

Title: Considering more complex covariate responses improves predictive performance in dynamic occupancy models

Authors: Saoirse Kelleher¹, Gurutzeta Guillera-Arroita^{2,1}, Marc Kéry³, Jane Elith¹, Natalie J. Briscoe¹

¹University of Melbourne, School of Agriculture, Food, and Ecosystem Sciences,
Parkville, Victoria, Australia

²Pyrenean Institute of Ecology, Spanish National Research Council (CSIC), Jaca, Spain

³Swiss Ornithological Institute, 6204 Sempach, Switzerland

Corresponding author: Saoirse Kelleher (Saoirse.kelleher@unimelb.edu.au)

Open research statement: Data and code used for this article is at

<https://figshare.com/s/1f85141930712897e5ef>. A permanent DOI has been reserved at
10.26188/32284755.

Keywords: Covariate selection; distribution models; model complexity; occupancy dynamics;
occupancy models; predictive performance

Abstract

1 Finding the right level of model complexity is a key challenge in most contexts, including
2 ecology where data are often messy and underlying patterns complex. Dynamic occupancy
3 models (DOMs) are increasingly used for modelling species distributions. DOMs have
4 advantages over traditional species distribution models because their hierarchical structure
5 allows them to explicitly model changes in occupancy over time while accounting for imperfect
6 detection. The modelled processes (initial occupancy, colonisation, extinction, and detection) can
7 be described as a function of covariates to account for heterogeneity in occupancy and detection
8 dynamics. While this model structure is more complex than most alternative distribution models,
9 in practice users of DOMs tend to fit simple linear responses to covariates with only infrequent
10 use of non-linear responses or interactions. However, where data suffices and where ecologically
11 justified, permitting more complex relationships between covariates and species occurrence may
12 yield greater predictive power. Here, we assess how the complexity of covariate relationships
13 affects the predictive performance of DOMs fit to Swiss breeding bird data. We find that models
14 allowing quadratics and interactions tend to provide superior performance in predicting future
15 occurrence, and that some covariates are only informative predictors when represented using
16 non-linear relationships or interactions. Relying solely on linear relationships alone is not only
17 inconsistent with ecological theory, but also potentially costly for predictive ability. To improve
18 predictive power and the utility of these models for identifying important habitat and species at
19 risk of declines, users of dynamic occupancy models should carefully consider not only their
20 choice of candidate covariates, but also the functional forms of covariate relationships.

Introduction

Species distribution patterns across landscapes result from many interacting factors of varying importance. A frequent challenge in quantitative ecology is to build models that adequately capture key factors to provide useful insights into patterns of species occurrence, whether the ultimate objective is to identify what drives these patterns or to predict where species may occur under future conditions. All models are abstractions, and even the most intricate models do not approach the true complexity of natural systems. Determining precisely how much complexity is needed to build useful models is a key modelling challenge, and there are long-standing, unresolved debates over the relative merits of simple versus complex models in any given context (Merow et al. 2013, 2014, Lonergan 2014).

The term ‘complexity’ encompasses several aspects of a model, such as whether its structure is simple or contains multiple sub-models, the number of parameters that are estimated, and the form of the statistical relationships between covariates and response variables (Merow et al. 2014). For example, conventional correlative species distribution models (hereafter SDMs) often directly relate species occurrence records to covariates representing environmental variables, a relatively simple model structure that typically does not account for factors like observation error or non-equilibrium conditions (Elith et al. 2010, Kellner and Swihart 2014). At the same time, many of the common approaches used for fitting SDMs, such as MAXENT and boosted regression trees (Elith et al. 2008, Merow et al. 2013), can adopt flexible non-linear relationships allowing for complex functional responses to variation in environmental conditions.

The occupancy-detection modelling framework offers an increasingly popular alternative to SDMs for describing patterns of species distributions (Kéry 2011, Guillera-Aroita 2017,

Goldstein et al. 2024). Using a hierarchical model structure with sub-models for the probability that the species occurs at a site and the probability that the species is detected where present, occupancy-detection models can improve estimates of species occurrence by adjusting for imperfect detection (MacKenzie et al. 2002, Tyre et al. 2003). Dynamic occupancy models (hereafter DOMs) extend this approach over multiple time-steps by adding additional complexity that allows estimates of changes in species occupancy as a result of site colonisation and local extinction events (MacKenzie et al. 2003). The more complex model structure of DOMs provides advantages over SDMs in many contexts by accounting for common biases induced by observation error and non-equilibrium conditions. As a result, DOMs have been promoted as particularly useful for predicting future species distributions, a use-case where standard SDMs still remain the predominant approach despite their known limitations (Kéry et al. 2013, Briscoe et al. 2021). However, most DOMs often use relatively simple linear relationships between covariates and modelled processes (Kelleher et al. 2025, but see Joseph 2020) — this contrasts with the frequent inclusion of non-linear terms in SDMs, and the consequences of this discrepancy on predictive ability remains unexplored.

Many of the hypothetical ecological relationships represented in models are at least theoretically non-linear in nature, particularly when the area where surveys are conducted includes the full ecological niche of a species (Austin 2007). For instance, all organisms have a fundamentally parabolic relationship with temperature between their critical limits, although this relationship may be locally linear if only a portion of the thermal niche is observed. Interactions between environmental factors are also widespread in complex environments for both flora and fauna species at small and large scales (Austin and Meyers 1996, Smith 2013, Catford et al. 2022, Spake et al. 2023). Prior research in SDMs has indicated that models capable of fitting flexible

relationships between occupancy and environmental variables may offer stronger predictive performance when underlying distribution patterns are complex, provided that mechanisms are in place to control the level of flexibility, prevent over-fitting, and avoid unjustifiably complex responses (Evans and Cushman 2009, Santika et al. 2014, Valavi et al. 2022, 2023).

While discussions and explorations of the merits of simplicity versus complexity in covariate responses have been prominent in the SDM literature for decades, much less such work has been conducted in the context of DOMs. Given the similarity in applications and theoretical underpinnings, one might expect to find similar connections in the relationship between complexity and predictive performance. However, it is also conceivable that the differences in the complexity of the overarching model structures of DOMs and SDMs may modify these relationships and support the use of comparatively simpler covariate responses.

We used a long-term breeding-bird dataset to assess how incorporating greater complexity in covariate responses impacts the predictive performance of dynamic occupancy models. For a selection of 10 species with contrasting ecologies and dynamics, we fit models permitting i) only linear responses to covariates, ii) both linear and quadratic responses, iii) interactions between covariates, and iv) interactions and quadratics. Our evaluation addresses a key methodological knowledge gap, providing guidance to modellers seeking to apply this increasingly popular modelling approach to predict species distributions.

Methods

We used the original parameterisation of the dynamic occupancy model as described by MacKenzie et al. (2003). This model assumes that sites are surveyed across time (in multiple ‘primary occasions,’ e.g., years), and that there is some replication within primary occasions

(multiple ‘secondary occasions’ in at least some sites, most often from repeated surveys). The model describes the occupancy state of each site i during each primary occasion t as either occupied ($z_{i,t} = 1$) or unoccupied ($z_{i,t} = 0$). For the first primary occasion, site occupancy is described as a Bernoulli trial with probability of initial occupancy $\psi_{i,t=1}$. Changes in site occupancy across time are described using a first-order Markov process so that for subsequent primary occasions, site occupancy is determined by the site’s occupancy state in the previous primary occasion and the probabilities of colonisation $\gamma_{i,t}$ and local extinction $\epsilon_{i,t}$:

$$z_{i,t=1} \sim \text{Bern}(\psi_1)$$

$$z_{i,t} \sim \text{Bern}(\gamma_{i,t-1}(1 - z_{i,t-1}) + (1 - \epsilon_{i,t-1})z_{i,t-1})$$

During secondary occasions j within a primary occasion t , a species is either detected at a site i ($y_{i,t,j} = 1$) or not ($y_{i,t,j} = 0$) depending on the occupancy state and the probability of detection $p_{i,t,j}$:

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

The standard DOM assumes that no false positive detections occur, and that the occupancy state of a site does not change within, but only between primary occasions. Each of the four basic parameters (referred to as ψ_1 , γ , ϵ , and p) can be modelled with respect to covariates through a logit link function, such that a typical DOM can be thought of as consisting of four connected generalised linear models (MacKenzie et al. 2017). In this study, we fit DOMs by maximum likelihood estimation in *R* using the *unmarked* package version 1.5.0.9002 (Kellner et al. 2023, R Core Team 2023).

Breeding bird data

We used freely available data from the Swiss Ornithological Institute’s long running breeding bird study (Strebel et al. 2026), which includes detection histories for 267 sites across

108 Switzerland (41,285 km²) that have been surveyed 2-3 times a year since 1999. Each site
109 consists of a 1x1 kilometre grid cell. During surveys, trained volunteers walk 4-6 km long
110 transects covering all habitat types within the site and record all observed bird species. Extended
111 details of the dataset are available in Schmid et al. (2004 and Strebel et al. (2026). For this study,
112 we fit models for ten species selected based on their variation in life history characteristics,
113 known distributions in Switzerland, and observed trends in occupancy. For each species, we
114 discarded any surveys prior to a species-specific Julian date corresponding to the end of spring
115 migration to avoid including surveys conducted before arrival or detections of transient, non-
116 breeding individuals. We used 19 years of detection data: 2000-2008 were used as training data
117 to fit models, with 2009-2018 retained as testing data for assessing model predictive
118 performance.

119 *Candidate covariates*

120 We developed models for all species using the same pool of candidate covariates. We included
121 five covariates derived from the CORINE land cover database (Copernicus Land Monitoring
122 Service 2018), representing the percent cover of arable land, artificial surfaces, forest, grassland,
123 or heterogeneous land cover within a 3x3 kilometre square centred on each survey site. We used
124 this larger window to capture information about the habitat surrounding each site at a distance
125 relevant to the ecology of mobile bird species. We also included the distance to the nearest body
126 of water (also from CORINE), measured from the centre of each site. As CORINE data is
127 updated every six years, we linearly interpolated each of these covariates for intervening years .
128 To capture climatic factors, we calculated two Bioclim covariates using CHELSA monthly
129 climatologies (Karger et al. 2017, 2021): Bio10 summarises the mean temperature of the
130 warmest quarter, and Bio12 summarises total annual precipitation; each was averaged between

year t and $t-1$ to account for lagged climate impacts. We also derived mean normalised derived vegetation index (NDVI) and the standard deviation of NDVI for each site over year t and $t-1$ using Copernicus land monitoring service data (Copernicus Land Monitoring Service 2021). We calculated the mean elevation of each site using the Swiss digital elevation map (Federal Office of Topography swisstopo 2004). We standardised all covariates using the mean and standard deviation of the covariates at sites during the training period (2000-2008). Covariates fell into either land cover, climate or detection-related categories that determined how they would be incorporated into the models.

Modelling framework

We compared the following four sets of assumptions regarding the potential complexity of covariate responses on the parameters in the occupancy submodel (Ψ_1 , γ , and ϵ):

- *Linear-only (LO)*: Permitting only linear relationships between covariates and the occupancy-related processes, e.g. $\psi_{i,1} = \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Frst}} * \text{Frst}_i)$
- *Linear/Quadratic (LQ)*: Permitting a linear and a quadratic response for each covariate, e.g. $\psi_{i,1} = \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Frst}} * \text{Frst}_i + \beta_{\text{Frst}^2} * \text{Frst}_i^2)$. Where the quadratic form was used for a covariate, the linear term for the covariate was also included to respect rules of marginality in linear models (McCullagh and Nelder 1989).
- *Linear/Interaction (LI)*: Permitting the linear response and interactions between a subset of the covariates, e.g., $\psi_{i,1} = \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Frst}} * \text{Frst}_i + \beta_{\text{Arbl}} * \text{Arbl}_i + \beta_{\text{Frst} * \text{Arbl}} * \text{Frst}_i * \text{Arbl}_i)$. Where an interaction was used, the two corresponding linear terms were also included.

- *Linear/Quadratic/Interaction (LQI)*: Permitting linear responses, quadratic responses, and the same subset of interactions as in 3. E.g., $\psi_{i,1} = \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Frst}} * \text{Frst}_i + \beta_{\text{Frst}^2} * \text{Frst}_i^2 + \beta_{\text{Arbl}} * \text{Arbl}_i + \beta_{\text{Frst} * \text{Arbl}} * \text{Frst}_i * \text{Arbl}_i)$.

Our focus was on predicting occupancy dynamics, rather than detectability. For the detection parameter p we considered five pre-defined formulations following prior research on this dataset:

$$p \sim \text{logit}(\alpha_{\text{intercept}}); p_{i,t,j} \sim \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Date}} * \text{Date}_{i,t,j}); p_{i,t,j} \sim \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Date}} * \text{Date}_{i,t,j} + \beta_{\text{Date}^2} * \text{Date}_{i,t,j}^2); p_{i,t,j} \sim \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Date}} * \text{Date}_{i,t,j} + \beta_{\text{Elev}} * \text{Elev}_i + \beta_{\text{Date} * \text{Elev}} * \text{Date}_{i,t,j} * \text{Elev}_i); \text{ and } p_{i,t,j} \sim \text{logit}(\alpha_{\text{intercept}} + \beta_{\text{Date}} * \text{Date}_{i,t,j} + \beta_{\text{Date}^2} * \text{Date}_{i,t,j}^2 + \beta_{\text{Elev}} * \text{Elev}_i + \beta_{\text{Date} * \text{Elev}} * \text{Date}_{i,t,j} * \text{Elev}_i)$$

For each species, we conducted a model selection process using each of the four sets of assumptions and a forward-stepwise selection process. This approach is also among the most common covariate selection strategies for DOMs in the literature (Morin et al. 2020, Kelleher et al. 2025). Commencing with the intercept-only model $p(\cdot)\Psi_1(\cdot)\Upsilon(\cdot)\boldsymbol{\epsilon}(\cdot)$, we first fit each of the five candidate formulae for p and fixed p to the formula that resulted in the lowest AIC. Next, we tested the addition of each eligible covariate under a given set of assumptions to the formula for Ψ_1 , retaining the covariate which produced the model with the lowest AIC and repeating until further additions no longer reduced AIC. The same process was then repeated for Υ , and finally $\boldsymbol{\epsilon}$, until a final model was reached. Each attempted formula was run with at least six sets of starting values; if the formula could not be fit without estimation issues, up to 20 sets of starting values were tried before moving on.

Where permitted, interactions were considered between any two climate covariates and any combination of landuse and climate covariates (Table 1). We precluded any interactions between

two land cover variables because these were considered less likely and their exclusion reduced the number of possible interactions to be considered. Where any interaction was included in a model, each of the corresponding linear terms was also included. Sets of assumptions that permitted either quadratic terms or interactions did not preclude the exclusive inclusion of linear terms that produced lower AICs, i.e., the *LQ* set of assumptions could have produced a model with exclusively linear terms if that formulation produced the lowest AIC.

After identifying a final model for each species and set of assumptions, we predicted the probability of occupancy for each site from 2000-2019. We then used predicted detection probabilities to calculate the probability of recording at least one observation of the species at a site in a year, given the same survey dates that were actually used during that period. We evaluated predictive performance by comparing the probability of recording at least one detection against whether the species was actually observed at that site and year because occupancy is a latent state that is not directly observed. We measured site-level predictive performance using four metrics recommended by Abrego and Ovaskainen (2023), reflecting various aspects of classification accuracy - the area under the receiver operating curve (AUC), True Skill Statistic (TSS), Cohen's Kappa, and Tjur's R^2 . We also included an additional metric, the Brier score, which is widely used in the machine learning literature to assess classification models and is roughly analogous to the root-mean-squared error (RMSE; Rufibach 2010).

Results

The number of terms included in each model varied strongly based on the maximum permitted complexity (Figure 1). *LO* models used the fewest coefficients, with a median of 10.5 estimated coefficients (the sum of all bars in Figure 1). *LQ* models used a median of 18.5 parameters and *LI* models used a median of 37 parameters. The *LQI* models used the most parameters with a

median of 41.5 (Figure 1). Where they were available, interactions were generally included with greater frequency than quadratics, i.e., *LI* used more interactions than *LQ* used quadratics for most species. Regardless, in the *Linear/Quadratic/Interaction* models, both quadratic terms and interaction terms were included for occupancy processes for all species.

The specific covariates included in each of the models also varied by the types of complexity permitted. Figure 2 shows whether each model included the regression coefficient for the linear term of a covariate, regardless of whether a quadratic or interaction term for that covariate was also included. In many cases, a covariate was not included in the simplest *LO* model for a species despite being included in the *LI* or *LQI* models. This can be observed in the case of the white-throated dipper, for example; of the 30 possible linear terms (10 for each of ψ_1 , γ , and ϵ), 10 were used in the *LO* model but 22 were used in the *LQI* model.

Predictions of the number of sites with detections in each year varied among models fit under different complexity assumptions (Figure 3). In some cases, the divergence in the predicted trends was considerable — for example, the *Linear-only* model for the Whinchat failed to capture the downward trajectory in the later part of the testing period, whereas the other three models were in closer alignment with the observed trajectory. In other cases, all four models made similar predictions, as with the Eurasian Green Woodpecker or Mallard. None of the four models produced uniformly superior predictions as measured by RMSE during the testing period.

As expected, the *Linear/Quadratic/Interaction* models reached the lowest AIC for all species except the Mallard (Table 2). Conceptually the possible *LI* models are a subset of possible *LQI* models and thus the lowest-possible AIC model should be available via the *LQI* method; this

model was not reached by the LQI method for the mallard due to idiosyncrasies in the multi-stage covariate selection procedure. The *Linear-only* and *Linear/Quadratic* models each had a ΔAIC of at least 30 and thus predicted far less well.

On average, the *Linear-only* models offered the lowest performance (mean rank 3.34 across species and metrics), followed by the *Linear/Quadratic* models (2.86) and *Linear/Interaction* models (1.94) (Figure 4). *Linear/Quadratic/Interaction* models gave the highest average performance (1.86) though they were not uniformly superior, having the lowest performance on 8/50 metrics (5 metrics x 10 species).

While there were reductions in predictive performance between the training and testing period, models fit under higher versus lower permitted levels of complexity did not appear to suffer greater drops in performance as might be expected had overfitting occurred (Figure 5). Where the higher permitted complexity models achieved poorer predictive performance on testing data than the simpler models, differences in AUC and other metrics were relatively small. In other cases, like the White-throated Dipper or Red Crossbill, the *Linear/Quadratic/Interaction* models made markedly better predictions than the simplest *Linear-Only* models based on testing AUC.

While the accuracy of spatial predictions could not directly be assessed at unsurveyed locations, comparisons of models fit under the lowest and highest levels of permitted complexity demonstrate that the outputs of the *Linear-Only* and *Linear/Quadratic/Interaction* models differed meaningfully for some species (Figure 6). Large regions of Switzerland had divergent predicted occupancy levels between the two models, in some cases predicting high probability of occupancy (>70%) within a site for one model while predicting a low probability of occupancy (<30%) in the other.

Discussion

Widely acknowledged trade-offs between the ability of complex models to capture patterns and the generalisability of simpler models remain an important consideration in ecological modelling (Merow et al. 2014). Including complex responses to covariates that are not based on sound ecological rationale risks over-fitting and poor extrapolation, and in some ecological models simpler models yield superior predictive performance (Ward et al. 2014). This is particularly true when the available data is limited and noisy, as is the case with many ecological applications (Ward et al. 2014). However, the amount of data available for modelling has grown considerably over recent decades, and DOMs are now routinely fit to large quantities of data from study systems that cover large areas (Kelleher et al. 2025). Despite this, it is very common for DOM implementations to consider no non-linear relationships or interactions between covariates on any model parameter (Kelleher et al. 2025). The results presented here suggest that preventing these more complex relationships between the environment and the occupancy dynamics can limit the predictive performance of DOMs, even in a relatively small country like Switzerland., and that decisions regarding the complexity of modelled responses can meaningfully alter predictions of species trends and spatial occurrence across the landscape.

We found that DOMs that permitted either quadratic relationships or interactions between covariates during model selection offered greater predictive performance than those allowing only simple linear relationships. In particular, models that considered interactions between covariates consistently ranked higher on metrics of relative model fit like AIC (Table 2). For this study system and dataset, it appears clear that to maximise model fit and predictive performance it is necessary to consider relationships beyond simple linear trends. This finding mirrors previous work from both the DOM and the SDM literature, which has often found that

permitting more complex responses improved predictive performance (Joseph 2020, Valavi et al. 2022, 2023).

Models that permitted more flexible covariate responses also used a substantially larger number of terms than the simplest linear-only models. While this is expected given the increase in the number of possible terms, it is notable that even if only coefficients for *linear* terms are counted, more covariates were included when more complexity was permitted (Figure 2). It appears that without considering interactions or quadratic terms, limited information was contained in the linear terms alone for many covariates, resulting in the selection of even simpler models. The contributions of many covariates seem to be realised only when a term is included in an interaction or quadratic relationship that may more realistically represent its underlying relationship. This is seen particularly clearly in cases where specific covariates were included in the *Linear/Interaction* model for a species but not the *Linear-only* model (Figure 2). In an applied context, this could easily be interpreted as a given covariate being uninformative where in reality the functional form was simply misrepresented. When considering possible model formulations, it is necessary to consider not only which covariates to include but also which functional forms could reasonably reflect the underlying ecological theory (Austin 2002, 2007).

The inclusion of interactions between covariates was more impactful than the inclusion of quadratics terms, although both improved over *Linear-only* models. It is conceivable that this is in part specific to this case study, given the Swiss topography and the way that elevation shapes nearly all aspects of the environment. Because of this, in some cases the linear term for one covariate may capture the quadratic form of another relatively well. For example, areas with a high proportion of forest cover tend to fall between low elevation valleys and high elevation alpine regions. These areas also tend towards intermediate summer temperatures. As a result, an

expected quadratic response to temperature could be reasonably approximated by the forest cover covariate, reducing the need to include quadratic forms to achieve optimal performance. Where the environmental context differs – for instance in large continental-scale studies where elevation is not a dominant factor – quadratic terms may play a greater role.

We found that initial occupancy typically used the fewest terms, despite being the first occupancy process to be fixed in our sequential model selection protocol. Conversely, despite being fixed last, extinction was on average fitted with the highest number of terms. Whilst we do not know the underlying reason for this pattern, it is possible that initial occupancy is less important for highly mobile species like birds with high turnover, or our covariate pool was better suited for capturing the colonisation and local extinction parameters. Alternatively, this pattern could be an artefact of the order in which parameters were fixed — prior work by Morin et al. (2020) found that varying the parameter-fixing order can impact the final model selected.

Allowing higher model complexity often raises concerns about over-fitting, particularly if the possible relationships are not well grounded in realistic ecological hypotheses. Most DOMs fit by maximum-likelihood estimation are selected using AIC, a scoring rule which is intended to balance model complexity and relative fit (although it is certainly still possible to over-fit models fit by AIC, ((Anderson et al. 1994, Aho et al. 2014)). In our case study, we also assessed the predictive performance decay on training data versus testing data and did not observe evidence of over-fitting in the models that permitted greater complexity. This could be, at least partly, because we constrained the possible set of interactions to avoid less realistic combinations such that the remaining terms were generally plausible. However, the levels of complexity tested here are well above those seen in most applications of DOMs. A recent review found that fewer than half of DOMs in the published literature use either quadratic terms or interactions, with most

resembling the *Linear-only* models fit in this case study (Kelleher et al. 2025). In a more realistic application, an informed modeller could likely further refine the list of plausible quadratic relationships or interactions to better align with a species ecology and avoid spurious relationships, or conversely could include more species-specific covariates that we did not have available to further reduce the risk of over-fitting and enhance predictive performance. Retaining data for testing or applying cross-validation techniques is another reasonable precaution when conducting covariate selection for predictive models (Yates et al. 2023).

As demonstrated by our results, decisions on the types of covariate complexity to include can have noticeable implications on model predictive performance. While the models proposed and fit in this study *do* use more parameters than most common DOMs, the relationships that are modelled remain comparatively simple relative to species distribution models fit with common tools such as Maxent. The apparent tendency in the DOM literature has been to avoid over-fitting at all costs by erring towards simplicity at each stage of the model building process, often by considering only a restricted pool of covariates and following that with additional complexity penalties during AIC-based model selection (Kelleher et al. 2025). While working to avoid over-fitting is well-intended, one can just as easily end with an *under-fit* model by over-zealously restricting complexity despite sound ecological justification (Soley-Guardia et al. 2024). As seen here and throughout the SDM literature, for complex systems a sizeable number of variables may be necessary to achieve the best available predictive model (Valavi et al. 2023, Soley-Guardia et al. 2024). It should be noted that approaches to covariate selection should vary depending on the study objective (Tredennick et al. 2021), but when developing DOMs for prediction, modellers should keep the ubiquity of complex relationships in ecology in mind and fully consider the possibility of non-linear relationships and interactions in their study systems (Austin 2007,

Catford et al. 2022, Spake et al. 2023). This is particularly important when using large datasets similar to the Swiss Breeding Bird Survey used here. In these situations, the level of ecological realism that could be incorporated without risking over-fitting may well exceed the maximum level attempted following current modelling conventions. As available data and model applications continue to shift, it is necessary to revisit common approaches to model building to ensure we continue to build the most robust, useful, and realistic models possible.

References

- Abrego, N., and O. Ovaskainen. 2023. Evaluating the predictive performance of presence-absence models: Why can the same model appear excellent or poor? *Ecology and Evolution* 13:e10784.
- Aho, K., D. Derryberry, and T. Peterson. 2014. Model selection for ecologists: The worldviews of AIC and BIC. *Ecology* 95:631–636.
- Anderson, D. R., K. P. Burnham, and G. C. White. 1994. AIC Model Selection in Overdispersed Capture-Recapture Data. *Ecology* 75:1780–1793.
- Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200:1–19.
- Austin, M. P. 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling* 157:101–118.
- Austin, M. P., and J. A. Meyers. 1996. Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. *Forest Ecology and Management* 85:95–106.
- Briscoe, N. J., D. Zurell, J. Elith, C. König, G. Fandos, A.-K. Malchow, M. Kéry, H. Schmid, and G. Guillera-Arroita. 2021. Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. *Global Change Biology* 27:4269–4282.

354 Catford, J. A., J. R. U. Wilson, P. Pyšek, P. E. Hulme, and R. P. Duncan. 2022. Addressing
 355 context dependence in ecology. *Trends in Ecology & Evolution* 37:158–170.
 356 Copernicus Land Monitoring Service. (n.d.). CORINE land cover.
 357 Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. *Methods*
 358 in *Ecology and Evolution* 1:330–342.
 359 Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees.
 360 *Journal of Animal Ecology* 77:802–813.
 361 Evans, J. S., and S. A. Cushman. 2009. Gradient modeling of conifer species using random
 362 forests. *Landscape Ecology* 24:673–683.
 363 Evans, M. R., V. Grimm, K. Johst, T. Knuuttila, R. de Langhe, C. M. Lessells, M. Merz, M. A.
 364 O'Malley, S. H. Orzack, M. Weisberg, D. J. Wilkinson, O. Wolkenhauer, and T. G. Benton.
 365 2013. Do simple models lead to generality in ecology? *Trends in Ecology & Evolution* 28:578–
 366 583.
 367 Federal Office of Topography swisstopo. (n.d.). DHM 25 - the digital height model of
 368 switzerland.
 369 Goldstein, B. R., A. G. Keller, K. L. Calhoun, K. J. Barker, F. Montealegre-Mora, M. W. Serota,
 370 A. Van Scoyoc, P. Parker-Shames, C. L. Andreozzi, and P. de Valpine. 2024. How do ecologists
 371 estimate occupancy in practice? *Ecography* n/a:e07402.
 372 Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities
 373 under imperfect detection: advances, challenges and opportunities. *Ecography* 40:281–295.
 374 Joseph, M. B. 2020. Neural hierarchical models of ecological populations. *Ecology Letters*
 375 23:734–747.

376 Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E.
 377 Zimmermann, H. P. Linder, and M. Kessler. 2017. *Climatologies at high resolution for the*
 378 *earth's land surface areas*. *Scientific Data* 4:170122.
 379 Kelleher, S., G. Guillera-Arroita, J. Elith, and N. J. Briscoe. 2025. Twenty years of dynamic
 380 occupancy models: A review of applications and look to the future. *Ecography*.
 381 Kellner, K. F., A. D. Smith, J. A. Royle, M. Kéry, J. L. Belant, and R. B. Chandler. 2023. The
 382 unmarked R package: Twelve years of advances in occurrence and abundance modelling in
 383 ecology. *Methods in Ecology and Evolution* 14:1408–1415.
 384 Kellner, K. F., and R. K. Swihart. 2014. *Accounting for Imperfect Detection in Ecology: A*
 385 *Quantitative Review*. *PLOS ONE* 9:e111436.
 386 Kéry, M. 2011. Towards the modelling of true species distributions. *Journal of Biogeography*
 387 38:617–618.
 388 Kéry, M., G. Guillera-Arroita, and J. J. Lahoz-Monfort. 2013. *Analysing and mapping species*
 389 *range dynamics using occupancy models*. *Journal of Biogeography* 40:1463–1474.
 390 Lonergan, M. 2014. *Data availability constrains model complexity, generality, and utility: A*
 391 *response to evans et al*. *Trends in Ecology & Evolution* 29:301–302.
 392 MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003.
 393 *Estimating site occupancy, colonization, and local extinction when a species is detected*
 394 *imperfectly*. *Ecology* 84:2200–2207.
 395 MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A.
 396 Langtimm. 2002. *Estimating Site Occupancy Rates When Detection Probabilities Are Less Than*
 397 *One*. *Ecology* 83:2248–2255.

398 MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2017.
 399 Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence.
 400 Elsevier Science & Technology, San Diego, UNITED STATES.
 401 Merow, C., M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, W.
 402 Thuiller, R. O. Wüest, N. E. Zimmermann, and J. Elith. 2014. What do we gain from simplicity
 403 versus complexity in species distribution models? *Ecography* 37:1267–1281.
 404 Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. A practical guide to MaxEnt for modeling
 405 species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058–
 406 1069.
 407 Morin, D. J., C. B. Yackulic, J. E. Diffendorfer, D. B. Lesmeister, C. K. Nielsen, J. Reid, and E.
 408 M. Schaubert. 2020. Is your ad hoc model selection strategy affecting your multimodel inference?
 409 *Ecosphere* 11:e02997.
 410 R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for
 411 Statistical Computing, Vienna, Austria.
 412 Rufibach, K. 2010. Use of brier score to assess binary predictions. *Journal of Clinical*
 413 *Epidemiology* 63:938–939.
 414 Santika, T., C. A. McAlpine, D. Lunney, K. A. Wilson, and J. R. Rhodes. 2014. Modelling
 415 species distributional shifts across broad spatial extents by linking dynamic occupancy models
 416 with public-based surveys. *Diversity and Distributions* 20:786–796.
 417 Schmid, H., N. Zbinden, and V. Keller. 2004. Überwachung der Bestandsentwicklung häufiger
 418 Brutvögel in der Schweiz.
 419 Smith, A. B. 2013. The relative influence of temperature, moisture and their interaction on range
 420 limits of mammals over the past century. *Global Ecology and Biogeography* 22:334–343.

421 Soley-Guardia, M., D. F. Alvarado-Serrano, and R. P. Anderson. 2024. Top ten hazards to avoid
 422 when modeling species distributions: a didactic guide of assumptions, problems, and
 423 recommendations. *Ecography* 2024:e06852.

424 Spake, R., D. E. Bowler, C. T. Callaghan, S. A. Blowes, C. P. Doncaster, L. H. Antão, S.
 425 Nakagawa, R. McElreath, and J. M. Chase. 2023. Understanding ‘it depends’ in ecology: a guide
 426 to hypothesising, visualising and interpreting statistical interactions. *Biological Reviews* 98:983–
 427 1002.

428 Strebel, N., S. Wechsler, R. Bühler, G. Häfliger, V. Keller, M. Kéry, C. Rogenmoser, M. Spiess,
 429 K. Varga, B. Volet, N. Zbinden, and H. Schmid. 2026. Data of the swiss common breeding bird
 430 monitoring program. *Ecology* in press.

431 Valavi, R., J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita. 2023. Flexible species
 432 distribution modelling methods perform well on spatially separated testing data. *Global Ecology*
 433 and *Biogeography* 32:369–383.

434 Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. Predictive performance
 435 of presence-only species distribution models: a benchmark study with reproducible code.
 436 *Ecological Monographs* 92:e01486.

437 Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: a meta-
 438 analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos*
 439 123:652–661.

440 Yates, L. A., Z. Aandahl, S. A. Richards, and B. W. Brook. 2023. Cross validation for model
 441 selection: A review with examples from ecology. *Ecological Monographs* 93:e1557.

Table 1: Candidate covariates for all species, with definitions and details on how they vary.

Covariates fell into either landcover, climate, or detection categories, determining which parameters they were considered for and which interactions were considered

Covariate	Description	Variation	Parameters	Group
% Arable (<i>Arbl</i>)	Percentage arable land cover in 3x3km square centred on site.	Site, Year	ψ_1, γ, ϵ	Land cover
% Artificial (<i>Artf</i>)	Percentage artificial surface cover in 3x3km square centred on site.	Site, Year	ψ_1, γ, ϵ	Land cover
% Forest (<i>Frst</i>)	Percentage forest cover in 3x3km square centred on site.	Site, Year	ψ_1, γ, ϵ	Land cover
% Grassland (<i>Gras</i>)	Percentage grassland cover in 3x3km square centred on site.	Site, Year	ψ_1, γ, ϵ	Land cover
% Heterogeneous (<i>Hetr</i>)	Percentage heterogenous land cover in 3x3km square centred on site (e.g., pastures and vineyards).	Site, Year	ψ_1, γ, ϵ	Land cover
Distance to water (<i>Watr</i>)	Distance from centre of site to nearest body of water.	Site, Year	ψ_1, γ, ϵ	Land cover
NDVI ($\mu NDVI$)	Mean NDVI within site, averaged between year t and $t-l$.	Site, Year	ψ_1, γ, ϵ	Climate
Std. deviation of NDVI ($\sigma NDVI$)	Standard deviation of NDVI throughout year t and $t-l$	Site, Year	ψ_1, γ, ϵ	Climate
BIOCLIM 10 ($\mu Temp$)	Mean temperature in warmest quarter within site, average between year t and $t-l$.	Site, Year	ψ_1, γ, ϵ	Climate
BIOCLIM 12 ($\mu Rain$)	Mean annual precipitation within site, averaged between year t and $t-l$.	Site, Year	ψ_1, γ, ϵ	Climate
Elevation (<i>Elev</i>)	Average elevation of site.	Site	p	Detection
Survey date (<i>Date</i>)	Julian date on which survey was conducted.	Survey	ρ	Detection

Table 2: ΔAIC for models fit with each set of complexity restrictions. Models with higher values predict less well and the model with a value of 0 is best-predicting in a set.

Species	LO	LQ	LI	LQI
Crossbill	199.2	54.2	39.9	0.0
Dipper	54.7	54.7	20.3	0.0
Kestrel	78.8	60.7	19.0	0.0
Kite	43.0	33.0	5.2	0.0
Mallard	52.4	47.2	0.0	3.2
Shrike	74.9	37.9	16.3	0.0
Skylark	131.6	49.3	6.9	0.0
Warbler	90.1	57.8	31.3	0.0
Whinchat	62.3	40.3	22.1	0.0
Woodpecker	101.7	59.1	22.6	0.0

Figure captions

Figure 1: The number of terms (regression coefficients) used in the lowest-AIC models for initial occupancy, colonisation, and extinction probabilities for each of ten avian species, under four alternative sets of assumptions regarding the possible complexity of covariate responses. Where used, linear terms are indicated in teal, quadratic terms in purple, and interaction terms in gold.

Figure 2: Inclusion of each covariate in the models for ten focal species fit under each level of permitted complexity. Coloured rings indicate that the regression coefficient for the linear term of a covariate was included in the final model fit at the corresponding level, the regression coefficients for quadratics or interactions are not presented here. For example, the gold ring for the Water in the Crossbill's Ψ_1 column indicates that the LQ model included the term $\beta_{\text{water}} * \text{water}_{i,t}$ in the initial occupancy formula.

Figure 3: Predicted number of sites with detections of each species in each year. Dotted lines indicate the observed number of sites; years to the left of the red line were used for training and years to the right were reserved for testing.

Figure 4: Performance of models for ten species fit under each level of permitted complexity as captured by five metrics of predictive performance.

Figure 5: Difference between testing- and training-set AUC for models fit under four different levels of permitted complexity.

Figure 6: Comparison of predicted occupancy across the full study period using models fit under the Linear-Only (red) and Linear/Quadratic/Interaction (blue) models. Colour intensity reflects the magnitude of the prediction.

Figures

Figure 1

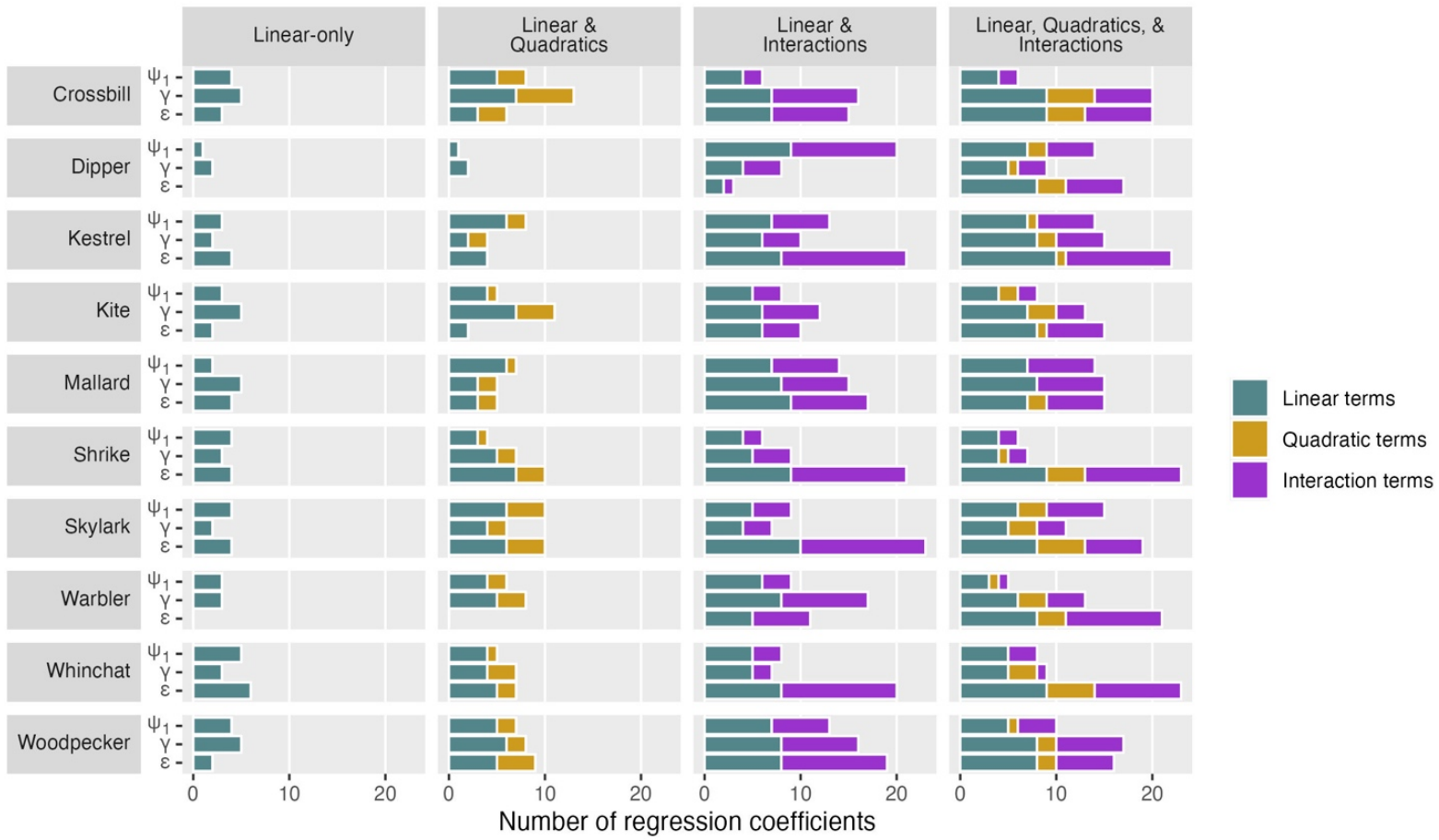


Figure 2

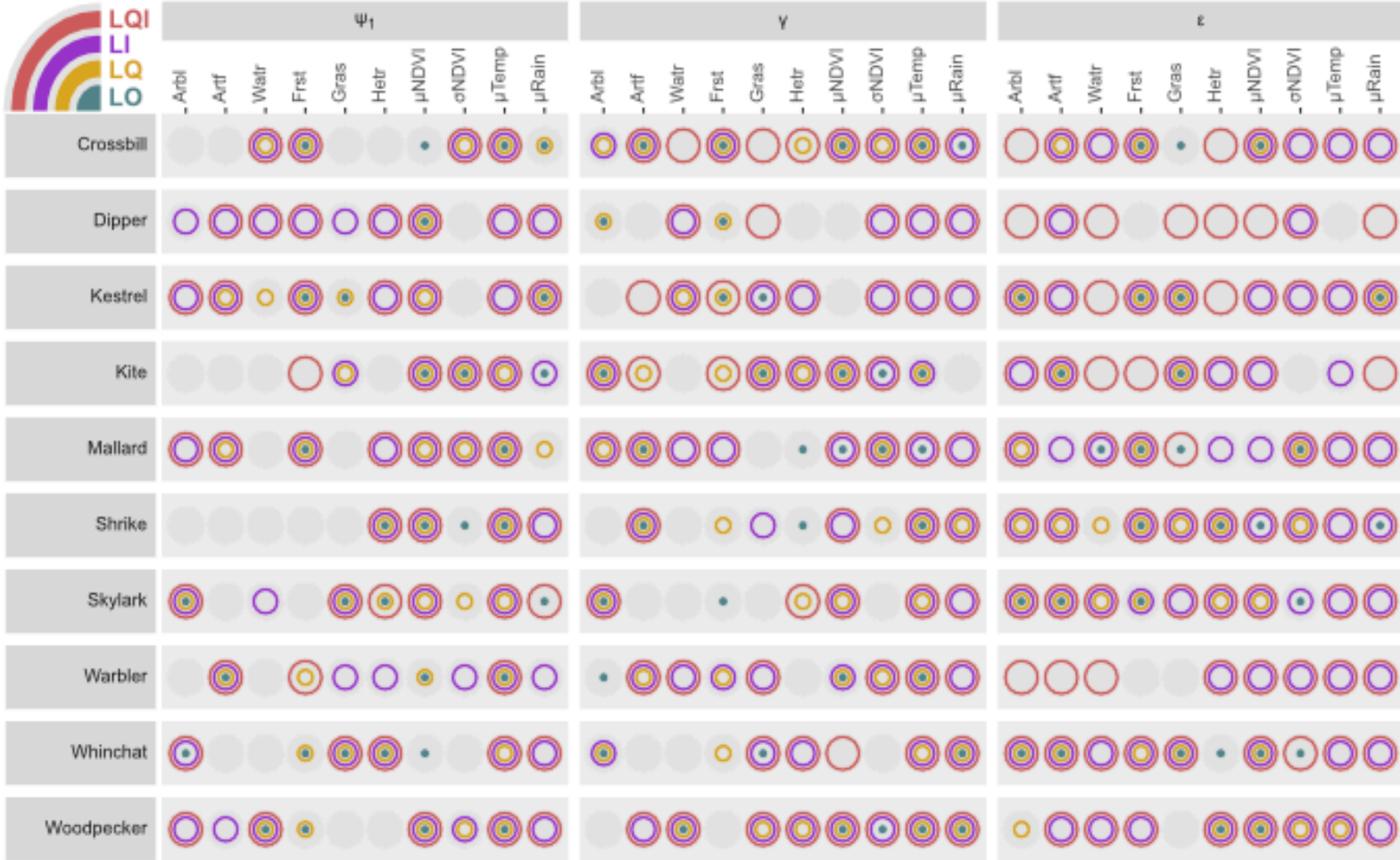


Figure 3

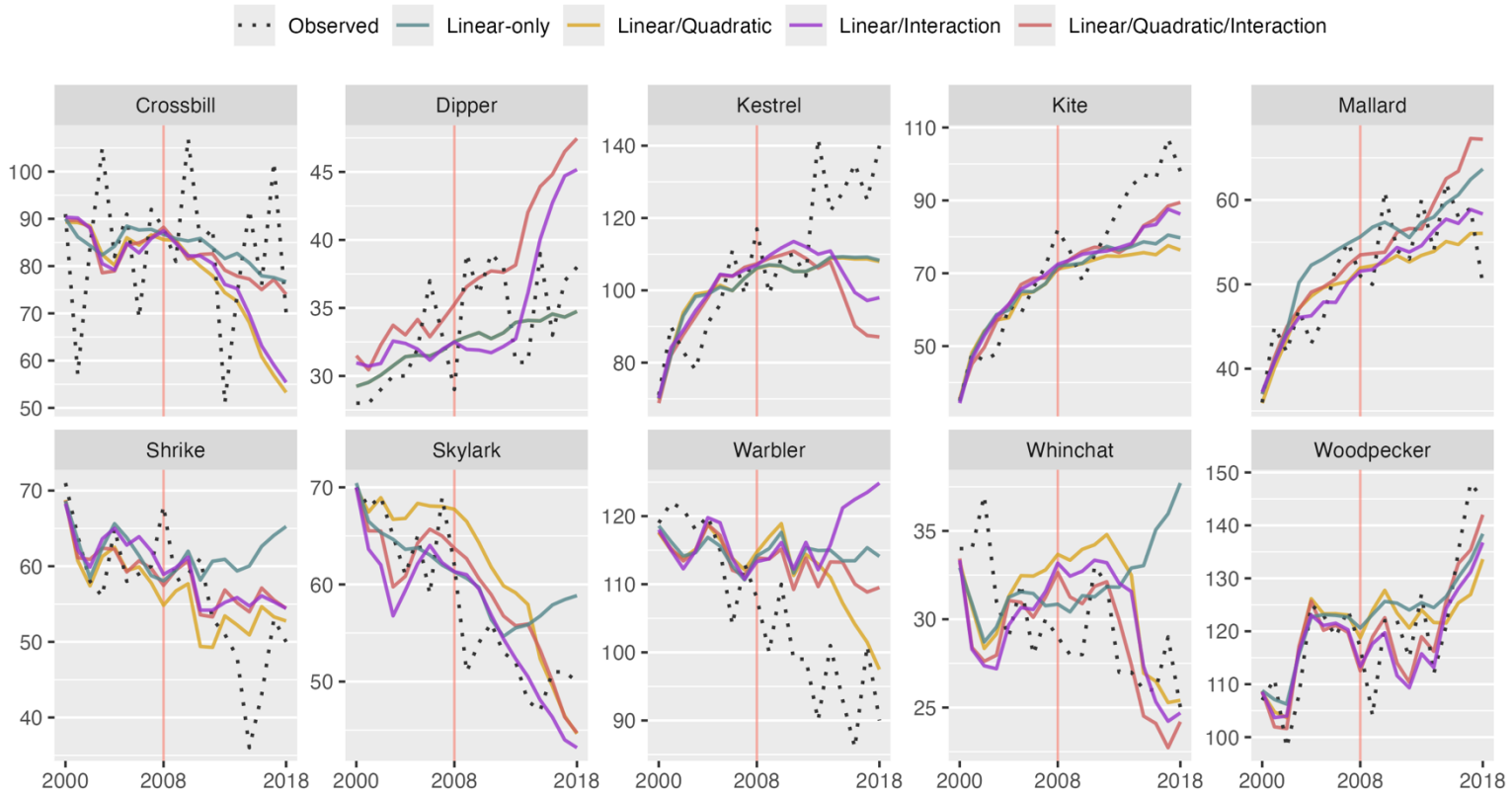


Figure 4

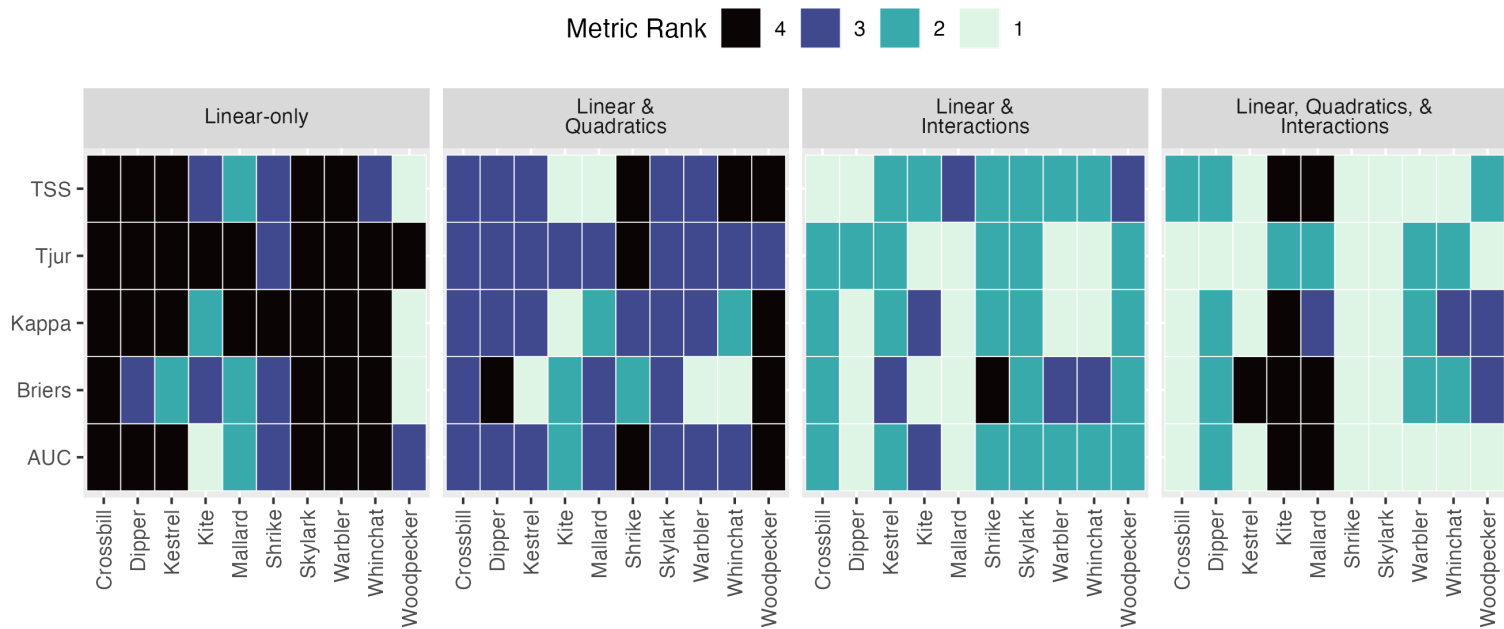


Figure 5

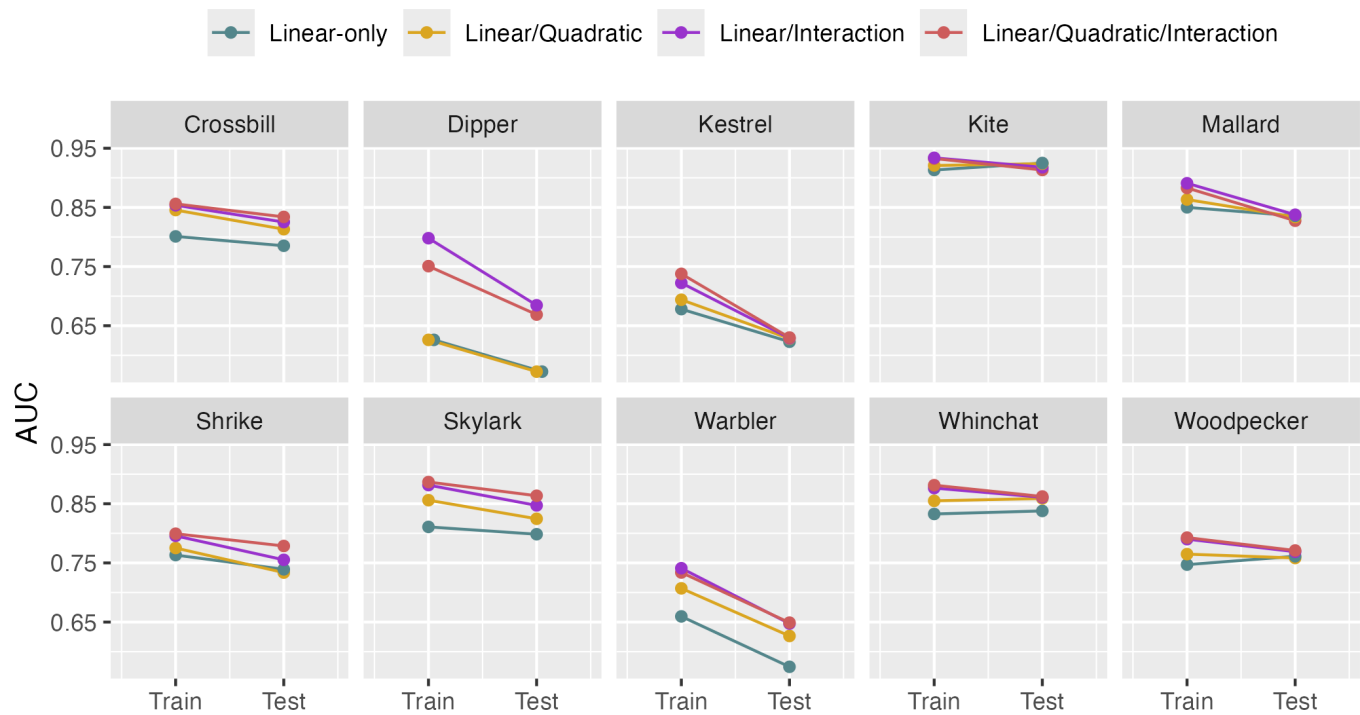


Figure 6

