macular degeneration. - *Biometrics* 68: 587–597.
- Copernicus Land Monitoring Service 2018. CORINE Land Cover.
- Copernicus Land Monitoring Service 2021. Normalised Difference Vegetation Index 1999-2020 (raster 1 km), global, 10-daily - version 3.
- Elith, J., Leathwick, J. R. and Hastie, T. 2008. A working guide to boosted regression trees. - *J. Anim. Ecol.* 77: 802–813.
- Federal Office of Topography swisstopo 2004. DHM 25 - the digital height model of Switzerland.
- Génard, M. and Lescourret, F. 1987. The Common Crossbill *Loxia curvirostra* in the Pyrenees: some observations on its habitats and on its relations with conifer seeds. - *Bird Study* 34: 52–63.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. - *Sci. Data* 4: 170122.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M. 2021. Climatologies at high resolution for the earth's land surface areas.
- Kellner, K. 2015. jagsUI: a wrapper around rjags to streamline JAGS analyses. - R Package Version in press.
- Kellner, K. F., Smith, A. D., Royle, J. A., Kéry, M., Belant, J. L. and Chandler, R. B. 2023. The unmarked R package: Twelve years of advances in occurrence and abundance modelling in ecology. - *Methods Ecol. Evol.* 14: 1408–1415.
- Morin, D. J., Yackulic, C. B., Diffendorfer, J. E., Lesmeister, D. B., Nielsen, C. K., Reid, J. and Schaub, E. M. 2020. Is your ad hoc model selection strategy affecting your multimodel inference? - *Ecosphere* 11: e02997.
- Plummer, M., Stukalov, A. and Denwood, M. 2016. rjags: Bayesian graphical models using MCMC. - R Package Version in press.
- Schmid, H., Zbinden, N. and Keller, V. 2004. Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz.

Strebel, N., Wechsler, S., Bühler, R., Häfliger, G., Keller, V., Kéry, M., Rogenmoser, C., Spiess, M., Varga, K., Volet, B., Zbinden, N. and Schmid, H. 2026. Data of the Swiss Common Breeding Bird Monitoring Program. - Ecology in press.

Swiss Federal Office of the Environment 2025. Forest mix rate NFI.

van Erp, S., Oberski, D. L. and Mulder, J. 2019. Shrinkage priors for Bayesian penalized regression. - J. Math. Psychol. 89: 31–50.