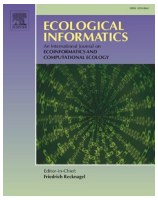
Ecological Informatics 71 (2022) 101803

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Transferable species distribution modelling: Comparative performance of Generalised Functional Response models

Shaykhah Aldossari [a,*], Dirk Husmeier [a], Jason Matthiopoulos [b]

[a] School of Mathematics & Statistics, University of Glasgow, Glasgow G12 8QQ, Scotland, UK
[b] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

ARTICLE INFO

Keywords:
Generalised linear models
Habitat selection
Predictive species distribution models
Radial basis functions
Random forests
Transferability

ABSTRACT

Predictive species distribution models (SDMs) are becoming increasingly important in ecology, in the light of rapid environmental change. However, the predictions of most current SDMs are specific to the habitat composition of the environments in which they were fitted. This may limit SDM predictive power because species may respond differently to a given habitat depending on the availability of all habitats in their environment, a phenomenon known as a functional response in resource selection. The Generalised Functional Response (GFR) framework captures this dependence by formulating the SDM coefficients as functions of habitat availability. The original GFR implementation used global polynomial functions of habitat availability to describe the functional responses. In this study, we develop several refinements of this approach and compare their predictive performance using two simulated and two real datasets. We first use local radial basis functions (RBF), a more flexible approach than global polynomials, to represent the habitat selection coefficients, and balance bias with precision via regularization to prevent overfitting. Second, we use the RBF-GFR and GFR models in combination with the classification and regression tree CART, which has more flexibility and better predictive powers for non-linear modelling. As further extensions, we use random forests (RFs) and extreme gradient boosting (XGBoost), ensemble approaches that consistently lead to variance reduction in generalization error. We find that the different methods are ranked consistently across the datasets for out-of-data prediction. The traditional stationary approach to SDMs and the GFR model consistently perform at the bottom of the ranking (simple SDMs underfit, and polynomial GFRs overfit the data). The best methods in our list provide non-negligible improvements in predictive performance, in some cases taking the out-of-sample R2 from 0.3 up to 0.7 across datasets. At times of rapid environmental change and spatial non-stationarity ignoring the effects of functional responses on SDMs, results in two different types of prediction bias (under-prediction or mis-positioning of distribution hotspots). However, not all functional response models perform equally well. The more volatile polynomial GFR models can generate biases through over-prediction. Our results indicate that there are consistently robust GFR approaches that achieve impressive gains in transferability across very different datasets.

1. Introduction

As the complexity of questions related to conservation and ecosystem management begins to outstrip our ability to collect detailed spatial and temporal data (Fordham et al., 2016; Kindsvater et al., 2018), we have come to rely on more sophisticated statistical methodologies for interpolating between locations, times and taxonomic groups and for predicting into the future. Predictive models of species distributions, in particular, play an increasingly important role as organisms respond to accelerating changes in climate and land use (Evans et al., 2012; Houlahan et al., 2017; Maris et al., 2018; Mouquet et al., 2015; Sequeira et al., 2018; Travers et al., 2019; Yates et al., 2018).

The demand for transferable models (i.e. models that can predict accurately in environments very different to those used for model fitting - Yates et al., 2018) has led to the realisation that statistical species distribution models (SDMs) are currently not fit-for-purpose, particularly in the case of animal species (Austin, 2002; Bahn and McGill, 2013; Barbet-Massin et al., 2018; Márcia Barbosa et al., 2009; Dormann, 2007; Ehrlén and Morris, 2015; Randin et al., 2006; Tessarolo et al., 2021; Torres et al., 2015; Zurell et al., 2009). A key challenge with the

* Corresponding author.
E-mail address: s.aldossari.1@research.gla.ac.uk (S. Aldossari).

https://doi.org/10.1016/j.ecoinf.2022.101803
Received 9 June 2022; Received in revised form 29 August 2022; Accepted 30 August 2022
Available online 8 September 2022

transferability of SDMs is that species, particularly animals, respond differently to a particular habitat depending on the availability of other habitats in their environment (Boyce and McDonald, 1999; Mysterud and Ims, 1999). Fig. 1 in Matthiopoulos et al. (2011) illustrated this problem when a generalized linear model was applied using a simple animal in a particular environment. The model fits well in the same environment, but provides poor predictions of habitat use when placed in a different environment. This process, termed a functional response in habitat selection, is the result of complex mechanistic interactions between habitat availability and animal behaviour (Matthiopoulos et al., 2011; Mauritzen et al., 2003; Mysterud and Ims, 1998). This is difficult to capture with standard statistical models because the estimated parameters of SDMs are specific to the environmental settings where these models were fitted. The consequence of functional responses is that unless the environmental context is explicitly taken into account, spatial predictions can be increasingly inaccurate as the prediction settings diverge from the model fitting environmental profiles (Paton and Matthiopoulos, 2016). Functional responses in habitat selection are detectable in real datasets. Fig. 6 in Bjørneraas et al. (2012) shows how moose in Norway used nine different habitat types based on relative habitat availability. Two species of birds show positive functional responses to three treatments of habitat (see Fig. 2 and Table 3 in Gillies and St. Clair (2010)). Functional responses to pastures were detected in a telemetry dataset containing 62 red deer in Norway (Godvik et al., 2009). When pasture was rare, the selection of pastures was increased, but the selection of pasture decreased with increasing pastures availability (Godvik et al., 2009). Functional responses are detectable by many different methods, but the exact nature of the response depends on the statistical methodology that is used to capture it. For example, two species of conservation concern, Canada lynx in the United States and woodland caribou in Canada, were used to evaluate four different functional response approaches (Holbrook et al., 2019). Habitat use in the additive scale, habitat use model in the multiplicative scale, habitat selection with resource selection function RSF, and habitat selection with the interaction of RSF. There was a variation among these approaches with regard to evaluating the functional response (see Fig. 3 in Holbrook et al., 2019). Some approaches show increases in habitat use by Canada lynx with increasing advanced regenerating forest availability while other approaches show the opposite effect. Some approaches demonstrated no functional response. The same variation in results occurred when testing woodland caribou habitat use in response to linear features. Differences resulting from different implementations highlight the importance of investigating the robustness of functional response models, our objective in this paper.

Different approaches have been proposed to model functional responses in habitat selection, ranging from single-habitat models of usage as a function of availability (Mysterud and Ims, 1998) to writing SDM coefficients as functions of the availability of all habitats (Boyce and McDonald, 1999). The need to account for functional responses is clearly demonstrated by the efficacy of approaches that do not use any model of functional response but simply recognise the distinction between different environmental scenarios by means of random effects (Gillies et al., 2006). The generalized functional response GFR approach (Matthiopoulos et al., 2011) uses a function of availability to represent the SDM's coefficients. The coefficients of the GFR are modeled by functions of local habitat availability using a polynomial function approach (Matthiopoulos et al., 2011; Matthiopoulos et al., 2019). The GFR model is ultimately structured using the local value of the habitat covariates, moments (e.g., the means) from the distribution of the habitat covariates, and the pairwise interactions between these terms (Matthiopoulos et al., 2011). The GFR is an example of a varying-coefficient model, an extension of generalized regression models with coefficients written as functions of other variables (Hastie and Tibshirani, 1993).

The approach taken in Matthiopoulos et al. (2011) was to model changes in each of the SDM coefficients via a global polynomial, motivated by the fact that under fairly general regularity conditions, any smooth function can be approximated by a Taylor series. The practical problem, however, is that this power series expansion with its polynomial coefficients has to be learned from data. Taking a high polynomial order leads - for limited and noisy data - to potential over-fitting (and poor transferability). Standard approaches therefore aim to find the adequate degree of model complexity, e.g. via cross-validation or based on information criteria, such that for small data sets and high noise levels, less complex models are preferred.

However, for a global polynomial function, controlling model complexity e.g. by restricting the number of adjustable model parameters, implies a truncation of the polynomial order and a limitation of the degree of non-trivial differentiability. This is methodologically inconsistent: the highest polynomial order and the degree of non-trivial differentiability are an intrinsic feature of the systems under investigation and must not be dictated by the quantity and quality of the available data. The aim of the present paper is to build on the GFR approach proposed by Matthiopoulos et al. (2011), by replacing the global polynomial expansion by several more recent methods from multivariate statistics and machine learning such that the logical inconsistency outlined above is avoided.

To address the limitations of the original GFR model proposed in Matthiopoulos et al. (2011), we adapt three state-of-the-art flexible regression paradigms to model the habitat selection coefficients. The first approach is based on a radial basis function (RBF) expansion, as e.g. reviewed in Chapter 5 of Bishop (1995), and we refer to this model as the RBF-GFR model. Next, we combine classification and regression trees (CART), reviewed e.g. in Chapter 9 of Hastie et al. (2008) or Section 16.2 in Murphy (2012), with both the original GFR model and our RBF-GFR model. We refer to these models as GFR-CART and RBF-GFR-CART, respectively. We finally create model ensembles, based on random forests (RFs) trained with bagging (see Chapter 15 in Hastie et al., 2008) or boosting (see e.g. Chapter 16 in Hastie et al., 2008 or Section 16.4 in Murphy, 2012). We refer to these ensembles with the suffix "RF" or "XGBoost". An overview of our methods can be found in Fig. 1. The

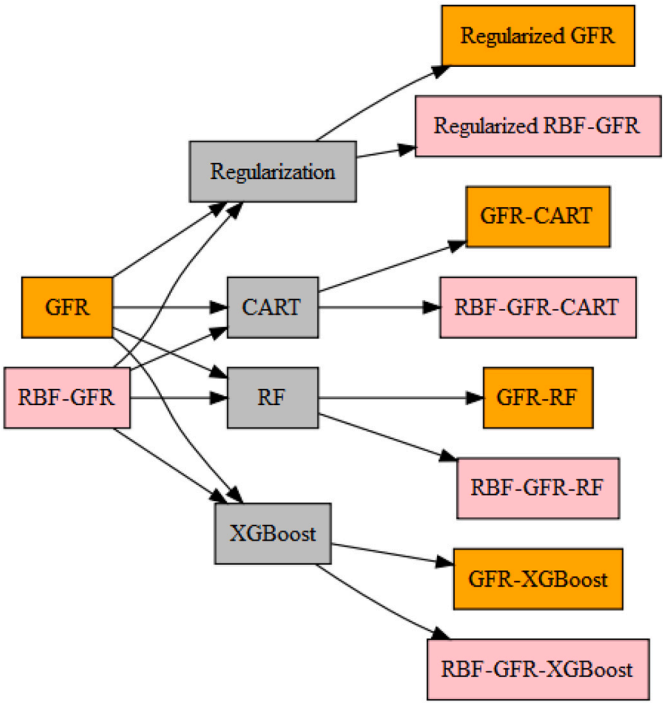


Fig. 1. A diagram showing the relationship between all models proposed in this study. The orange boxes refer to the GFR model and the extensions of the GFR model while the pink boxes are the RBF-GFR model and its extensions. The gray boxes are the methods that were used to combine to the GFR and RBF-GFR models.

current work is based on preliminary work published in two conference proceedings (Aldossari et al., 2020; Aldossari et al., 2021) with substantial methodological expansion and new applications.

Our article is structured as follows: We review the original GFR model from Matthiopoulos et al. (2011) in Section 2 and describe our new methods in Section 3. Details of study design, investigations and performance evaluation can be found in Section 4. Section 5 provides an overview of two simulated and two real-world data sets used for a comparative evaluation of the various methods. The results of this evaluation are presented in Section 6. We finish with our general dissection and conclusions in Section 7. To keep the main text sufficiently concise, we have relegated some methodological details along with more comprehensive simulation results to the online Supplementary material.

2. Background

To explain GFRs more precisely, we need to introduce this class of models mathematically, starting from the basic form of modern SDMs. Both of the dominant methods for modelling species distributions Maximum Entropy (Phillips et al., 2006) and Resource Selection Functions (Boyce and McDonald, 1999) use the exponential formulation of a predictor function of environmental covariates $\mathbf{x} = (x_1, \ldots, x_I)$ to describe the association between a species and its environment:

$$h\left(\mathbf{x}\right) = \exp\left(\sum_{i=1}^{I} \beta_i x_i\right) \tag{1}$$

where $h(\mathbf{x})$ denotes habitat preference and the coefficients β_i are assumed fixed. Although the summand shown above is the prototypical expression for a linear predictor, the approach may be augmented with customary extensions such as higher-order polynomial terms, interactions, or generalised additive terms. The coefficients of the linear predictor are either estimated by likelihood (e.g., RSFs) or entropy criteria (e.g., MaxEnt), and they are fitted to different types of data (e.g., telemetry, survey etc.) by different link functions. Modern inference has unified all the existing approaches to different types of data under the framework of Inhomogeneous Point Processes (IPP), and has therefore tended to interpret the quantity in Eq. (1) as the intensity function of the IPP (reviewed in Matthiopoulos et al., 2020a).

Habitat selection models formalized in this way are known to have low transferability to new environmental scenarios (Boyce and McDonald, 1999; Mysterud and Ims, 1999). The predictions from Eq. 1 rely on the assumption that organisms use habitats in proportion to their preference and that preference does not change when habitat availability changes. This assumption is born of statistical convenience, not biological reality. In practice, any changes in habitat availability lead to disproportionate changes in a species' response, a phenomenon known as a functional response in resource selection (Mysterud and Ims, 1998).

Boyce and McDonald (1999) argued that functional responses could be captured by relaxing the stationarity of the fixed coefficients $\beta_i \in \mathbb{R}$. They proposed in particular that multiple sampling instances (e.g., different home ranges from multiple individuals) are examined, to see how the coefficients of the SDM shift in response to changes in habitat availability from one sampling instance to the other. Hence, in their original GFR model, Matthiopoulos et al. (2011) pooled data from multiple sampling instances and allowed the β_i to vary as flexible functions $\gamma_i(\mathbf{x})$ weighted by habitat availability $f_b(\mathbf{x})$. In the case of continuous environmental space:

$$\beta_{i,b} = \int \gamma_i\left(\mathbf{x}\right) f_b\left(\mathbf{x}\right) d\mathbf{x}, \tag{2}$$

where $f_b(\mathbf{x})$ is a probability density function describing the availability of habitat $\mathbf{x}$ in the b^{th} sampling instance. A sampling instance represents an environmental scenario defined in a biological way as the environment experienced by the study animals during an appropriate spatio-temporal frame of accessibility (Matthiopoulos et al., 2020b). For example, a sampling instance could represent the spatial domain of a well-mixing subpopulation during a given year. A sampling instance could represent different years for the same population or different subpopulations in the same year (i.e., a space-for-time substitution in sampling effort is possible). An approximate, discretised version of this formulation, uses summation:

$$\beta_{i,b} = \sum_{n=1}^{N} \gamma_i\left(\mathbf{x}_n\right) f_b\left(\mathbf{x}_n\right), \tag{3}$$

where n encodes for a specific habitat. Intuitively, the function $\gamma_i(\mathbf{x})$ describes the change in the SDM's slope for the i^{th} covariate (i.e., $\beta_{i,b}$), generated by a unitary increase in the availability of the n^{th} habitat type.

Matthiopoulos et al. (2011) used a polynomial function to formulate the $\gamma_i(\mathbf{x})$ for each environmental variable $(\mathbf{x})$:

$$\gamma_i\left(\mathbf{x}\right) = \sum_{j=1}^{I} \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} x_j^m \tag{4}$$

where the coefficient of $\gamma_i(\mathbf{x})$ for the m^{th} power of the j^{th} variable is $\delta_{i,j}^{(m)}$. This derivation leads to the following expression for the β's:

$$\beta_{i,b} = \gamma_{i,0} + \sum_{j=1}^{I} \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} E\left[X_j^m\right]_b \tag{5}$$

where M_j is an integer order parameter and $E[X_j^m]_b$ is the mth moment of the covariate j calculated for the conditions prevailing in the bth sampling instance. Furthermore, $\gamma_{i,0}$ is an intercept corresponding to the scenario of zero expectations. If at least the first two moments of X are zero (corresponding to zero mean and variance for X), this implies that the environmental variable has its baseline value, uniformly across accessible space.

Let z denote a vector composed of all elements $\{x_j^m\}$ and $\{E[X_j^m]_b\}$. Using the polynomial function approach described above, habitat preference $h(z)$ can be expressed as a function of fixed effects of covariates and pairwise interactions between covariates and their moments:

$$h\left(z;\boldsymbol{\theta}\right) = \exp\left\{ \gamma_{0,0} + \sum_{i=1}^{I} \left(\sum_{m=0}^{M_i} \delta_{0,i}^{(m)} E\left[X_i^m\right]_b + \gamma_{i,0} x_i + x_i \sum_{j=1}^{I} \times \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} E\left[X_j^m\right]_b \right) \right\} \tag{6}$$

where $\boldsymbol{\theta}$ is a parameter vector composed of the parameters $\gamma_{i,0}$ and $\delta_{i,j}^{(m)}$. The vector z combines habitat variables x_i and their expectations $E[X_i^m]$. Although the GFR model has been shown to achieve better predictive performance than the conventional GLM model of Eq. (1), it suffers from various limitations. The degree of nonlinear complexity and smoothness is restricted in advance: the functions in Eq. (6) are only M_j times differentiable, where the $M_j's$ are the highest polynomial orders. A complex function with a high degree of differentiability thus requires a large number of parameters, which renders the approach susceptible to over-fitting. Restricting the maximum polynomial order commensurately with the training set size leads to the paradox situation that the functional complexity of the habitat preference coefficients, which is an inherent property of the species and the habitat under investigation, becomes contingent on the arbitrariness of the data acquisition process. Moreover, while the degree of smoothness and model complexity is allowed to vary with respect to the choice of environmental variable, it is assumed to apply globally to the entire input domain.

3. Methods

The shortcomings described in the previous section are well-known in the statistics and machine learning communities, and various flexible regression methods have been developed to address them (see e.g., Hastie et al., 2008). In the following section, we review three such solutions, based on basis function expansions, regression trees and model ensembles. We then describe how we have adapted them to improve the transferability of SDMs.

3.1. A Radial Basis Function (RBF) model

A radial basis function (RBF) expansion is a widely used flexible regression approach that disentangles the number of adaptable parameters (defined by the number of basis functions) from the degree of differentiability (determined by the functional form of the basis functions). For a review, see e.g. Chapter 5 in Bishop (1995).

Instead of the polynomial function that was used in the original GFR model, we use a basis function expansion to model the generalised functional response coefficients $\gamma_i(\mathbf{x})$:

$$\gamma_i\left(\mathbf{x}\right) = \sum_j \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} \phi\left(x_j, \boldsymbol{\theta}_{j,m}\right) = \sum_j \sum_m \delta_{i,j}^{(m)} \phi\left(x_j, \boldsymbol{\theta}_{j,m}\right) \tag{7}$$

where ϕ is a basis function (e.g. splines, wavelets, basis functions of a reproducing kernel Hilbert space, etc.) with parameters $\boldsymbol{\theta}_{j,m}$, chosen to represent known functional characteristics, and the sum over m going from 0 to M_j.

We choose the following standard radial basis function (RBF) for $\gamma_i(\mathbf{x})$ (see Bishop, 1995):

$$\gamma_i\left(\mathbf{x}\right) = \sum_j \sum_m \delta_{i,j}^{(m)} \exp\left(-\frac{1}{2} \frac{\left(x_j - \xi_{j,m}\right)^2}{\sigma_{j,m}^2} \right) \tag{8}$$

where $\xi_{j,m}$ is the center of the mth basis function for the jth covariate and $\sigma_{j,m}$ is its bandwidth parameter. We follow Matthiopoulos et al. (2015) and model the probability distribution $f(\mathbf{x})$ (i.e., the habitat availability characterising a sampling instance) with a Gaussian mixture model:

$$f\left(\mathbf{x}\right) = \sum_{k=1}^{K} w_k N\left(\mathbf{x} | \boldsymbol{\mu}_k, \mathbf{C}_k\right) \tag{9}$$

where K is the number of mixture components, w_k is the mixing weight of the kth component, $\boldsymbol{\mu}_k$ defines its centre and $\mathbf{C}_k$ is the covariance matrix. We assume that it is implied that f, w_k, $\boldsymbol{\mu}_k$ and $\mathbf{C}_k$ are all specific to a sampling instance, and therefore that the subscript b is omitted for all these quantities, to simplify notation.

These modelling choices for $\gamma_i(\mathbf{x})$ and f lead to the following expression for the habitat selection coefficients of the SDM (see Supplement A.1 for a derivation):

$$\beta_i = \gamma_{i,0} + \sum_j \sum_m \delta_{i,j}^{(m)} I_{j,m} \tag{10}$$

A comparison with Eq. (6) shows that the new RBF-GFR model replaces $E\left[\mathbf{x}_j^m\right]_b$ from the original GFR model in Matthiopoulos et al. (2011) by $I_{j,m}$:

$$I_{j,m} = \sum_k w_k \frac{\sigma_{j,m}}{\sqrt{\left(\sigma_{j,j}^2 + \sigma_{j,m}^2\right)}} \cdot \exp\left[-\frac{1}{2} \left(\frac{\left(\mu_{j,k} - \xi_{j,m}\right)^2}{\sigma_{j,j}^2 + \sigma_{j,m}^2} \right) \right] \tag{11}$$

Defining z sightly differently from before in order to account for that, namely composed of all elements in $\{x_i\}$ and $\{I_{j,m}\}$, and inserting Eq. (10) into (1), we get the following model of habitat preference:

$$h\left(z;\boldsymbol{\theta}\right) = \exp\left\{ \gamma_{0,0} + \sum_{j=1}^{I} \sum_{m=0}^{M_j} \delta_{0,j}^{(m)} I_{j,m} + \sum_{i=1}^{I} \left(\gamma_{i,0} + \sum_{j=1}^{I} \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} I_{j,m} \right) x_i \right\} \tag{12}$$

The parameters of the RBFs, $\xi_{j,m}$ and $\sigma_{j,m}$, need to be determined in advance to find $I_{j,m}$ in Eq. (11). Various methods have been suggested in the literature; see e.g. Bishop (1995). We set the centres of the basis functions $\xi_{j,m}$'s of the jth environmental covariate to be the quantile of the jth environmental covariate. The bandwidth parameter $\sigma_{j,m}$ is the larger of the differences between the m-quantile and $(m-1)$-quantile and the difference between the $(m+1)$-quantile and m-quantile. For example, if the number of basis functions is 3, then the first quantile, median, and the third quantile are the centres of the basis functions. Specifically, $\sigma_{j,2}$ is the larger of the difference between the first and second quantiles and the difference between the second and third quantiles. We note that the vector z, which characterizes the habitat, is usually given a separate subscript, z_k, where k denotes a particular plot or geographical patch where species counts are taken. We have avoided this subscript in the present section to avoid clutter in the notation, but will make use of it in subsequent sections.

3.2. Calibration and regularization

The original GFR and RBF-GFR models are types of GLMs whose parameters can be estimated via maximum likelihood (ML) (Hastie et al., 2016). In the GFR and RBF-GFR models, the ML approach aims to maximize the likelihood function $L(\boldsymbol{\theta}, z, \boldsymbol{y})$ where $\boldsymbol{\theta}$ is a parameter vector composed of the parameters $\gamma_{i,0}$ and $\delta_{i,j}^{(m)}$ in Eq. (6) for the GFR model and Eq. (12) for the RBF-GFR model. Note that z is a vector combining habitat variables x_i and either their expectation values, $E\left[X_i^m\right]$, or the derived quantities $I_{i,m}$ defined in Eq. (11). The variables z_i are readily available from the observed data. On fitting a Gaussian mixture model to the explanatory data, we obtain the quantities $I_{i,m}$ from Eq. (11). The vector $\boldsymbol{y} = (y_1, \ldots, y_N)$ contains species observations, where N denotes the number of patches where species counts are taken. Depending on the study, the elements of this vector, y_n, can be binary use/availability indicators or count data. The equivalence between grid count, use-availability and point-process data has been demonstrated in the literature, on the basis of their corresponding likelihood functions (Aarts et al., 2012; Renner and Warton, 2013; Warton and Aarts, 2013; Warton and Shepherd, 2010). Although some of the datasets we used in this study took the form of binary (0/1) values, they were nevertheless equivalent to abundance models. Our analyses stayed firmly in the area of abundance rather than occupancy models. Occupancy models (the recording of the presence of a species, regardless of its abundance) also results in binary data, but involve loss of information on abundance and although they are a widely used type of analysis (MacKenzie et al., 2017), they pose additional analytical challenges that were outside the remit of this paper.

To optimize the vector parameter $\boldsymbol{\theta}$ in Eq. (6) for the original GFR model and Eq. (12) for the RBF-GFR model using the maximum likelihood estimator, we have to find the values of the model parameters that maximize the likelihood function $L(\boldsymbol{\theta}, z, \boldsymbol{y})$.

If the response variable for a data set is a binary species use/availability indicator, that is $y_n \in \{0, 1\}$, the Bernoulli model is used in each site with probability of use $p = Pr(\boldsymbol{y} = 1 | z; \boldsymbol{\theta})$ and probability of availability $1 - p = Pr(\boldsymbol{y} = 0 | z; \boldsymbol{\theta})$ where:

$$\begin{aligned} p = Pr(y_n = 1 | \boldsymbol{z}_n; \boldsymbol{\theta}) &= \frac{1}{1 + e^{-h(\boldsymbol{z}_n; \boldsymbol{\theta})}}; \\ Pr(y_n = 0 | \boldsymbol{z}_n; \boldsymbol{\theta}) &= 1 - p \end{aligned} \tag{13}$$

where the subscript n denotes a geographical patch or plot where species counts are taken. The log likelihood is:

$$l\left(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{y}\right) = \frac{1}{N}\sum_{n=1}^{N}\log Pr\left(y_n|\boldsymbol{z}_n;\boldsymbol{\theta}\right) \tag{14}$$

For species abundance levels, the Poisson distribution is used with mean parameter equals to $h(\boldsymbol{z};\boldsymbol{\theta})$ in Eq. (6) for the GFR model and Eq. (12) for the RBF-GFR model. If y_n is the number of species in cell n, then the log likelihood is:

$$l\left(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{y}\right) = \sum_{n=1}^{N}\log\left(\frac{h(\boldsymbol{z}_n;\boldsymbol{\theta})^{y_n}e^{-h(\boldsymbol{z}_n;\boldsymbol{\theta})}}{y_n!}\right) \tag{15}$$

Further note that for GLM-type models the ML equations have no closed-form solution, and an iterative optimization algorithm is therefore applied; see e.g. Section 4.3.3 in Bishop (2006) for details.

Parameter estimation with ML can be susceptible to over-fitting, especially for sparse and noisy data. This can be addressed with regularization, where a penalty term that quantifies model complexity is added to the log likelihood; see e.g. Section 3.2 in Bishop (2006). A particular form of regularization is ridge regression, where the model coefficients are shrunk towards smaller values by altering the ML criterion to include a penalty term based on the weighted L2 norm of the parameter vector:

$$l(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{y}) - \lambda\|\boldsymbol{\theta}\|^2 \tag{16}$$

The weighting factor λ is a regularization parameter (see e.g. Platt, 1999) which needs to be optimized. In the present work, we repeated the iterative optimization of the objective function in Eq. (16) for 100 discrete candidate values of λ chosen from a vector of equidistant values (Hastie et al., 2016), and then selected the value that minimizes the Bayesian information criterion (BIC) (Schwarz, 1978) as an established model selection score.

3.3. Classification and Regression Trees (CART)

To increase the flexibility of the previously discussed models, we integrate them into a classification and regression tree (CART). CARTs, originally proposed by Breiman et al. (1984), are based on a divide-and-conquer strategy, whereby the data space is divided into subsets defined by the decision rules of a bifurcating tree. The inner nodes of the tree, also called decision nodes, compare the values of selected input variables against certain decision thresholds. This confers a cybernetics-like interpretable logical structure to the model, which aids interpretability and explainability. All the decision rules, including the selection of variables and decision thresholds, are systematically learned from the data. For the corresponding statistical inference methods and regularization techniques (to prevent overfitting), we refer the reader to the statistical literature (e.g. Breiman et al., 1984 and Hastie et al., 2008). We also include a more detailed summary in Supplement A.2.

3.4. Random forests, bagging and boosting

We can combine classification and regressions trees into a model ensemble called a "random forest". This is based on the insight that the expected out-of-sample prediction error can be decomposed into a bias and a variance component (see e.g. Section 3.2 in Bishop (2006)), that for a flexible model, like CART, the main contribution to this error comes from the variance term, and that this variance term can be reduced in a model ensemble, provided the models are sufficiently uncorrelated (see e.g. Section 14.3 in Bishop, 2006).

To reduce the correlations between the individual members of the ensemble, we follow the "bootstrapping and aggregating" procedure proposed by Breiman (2001), also called "bagging", whereby CART models are repeatedly trained on different independent bootstrap replicates, and then aggregated in a model ensemble, via voting (for classification) or averaging (for regression). To further decrease the correlation between the individual CART models, the split rules at the inner nodes of the trees are limited to randomly selected subsets of the features as candidate sets. For further details, see Breiman (2001) and Chapter 15 in Hastie et al. (2008). In our work, we propose a new variant of random forests, where each leaf node of the trees in the ensemble is a GFR or GFR-RBF model.

An alternative to bagging is boosting, where the models in the ensemble are trained sequentially, using a weighted form of the data, where the weights depend on the previous model such that misclassified or poorly predicted instances get greater weights. For further details, see e.g. Section 14.3 in Bishop (2006) or Chapter 10 in Hastie et al. (2008). The boosting variant we use in our work is extreme gradient boosting (XGBoost), proposed by Chen and Guestrin (2016).

While the ensemble size for bagging is not particularly critical, provided it is sufficiently large (500 is a widely used default value), it does matter for boosting. In XGBoost, large ensemble sizes can cause over-fitting because the gradient technique focuses on the most difficult cases, which can be due to noise. To avoid the over-fitting issue in XGBoost, we use a nested k-fold cross-validation scheme. We split each dataset into 3 subsets: the tuning set (k-2 folds), validation set (1-fold), and test set (1-fold). For each choice of number of iterations and each fold, we train the model on the tuning set and monitor the performance on the validation set by calculating the out-of-sample prediction accuracy and taking the median of k-1 folds. This gives us k medians for each number of iterations, as explained in Algorithm 1 and shown in Fig. S2 in Supplement A.3.

Algorithm 1 Optimize the iteration number

```
procedure SPLIT THE DATASET TO K FOLDS (state)
    for each k-1 folds do
        for each number of iterations do
            Split the dataset to datasets: k-2 folds (tuning set), and 1-fold (validation dataset)
            Train the model using the tuning set and the number of iterations.
            Predict using the 1-fold validation set.
            Calculate the out-of-sample R^2.
        Calculate the median of k-1 out-of-sample R^2's for each number of iterations to pick the best number of iterations.
    return a matrix of medians for each fold (k folds) in rows and each number of iterations in columns.
```

4. Simulations and evaluation

4.1. Study design and investigations

In the first part of the comparative model assessment study, the original GFR model, reviewed in Section 2, was compared with the proposed RBF-GFR model described in Section 3.1. Both models depend on different complexity parameters. For the GFR model, it is necessary to define the polynomial order, as seen from Eq. (5). For the RBF-GFR model, we need to decide on the number of Gaussian mixture components, see Eq. (9), and the number of RBF basis functions, as seen from Eq. (11). We repeated the iterative optimization of the objective function from Eq. (12) for different choices of these complexity parameters, and then picked the one that minimized model selection score (BIC). For the number of Gaussian mixture components, we found the number of components that minimize the BIC score for each block, then used the average of the number of components of all blocks as the optimal number of Gaussian mixture components for the RBF-GFR model and its extensions.

For the best parameters thus selected, we compared the test set accuracies, quantified in terms of out-of-sample R-square scores, for the original GFR and the proposed RBF-GFR model.

CART and RF models were combined with the original GFR and RBF-GFR models, and all models were then compared against the original GFR and RBF-GFR models, as shown in Fig. 1. A standard CART algorithm, where each leaf is a separate GFR model or RBF-GFR model, was applied in each case, and the cost function in Eq. (27) in Supplement A.2

was used to grow the tree and find the best split variable for each iteration of the optimization algorithm. The tree was then pruned using 10-fold cross-validation based on the training set. The habitat usage of the test set was then predicted to measure the out-of-sample prediction scores. For the RF model, the number of trees had to be selected, as described in Section 3.4. In this case, a baseline of 500 trees was set, where each leaf in each tree was a separate original GFR model or RBF-GFR model.

The XGBoost model was used in combination with the original GFR and RBF-GFR models over several different numbers of iterations {2, 5, 10, 15, 20, 40, 80, 100, 200, 300, 400, 500}; Algorithm 1 was then used to determine the best number of iterations of XGBoost for use in all subsequent applications.

A comprehensive comparative evaluation of the methods discussed in this paper was carried out, with results as shown in Fig. 1 and Table S1 in Supplement A.4.

To summarise the comparative model assessments study, the models were then ranked by performance for each of the four data sets in turn, generating a "league table" of models. This table thus offers the ranks of all the models included in the study, as shown in Fig. 1. Based on the need to assess the quality of predictions, the predictions of animal habitat usage derived from the models used were presented in spatial maps and further visualisations of species abundance were generated.

4.2. Performance evaluation

To measure the out-of-sample performance (and hence, the transferability) of different models, we used the standard out-of-sample R^2 score, obtained by splitting the data set into two parts, for training and testing. The metric is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{17}$$

where y_i are the observations in the test set, $\bar{y}$ is their mean, and $\widehat{y_i}$ are the model predictions. We used k-fold cross-validation to calculate the R^2 score in this way, and summarised the performance by calculating the median out-of-sample R^2 score, obtained from all folds, as a more robust alternative to the mean (Leys et al., 2013). For species abundance counts, the following modified score based on the residual deviance is preferred, as discussed in Colin Cameron and Windmeijer (1996):

$$R^2_{DEV} = 1 - \frac{\sum_{i=1}^{n} \left\{ y_i log\left(y_i/\widehat{y_i}\right) - \left(y_i - \widehat{y_i}\right) \right\}}{\sum_{i=1}^{n} y_i log\left(y_i/\bar{y}\right)}, \tag{18}$$

In addition, the median-absolute deviation (MAD) of the out-of-sample R^2 scores was used to measure the variability of the R^2's obtained from each fold. MAD is more robust to outliers than the more widely used standard deviation. To aid interpretability, we follow the standard approach and apply a scale factor of $c = 1.4826$ to make the MAD identical to the standard deviation for normally distributed data (Leys et al., 2013):

$$MAD = median\left(\left|R^2 - \widetilde{R^2}\right|\right) \times c$$

where $\widetilde{R^2}$ is the median of out-of-sample R^2 scores.

Finally, we used the pseudo R^2 score to measure the proportion of the total variability explained by the model as follows:

$$R^2_P = 1 - \frac{Residual\ Deviance}{Null\ Deviance} = 1 - \frac{2l(\boldsymbol{\theta}_c, \boldsymbol{z}, \boldsymbol{y}) - 2l(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y})}{2l(\boldsymbol{\theta}_c, \boldsymbol{z}, \boldsymbol{y}) - 2l(\boldsymbol{\theta}_0, \boldsymbol{z}, \boldsymbol{y})}, \tag{19}$$

where Residual Deviance is twice the difference between the log-likelihood of the full model $l(\boldsymbol{\theta}_c, \boldsymbol{z}, \boldsymbol{y})$ and the proposed model $l(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y})$ whereas Null Deviance is twice the difference between the log-likelihood of the full model and the null model $l(\boldsymbol{\theta}_0, \boldsymbol{z}, \boldsymbol{y})$ (Smith and McKenna, 2013). The null model consists of the intercept only where it corresponds to the scenario of zero expectations for the GFR model or $I_{j,m}$ in Eq. (11) for the RBF-GFR model. The full model is where the number of parameters is equal to the number of data points.

5. Data

We applied all modelling approaches to four distinct datasets, outlined in Table 1.

5.1. Simulated datasets

5.1.1. Simulated dataset in Matthiopoulos et al. (2015)

Test data for this paper were derived from multiple simulated instances, each representing subpopulations of a species living in different landscapes. Each instance was obtained from a realisation of an individual-based simulation within a small (50x50 cell) spatial arena, where rudimentary energetics gave rise to simple demographic processes and population dynamics. Two spatially autocorrelated environmental variables (*food*, designated as a resource and *temperature*, set as a condition) were distributed across the arena. Individuals were programmed to move up gradients of environmental profitability (i.e., food richness moderated by temperature) and their movement was subject to perception error. The population size (N) associated with the entire sampling instance was also included to capture any density-dependent effects on the distribution of the animals.

The data set contained 20 landscape scenarios whose dynamics were modelled over 20 years, yielding 400 different sampling instances. Different subsets of this data set could thus be used to emulate realistic scenarios for sample size across time and space.

5.1.2. Simulated dataset from Matthiopoulos et al. (2011)

This dataset contains a simpler version of the above individual-based data that looks at only two resources (e.g., *food* and *cover*), deemed to be required in alternation. This simulation considered no demography or density-dependence. The simulated animals in this simplified version of the IBM climbed up gradients of food when they were hungry and then climbed up gradients of cover when they were sated. Feeding occurred through a Holling type II model (Matthiopoulos, 2011).

Table 1
Overview table of the datasets including the habitat variables, number of sample instances, and data size.

Dataset	Habitat variables	Sample number	Data size
Simulated Dataset in Matthiopoulos et al. (2015)	Food, temperature, and population size	400	200,000
Simulated Dataset in Matthiopoulos et al. (2011)	Food, and cover	20	50,000
Sparrow	Grass, bush, roof, and colony size	32	1,280
Wolf	Distance to high human use, distance to edge, slope, burnt, alpine, shrub, rock and herbaceous	11	18,042

5.2. Real-life datasets

5.2.1. Sparrow population dataset

The sparrow data set comprised habitat use and population size data first analysed in Matthiopoulos et al. (2019). These data were collected by the UK Royal Society for the Protection of Birds (RSPB) and the University of Glasgow during the breeding season in 2014 from 32 colonies in the United Kingdom. The sparrow habitat use variable in the data is represented by values of 1 and 0 based on the presence of a sparrow in each cell (strictly, a count, not an occupancy variable). The three main variables in the dataset, the estimated percentages of *grass*, *bush*, and *roof* for each cell were captured from Google Earth images. The response variables for this data set are binary count data, and the binomial log likelihood seen in Eq. (14) was thus used when fitting the original GFR model and the RBF-GFR model to the data using the habitat variables *grass*, *bush*, and *roof* as the main covariates in the models, along with *colony size*, set as count of the maximum number of males measured in each colony, which was used as an additional explanatory variable.

5.2.2. Wolf dataset

The dataset on telemetry of habitat use by wolves was used in Mathiopoulos et al. (2011) to fit the original GFR model. The analysis employed a use-availability approach representing response data as either 0 or 1. These data cover 11 wolves, that were members of five different packs. The dataset comprised three continuous variables, *distance to high human use*, *distance to edge*, and *slope*, and represented landcover as a factor with levels for *burnt*, *alpine*, *shrub*, *rock*, and *herbaceous*.

6. Results

The results from applying all the models included in the study, as shown Fig. 1 for the first simulated, second simulated, sparrow and wolf datasets are summarized in Fig. 2 and the details can be found in Supplements A.5–A.8.

6.1. Model ranking

The models were then ranked by performance for each of the four data sets in turn, as shown in Fig. 2. This table thus offers the ranks and the detailed out of data performance of all the models included in the study, as shown in Fig. 1. While none of the individual models consistently outperformed all other models across all data sets, a pattern did emerge whereby the ensemble methods, which use bagging or boosting for the creation of random forests, tend to outperform all other models as a class (namely the "class" of ensemble methods, as opposed to individual models). In particular, the combination of the proposed RBF-GFR and GFR models with bagging, as represented by the two models shown in the top rows of Fig. 2, consistently achieve ranks in the top 40% of the performance spectrum. This offers evidence of more stable performance than the non-ensemble models, while the latter show higher variability, as exemplified by the regularised RBF-GFR model, which appears as the best model for the second simulated data set, but as the third-worst model for the first simulated data set. Regularization was not applied to the individual models included in the ensembles, with the results thus suggesting that, in terms of improving out-of-sample generalisation performance, model averaging over ensembles offers an alternative to regularisation, confirming similar findings in Machine Learning literature (Sollich and Krogh, 1996). The combination of the proposed RBF-GFR model with random forests (RBF-GFR-RF) produced the best model overall, consistently achieving a place in the top three performance rankings. An important additional finding was that almost all the methods proposed in this study outperform the original GFR model from Matthiopoulos et al. (2011), which was the initial aim motivating the present work. As shown in Fig. 2, the GFR model never achieves a rank better than 6. R^2_{DEV} in Eq. (18) is generally a better behaved measurement than R^2 in Eq. (17) for count data as described in Section 4.2. We used R^2_{DEV} to calculate the out-of-sample predictive performance in these datasets as shown in Fig. S9 in Supplement A.9. However, the overall ranks using R^2_{DEV} are not different from the overall ranks using R^2 in Eq. (17) (comparing the average rank in Fig. 2 with ranks in Fig. S9 in Supplement A.9).

6.2. Visualising spatial predictions of models

The predictions of animal habitat usage derived from the models were presented in spatial maps and further visualisations of species abundance were generated using the second simulated dataset, which contained 20 sample instances, each containing 2,500 observations (50 x 50 arena). One map was reserved from the cross-validation scheme and predicted from the rest of the data. Predictions for sampling instance # 1 is shown for illustrative purposes in the main text, along with further results in in Supplement A.12.

Fig. 3 shows a heat map of species abundance and geographical predictions of abundance in terms of latitude and longitude for the ground truth and the various models shown in Fig. 1.

These results suggest that the RBF-GFR-RF model, which is overall

	Sim1	Sim2	Sparrow	Wolf	Average
RBF-GFR-RF	0.937	0.593	0.86	0.76	0.782
GFR-RF	0.936	0.443	0.73	0.769	0.719
RBF-GFR-XGBoost	0.94	0.535	0.861	0.354	0.673
GFR-XGBoost	0.944	0.491	0.834	0.405	0.669
RBF-GFR-CART	0.822	0.44	0.884	0.182	0.582
GFR-CART	0.821	0.235	0.619	0.222	0.474
Reg RBF-GFR	0.796	0.635	0.252	0.199	0.471
Reg GFR	0.796	0.359	0.241	0.234	0.408
GLM	0.731	0.256	0.265	0.215	0.367
GFR	0.837	-0.97	0.338	0.156	0.09
RBF-GFR	0.837	-2.4	0.356	0.219	-0.103

Fig. 2. Rank table of out-of-sample R^2's of the models using the two simulated, sparrow, wolf datasets and the average score of out-of-sample R^2. The shading of colours indicates the ranks of the models. For each column, the colour shading ranges from yellow to dark red, with yellow indicating the lowest score in the respective column, and dark red indicating the maximum value.

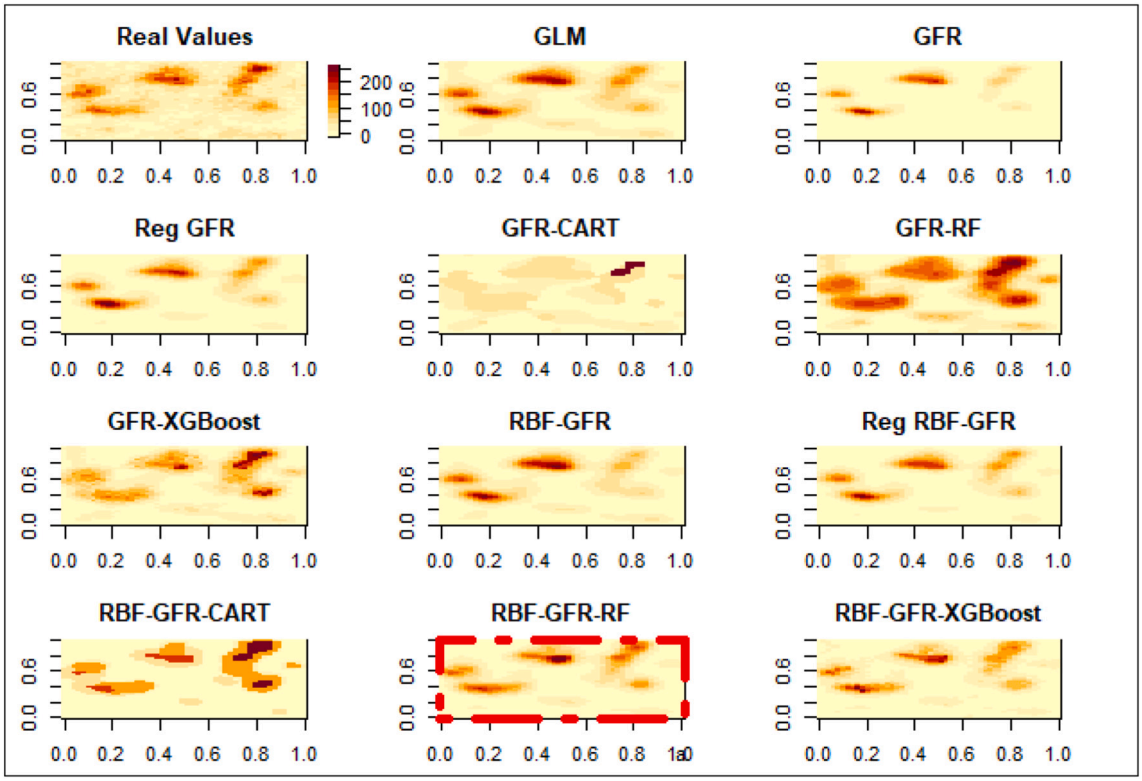


Fig. 3. A heat map of abundance and geographical predictions of the abundance of sample instance # 1 from the second simulated dataset in terms of geographical dimensions: latitude and longitude for the ground truth and the various models shown in Fig. 1. Light colours indicate low abundance levels, so that the abundance levels increase as the colour shading gets darker. We use the same output range for all models; all models were plotted on the same scale as the ground truth. The map with red borders is the best predictive model based on out-of-sample R^2.

the best model according to Fig. 2, also offers the best qualitative agreement with the ground truth for predicted spatial abundance profiles. The RBF-GFR-RF predictions faithfully reproduce the high-intensity hotspots near the top right corner of the map, around coordinates (0.8,0.8), as well as those near the left margin, around coordinates (0.2,0.4). The alternative models also tend to qualitatively capture the ground truth pattern, but these display larger deviations. For instance, the GLM model shows reasonable agreement with the ground truth, i.e. when plotted on the same scale as the ground truth; however, on its individually adjusted intensity scale (Fig. 4), the GLM predictions are systematically lower than the ground truth values, implying that the GLM model systematically underestimates extremes, while the GFR model shows an opposing trend, systematically overestimating extremes, as indicated by the white patches in Fig. 4. Furthermore, the out-of-sample R^2 scores shown in Table 2 of sample instance # 1 for the various models used to predict the heat maps in Figs. 3 and 4 illustrate our finding from the heat maps where the RBF-GFR-RF score is higher than the scores of other models.

6.3. Variable ranking

Colony size, measured by the maximum number of males in each colony, is the most important feature for the sparrow population, which was determined using the best two models overall, the RF approach in combination with the GFR and RBF-GFR models. The percentage of *bush*, on the other hand, has the lowest importance score compared to the other main variables, as seen in Fig. 5. The importance scores were calculated using the mean decrease in accuracy from permuting out-of-bag data; the mathematical details for this calculation can be found in Han et al. (2016). Both features positively affect the habitat suitability of the sparrows based on the GFR and RBF-GFR models.

For the wolf dataset, using the RF approach in combination with the RBF-GFR model, "*distance to high human use*" is the most important covariate, positively affecting the wolves' habitat preference. The *slope* and *distance to edges* strongly influence habitat preference; decreasing the *slope* or increasing the *distance to edge* increases habitat preference. In contrast, *rocks* have the lowest impact on the wolves' habitat preference, as seen in the right panel of Fig. 6. However, the *slope* is more important than the *distance to human use* based on the RF approach in combination with the GFR model, as seen in the left panel of Fig. 6.

7. Discussion and conclusions

Given their extensive applications to questions of anthropogenic change (Iturbide et al., 2018) it is imperative that the predictive ability of SDMs be assessed by transferring models built in one region or time to another spatiotemporal frame where the prevailing environmental conditions are different, and possibly outside the range of covariate values previously measured (Duque-Lazo et al., 2016; Elith and Leathwick, 2009). The discrepancy between phenomenological models and the complex biological mechanisms they try to capture leads to the existence of highly non-linear functional responses in species-habitat associations (Mysterud and Ims, 1998). These are especially complicated in the cases of animals with higher mobility and cognition. Despite the increasingly recognised challenges of model transferability (Petitpierre et al., 2017; Yates et al., 2018; Wenger and Olden, 2012; Peterson et al., 2003; Randin et al., 2006; Townsend Peterson et al., 2007; Márcia Barbosa et al., 2009; Sundblad et al., 2009; Wenger et al., 2011), particularly for purely statistical models such as SDMs, there is a dearth of methods for functional responses in SDMs and a lack of comparative validation of such methods with synthetic and real data.

Here, we have built on the suggestions of Boyce and McDonald (1999) and the early implementation of the GFR by Matthiopoulos et al. (2011). Our work addressed the lack of flexibility and control in the

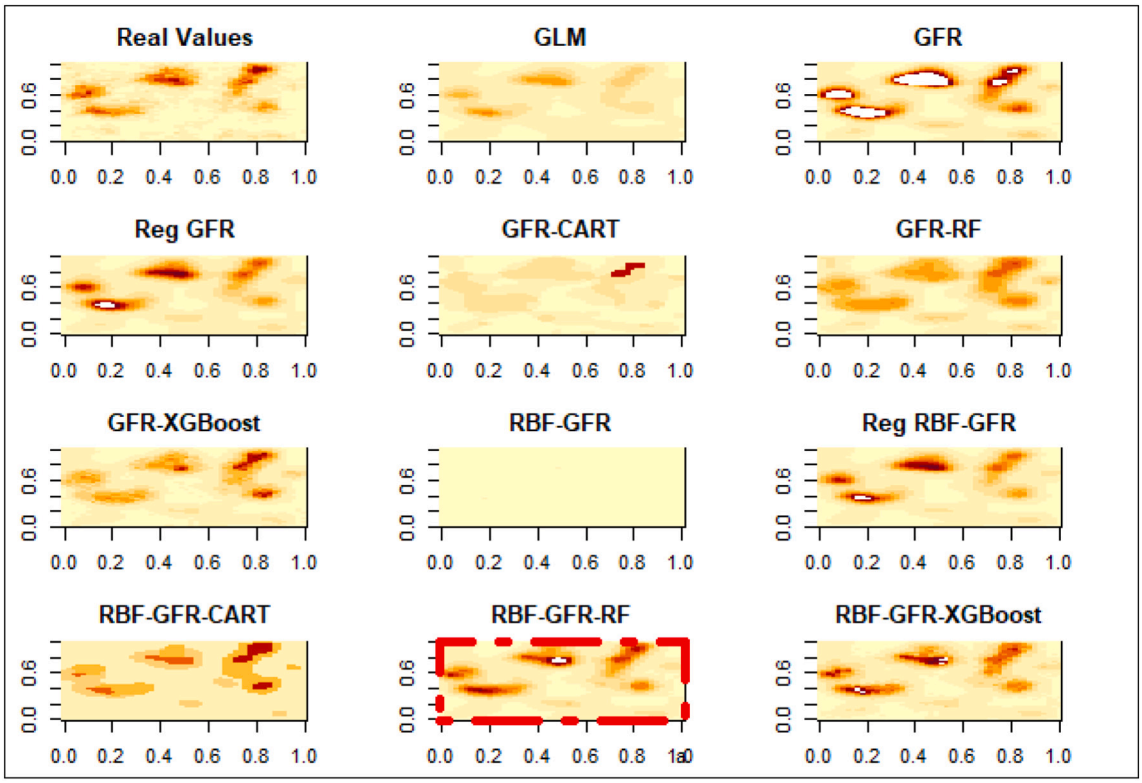


Fig. 4. A heat map of abundance and geographical predictions of the abundance of sample instance # 1 from the second simulated dataset in terms of geographical dimensions: latitude and longitude for the ground truth and the various models shown in Fig. 1. The colour range may be different for different models but the minimum and maximum values for which colours should be plotted are limited by the minimum and maximum numbers of the true values and model outputs that are larger than the maximum value of the truth are treated as missing values and are shown in white. The map with red borders is the best predictive model based on out-of-sample R^2.

Table 2
Out-of-sample R^2 scores for the various models shown in Fig. 1 of sample instance # 1 from the second simulated dataset.

Models	Out-of-sample R^2
GLM	0.612
GFR	−3.6
Reg GFR	0.695
GFR-CART	0.327
GFR-RF	0.673
GFR-XGBoost	0.764
RBF-GFR	−0.734
Reg RBF-GFR	0.742
RBF-GFR-CART	0.628
RBF-GFR-RF	0.769
RBF-GFR-XGBoost	0.713

original GFR model and investigated how alternative models might be implemented within the broader GFR framework. Replacing the global polynomial functions of the original GFR with radial basis functions as well as using the Gaussian mixtures approximation to approximate habitat availability allowed the RBF-GFR model to be more flexible than the original GFR model. This flexibility poses the risk of overfitting, the opposite behaviour to the rigidity of classic SDMs (Paton and Matthiopoulos, 2016). Overfitting is a fundamental issue for achieving transferable GFR models (Wenger and Olden, 2012). In this respect, we found regularization approaches to be effective in controlling overfitting in GFRs.

We have also explored the suggestion made by recent publications that machine-learning methods achieve better results than traditional likelihood-based methods (Heikkinen et al., 2012; Elith* et al., 2006; Lawler et al., 2006; Prasad et al., 2006). We achieved this by combining the RBF-GFR and GFR models with CART methods and, as a further extension, we use ensemble approaches, random forests (RF) and extreme gradient boosting (XGBoost). Ignoring structure dependence in data increases the susceptibility to overfitting and causes autocorrelations and non-independence of model residuals (Roberts et al., 2017). The block cross-validation approach addresses the autocorrelation of dataset structures (Roberts et al., 2017). We have implemented the block cross-validation approach to account for autocorrelation of dataset structures. The simulated dataset in Matthiopoulos et al. (2015) was simulated from multiple instances where each instance consisted of 500 observations representing a sub-population in a different landscape. The dataset has a spatial structure based on these scenarios. We have used these scenarios as dataset blocks when we applied the models. We have used a cross-validation approach based on these blocks (10-cross--validation where each fold contained 40 blocks) to measure the out-of-sample predicted performance. The simulated dataset in Matthiopoulos et al. (2011) is a simpler version of the Matthiopoulos et al. (2015) dataset consisting of 20 blocks (sample instances). We set the blocks to be the folds, meaning that we have used a 20-block cross-validation approach. In the sparrow population dataset, we have used the 32 colonies as the blocks and folds, resulting in 32-block cross-validation. The wolf dataset has a grouping structure based on the five packs that the wolves belong to. We have used five-blocks cross-validation based on these five packs when applying the models.

The resulting performance of the GFR, RBF-GFR and their extensions on four datasets showed that considerable gains could be achieved in predictive performance and that these gains were approximately consistent across data sets. Going from a global median regression GFR model to the radial basis function model, offered local flexibility in the functional response curves but generated only moderate improvements in out-of-sample R^2 score. However, combining the ensemble approach

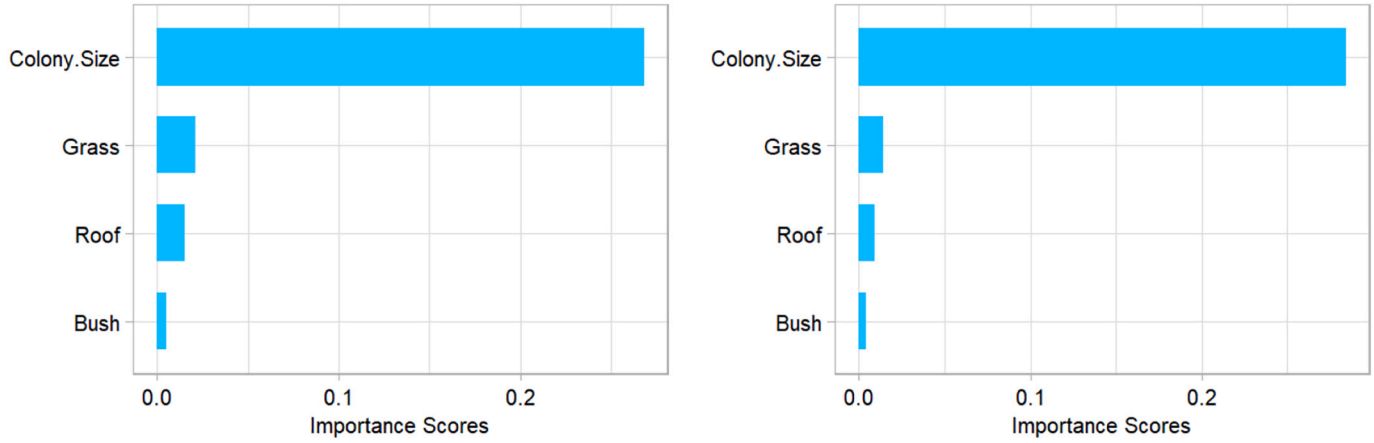


Fig. 5. Importance scores for the main variables in the sparrow population dataset, using the GFR-RF model in the left panel and the BF-GFR-RF model in the right panel.

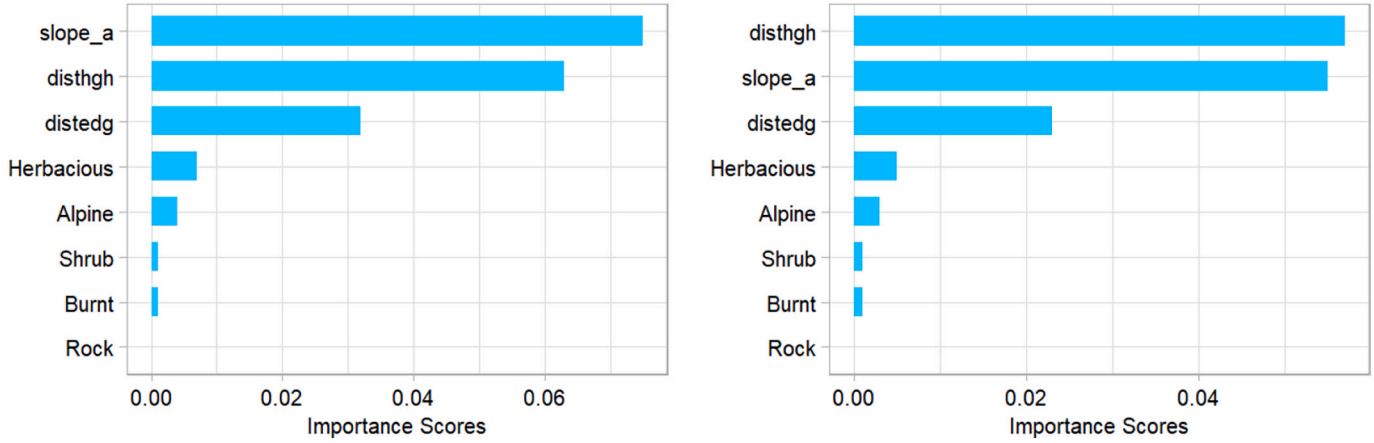


Fig. 6. Importance scores for the main variables in the wolf dataset, using the GFR-RF model in the left panel and the RBF-GFR-RF model in the right panel.

using bagging and boosting with the GFR and RBF-GFR models substantially improved the out-of-sample R^2 scores. In general, ensemble methods, such as bagging and boosting were consistently among the top-scoring models, with no evidence of overfitting, while for other models, performance varied more drastically. In essence, the original GLM model provides a much flatter version of the truth, while applying overly flexible extensions of the GFR model can increase the risk of exaggerating extremes in species distribution (i.e. over-predicting abundance hot-spots/peaks and under-predicting cold-spots/troughs). We also replicated the finding that model ensembles can perform the same role as regularization, buffering the models' predictions from such biases of exaggeration.

It has been clear that simple SDMs homogenise predictions (Paton and Matthiopoulos, 2016) and that polynomial GFRs can be overly volatile. The key message from our work is that using measures against overfitting (i.e. either regularisation or ensemble modelling) can give consistent and impressive improvements in out-of-data predictions, in some cases raising the R^2 from 0.25 to 0.85 (typical gains were from 0.35 to 80). This comes at a cost of implementation. The libraries required for fitting these models are not as user-friendly as the base GLM approaches, so we look forward to more automatic software workflows for functional responses in the future. A key advantage of such approaches is that regularisation is an efficient way of achieving parsimonious model so, in addition to GFR flexibility, it would simultaneously facilitate issues of covariate selection.

The differences in performance between different datasets are as interesting as the consistent features of Fig. 2, but considerably harder to explain. Improvements in predictive performance were most dramatic in the two real data sets (the sparrows and wolves), despite the fact that simulated data sets were designed to offer better adherence to the spatial stationarity of covariates and distributional assumptions made by the models fitted to those data. Several reasons have been mentioned for poor transferability in the literature. For example, the poorer information content of occupancy compared to abundance data (Yates et al., 2018), the definition of the scale of habitat availability (Márcia Barbosa et al., 2009; Paton and Matthiopoulos, 2016; Beyer et al., 1550), the ranging behaviour of the study species (Vanreusel et al., 2007; Yates et al., 2018; Wogan, 2016). The suggestion of stratifying the data (Yates et al., 2018) as a solution to the problem of varying conditions is addressed more efficiently under the auspices of the GFR family of models.

The work done here instills computational robustness into a method that has been previously shown to work and opens the avenue for further comparative studies and biological interpretation. Better visualisation methods for how regression coefficients in species-habitat association models adapt to changes in overall habitat composition provide a link between these de facto phenomenological models with some quintessentially mechanistic fields of environmental sciences, particularly behavioural and landscape ecology. For example, the analysis of the functions $\gamma_i(\mathbf{x})$ in Eq. (7), as derived from our GFR models, has clear parallels with models of consumer choice developed in the areas of ethology (Sih and Christensen, 2001) and the humanities (Raghavarao

et al., 2010).

Similarly, by extending the SDMs to account for regional environmental context, the models might provide clues about more holistic processes at the level of landscape ecology. Therefore, the GFR, an approach that begun with the sole aim of trying to improve predictive performance may, through the generation of new hypotheses for habitat selection, lead to new insights about fundamental biology at the level of the individual and the landscape.

Author contributions

The statistical methodology was jointly developed by all authors. Shaykhah Aldossari implemented the methods, ran the simulations and carried out the numerical analysis. All authors analysed the results, designed the simulation studies and contributed to writing the manuscript.

Data availability

The code that generated dataset in Matthiopoulos et al. (2015) can be found at the Supplementary material for this paper at https://doi.org/10.1890/14-2244.1.sm. Details on the simulated dataset from Matthiopoulos et al. (2011) are provided in Appendix A and Supplement 1 for this paper at https://doi.org/10.1890/10-0751.1. The sparrow population dataset can be found online in electronic Supplementary at https://dx.doi.org/10.6084/m9 (Matthiopoulos et al., 2019).

Declaration of Competing Interest

None.

Acknowledgments

This work was conducted as part of the PhD for Shaykhah Aldossari, funded by Saudi Arabia Cultural Bureau SACB in the UK. Dirk Husmeier was supported by a grant from the Leverhulme Trust, reference number RPG-2018-398. We thank R. MacLeod, C. Field and the volunteers of the RSPB for collecting the sparrow data. We thank M. Hebblewhite for the use of the wolf data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecoinf.2022.101803.

References

Aarts, Geert, Fieberg, John, Matthiopoulos, Jason, 2012. Comparative interpretation of count, presence–absence and point methods for species distribution models. Methods Ecol. Evol. 3 (1), 177–187.

Aldossari, Shaykhah, Matthiopoulos, Jason, Husmeier, Dirk, 2020. Statistical modelling of habitat selection. In: The 35th International Workshop on Statistical Modelling. Servicio Editorial de la Universidad del País Vasco.

Aldossari, Shaykhah, Husmeier, Dirk, Matthiopoulos, Jason, 2021. Generalized functional responses in habitat selection fitted by decision trees and random forests. In: The 3rd International Conference on Statistics: Theory and Applications (ICSTA'21). Avestia Publishing. https://doi.org/10.11159/icsta21.125.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157 (2–3), 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3.

Bahn, Volker, McGill, Brian J., 2013. Testing the predictive performance of distribution models. Oikos 122 (3), 321–331. https://doi.org/10.1111/j.1600-0706.2012.00299.x.

Barbet-Massin, Morgane, Rome, Quentin, Villemant, Claire, Courchamp, Franck, 2018. Can species distribution models really predict the expansion of invasive species? PloS One 13 (3), e0193085. https://doi.org/10.1371/journal.pone.0193085.

Márcia Barbosa, A., Real, Raimundo, Mario Vargas, J., 2009. Transferability of environmental favourability models in geographic space: the case of the Iberian desman (Galemys pyrenaicus) in Portugal and Spain. Ecol. Model. 220 (5), 747–754. https://doi.org/10.1016/j.ecolmodel.2008.12.004.

Beyer, Hawthorne L., Haydon, Daniel T., Morales, Juan M., Frair, Jacqueline L., Hebblewhite, Mark, Mitchell, Michael, Matthiopoulos, Jason, 2010. The interpretation of habitat preference metrics under use–availability designs. Philos. Trans. R. Soc. B: Biol. Sci. 365 (1550), 2245–2254. https://doi.org/10.1098/rstb.2010.0083.

Bishop, Christopher M., 1995. Neural Networks for Pattern Recognition. Clarendon Press.

Bishop, Christopher M., 2006. Pattern recognition and machine learning. Springer.

Bjørneraas, Kari, Herfindal, Ivar, Solberg, Erling Johan, Sæther, Bernt-Erik, Moorter, Bram van, Rolandsen, Christer Moe, 2012. Habitat quality influences population distribution, individual space use and functional responses in habitat selection by a large herbivore. Oecologia 168 (1), 231–243.

Boyce, Mark S, McDonald, Lyman L, 1999. Relating populations to habitats using resource selection functions. Trends Ecol. Evol. 14 (7), 268–272. https://doi.org/10.1016/S0169-5347(99)01593-1.

Breiman, L., Friedman, J., Olshen, R., 1984. Classification and Regression Trees. Wadsworth.

Breiman, Leo, 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Colin Cameron, A., Windmeijer, Frank AG, 1996. R-squared measures for count data regression models with applications to health-care utilization. J. Bus. Econ. Stat. 14 (2), 209–220. https://doi.org/10.1080/07350015.1996.10524648.

Chen, Tianqi, Guestrin, Carlos, 2016. XGBoost: a scalable tree boosting system. KDD. https://doi.org/10.1145/2939672.2939785, pp. 785–794.

Dormann, Carsten F, 2007. Promising the future? global change projections of species distributions. Basic Appl. Ecol. 8 (5), 387–397. https://doi.org/10.1016/j.baae.2006.11.001.

Duque-Lazo, J., Van Gils, H.A.M.J., Groen, T.A., Navarro-Cerrillo, R.M., 2016. Transferability of species distribution models: The case of Phytophthora cinnamomi in Southwest Spain and Southwest Australia. Ecol. Model. 320, 62–70. https://doi.org/10.1016/j.ecolmodel.2015.09.019.

Ehrlén, Johan, Morris, William F., 2015. Predicting changes in the distribution and abundance of species under environmental change. Ecol. Lett. 18 (3), 303–314. https://doi.org/10.1111/ele.12410.

Elith, Jane, Leathwick, John R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Elith, Jane, Graham*, Catherine H., Anderson, Robert P., Dudík, Miroslav, Ferrier, Simon, Guisan, Antoine, Hijmans, Robert J., Huettmann, Falk, Leathwick, John R., Lehmann, Anthony, et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Evans, Matthew R., Norris, Ken J., Benton, Tim G., 2012. Predictive ecology: systems approaches.

Fordham, Damien A., Resit Akçakaya, H., Alroy, John, Saltré, Frédérik, Wigley, Tom M. L., Brook, Barry W., 2016. Predicting and mitigating future biodiversity loss using long-term ecological proxies. Nat. Climate Change 6 (10), 909–916. https://doi.org/10.1038/nclimate3086.

Gillies, Cameron S, St. Clair, Colleen Cassady, 2010. Functional responses in habitat selection by tropical birds moving through fragmented forest. J. Appl. Ecol. 47 (1), 182–190.

Gillies, Cameron S., Hebblewhite, Mark, Nielsen, Scott E., Krawchuk, Meg A., Aldridge, Cameron L., Frair, Jacqueline L., Saher, D. Joanne, Stevens, Cameron E., Jerde, Christopher L., 2006. Application of random effects to the study of resource selection by animals. J. Anim. Ecol. 75 (4), 887–898. https://doi.org/10.1111/j.1365-2656.2006.01106.x.

Godvik, Inger Maren Rivrud, Loe, Leif Egil, Vik, Jon Olav, Veiberg, VebjØrn, Langvatn, Rolf, Mysterud, Atle, 2009. Temporal scales, trade-offs, and functional responses in red deer habitat selection. Ecology 90 (3), 699–710.

Han, Hong, Guo, Xiaoling, Hua, Yu., 2016. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In: 2016 7th ieee international conference on software engineering and service science (icsess). IEEE, pp. 219–224.

Hastie, Trevor, Tibshirani, Robert, 1993. Varying-coefficient models. J. Roy. Stat. Soc.: Ser. B (Methodol.) 55 (4), 757–779. https://doi.org/10.1111/j.2517-6161.1993.tb01939.x.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, 2008. The Elements of Statistical Learning. Springer.

Hastie, Trevor, Qian, Junyang, Tay, Kenneth, 2016. An introduction to glmnet.

Heikkinen, Risto K, Marmion, Mathieu, Luoto, Miska, 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography 35 (3), 276–288. https://doi.org/10.1111/j.1600-0587.2011.06999.x.

Holbrook, Joseph D., Olson, Lucretia E., DeCesare, Nicholas J., Hebblewhite, Mark, Squires, John R., Steenweg, Robin, 2019. Functional responses in habitat selection: clarifying hypotheses and interpretations. Ecol. Appl. 29 (3), e01852.

Houlahan, Jeff E., McKinney, Shawn T., Michael Anderson, T., McGill, Brian J., 2017. The priority of prediction in ecological understanding. Oikos 126 (1), 1–7. https://doi.org/10.1111/oik.03726.

Iturbide, Maialen, Bedia, Joaquín, Gutiérrez, José Manuel, 2018. Background sampling and transferability of species distribution model ensembles under climate change. Global Planet. Change 166, 19–29. https://doi.org/10.1016/j.gloplacha.2018.03.008.

Kindsvater, Holly K., Dulvy, Nicholas K., Horswill, Cat, Juan-Jordá, Maria-José, Mangel, Marc, Matthiopoulos, Jason, 2018. Overcoming the data crisis in biodiversity conservation. Trends Ecol. Evol. 33 (9), 676–688. https://doi.org/10.1016/j.tree.2018.06.004.

Lawler, Joshua J., White, Denis, Neilson, Ronald P., Blaustein, Andrew R., 2006. Predicting climate-induced range shifts: model differences and model reliability.

Glob. Change Biol. 12 (8), 1568–1584. https://doi.org/10.1111/j.1365-2486.2006.01191.x.
Leys, Christophe, Ley, Christophe, Klein, Olivier, Bernard, Philippe, Licata, Laurent, 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49 (4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013.
MacKenzie, Darryl I., Nichols, James D., Royle, J. Andrew, Pollock, Kenneth H., Bailey, Larissa L., Hines, James E., 2017. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier.
Maris, Virginie, Huneman, Philippe, Coreau, Audrey, Kéfi, Sonia, Pradel, Roger, Devictor, Vincent, 2018. Prediction in ecology: promises, obstacles and clarifications. Oikos 127 (2), 171–183. https://doi.org/10.1111/oik.04655.
Matthiopoulos, Jason, 2011. How to be a quantitative ecologist: the'A to R'of green mathematics and statistics. John Wiley & Sons.
Matthiopoulos, Jason, Hebblewhite, Mark, Aarts, Geert, Fieberg, John, 2011. Generalized functional responses for species distributions. Ecology 92 (3), 583–589. https://doi.org/10.1890/10-0751.1.
Matthiopoulos, Jason, Fieberg, John, Aarts, Geert, Beyer, Hawthorne L., Morales, Juan M., Haydon, Daniel T., 2015. Establishing the link between habitat selection and animal population dynamics. Ecol. Monogr. 85, 413–436. https://doi.org/10.1890/14-2244.1.
Matthiopoulos, Jason, Field, Christopher, MacLeod, Ross, 2019. Predicting population change from models based on habitat availability and utilization. Proc. R. Soc. B 286 (1901), 20182911. https://doi.org/10.1098/rspb.2018.2911.
Matthiopoulos, Jason, Fieberg, John, Aarts, Geert, 2020a. Species-habitat associations: spatial data, predictive models, and ecological insights.
Matthiopoulos, Jason, Fieberg, John, Aarts, Geert, Barraquand, Frédéric, Kendall, Bruce E., 2020b. Within reach? habitat availability as a function of individual mobility and spatial structuring. Am. Nat. 195 (6), 1009–1026. https://doi.org/10.1086/708519.
Mauritzen, Mette, Belikov, Stanislav E., Boltunov, Andrei N., Derocher, Andrew E., Hansen, Edmond, Wiig, Øystein, Yoccoz, Nigel, 2003. Functional responses in polar bear habitat selection. Oikos 100 (1), 112–124. https://doi.org/10.1034/j.1600-0706.2003.12056.x.
Mouquet, Nicolas, Lagadeuc, Yvan, Devictor, Vincent, Doyen, Luc, Duputié, Anne, Eveillard, Damien, Faure, Denis, Garnier, Eric, Gimenez, Olivier, Huneman, Philippe, et al., 2015. Predictive ecology in a changing world. J. Appl. Ecol. 52 (5), 1293–1310. https://doi.org/10.1111/1365-2664.12482.
Murphy, Kevin P, 2012. Machine learning: a probabilistic perspective. MIT press. https://doi.org/10.1080/09332480.2014.914768.
Mysterud, Atle, Ims, Rolf Anker, 1998. Functional responses in habitat use: availability influences relative use in trade-off situations. Ecology 79 (4), 1435–1441. https://doi.org/10.1890/0012-9658(1998)079[1435:FRIHUA]2.0.CO;2.
Mysterud, Atle, Ims, Rolf Anker, 1999. Relating populations to habitats. Trends Ecol. Evol. 14 (12), 489–490 doi:10.1.1.707.9728.
Paton, Robert Stephen, Matthiopoulos, Jason, 2016. Defining the scale of habitat availability for models of habitat selection. Ecology 97 (5), 1113–1122. https://doi.org/10.1890/14-2241.1.
Peterson, A. Townsend, Papes, Monica, Kluza, Daniel A., 2003. Predicting the potential invasive distributions of four alien plant species in North America. Weed Sci. 51 (6), 863–868. https://doi.org/10.1614/P2002-081.
Petitpierre, Blaise, Broennimann, Olivier, Kueffer, Christoph, Daehler, Curtis, Guisan, Antoine, 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. Glob. Ecol. Biogeogr. 26 (3), 275–287. https://doi.org/10.1111/geb.12530.
Phillips, Steven J., Anderson, Robert P., Schapire, Robert E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.
Platt, John, et al., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv. Large Margin Classifiers 10 (3), 61–74.
Prasad, Anantha M., Iverson, Louis R., Liaw, Andy, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9 (2), 181–199. https://doi.org/10.1007/s10021-005-0054-1.
Raghavarao, Damaraju, Wiley, James B., Chitturi, Pallavi, 2010. Choice-based conjoint analysis: models and designs. Chapman and Hall/CRC. https://doi.org/10.1201/9781420099973.
Randin, Christophe F., Dirnböck, Thomas, Dullinger, Stefan, Zimmermann, Niklaus E., Zappa, Massimiliano, Guisan, Antoine, 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. 33 (10), 1689–1703. https://doi.org/10.1111/j.1365-2699.2006.01466.x.
Renner, Ian W., Warton, David I., 2013. Equivalence of maxent and poisson point process models for species distribution modeling in ecology. Biometrics 69 (1), 274–281. https://doi.org/10.1111/j.1541-0420.2012.01824.x.
Roberts, David R., Bahn, Volker, Ciuti, Simone, Boyce, Mark S., Elith, Jane, Guillera-Arroita, Gurutzeta, Hauenstein, Severin, Lahoz-Monfort, José J., Schröder, Boris, Thuiller, Wilfried, et al., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40 (8), 913–929.
Schwarz, Gideon, 1978. Estimating the dimension of a model. The annals of statistics, pp. 461–464. doi:http://www.jstor.org/stable/2958889.
Sequeira, Ana M.M., Bouchet, Phil J., Yates, Katherine L., Mengersen, Kerrie, Caley, M. Julian, 2018. Transferring biodiversity models for conservation: opportunities and challenges. Methods Ecol. Evol. 9 (5), 1250–1264. https://doi.org/10.1111/2041-210X.12998.
Sih, Andrew, Christensen, Bent, 2001. Optimal diet theory: when does it work, and when and why does it fail? Anim. Behav. 61 (2), 379–390. https://doi.org/10.1006/anbe.2000.1592.
Smith, Thomas J., McKenna, Cornelius M., 2013. A comparison of logistic regression pseudo r2 indices. Multiple Linear Regression Viewpoints 39 (2), 17–26.
Sollich, Peter, Krogh, Anders, 1996. Learning with ensembles: how over-fitting can be useful. Neural Inf. Process. Syst. (NIPS) 8, 190–196 doi:https://proceedings.neurips.cc/paper/1995/file/1019c8091693ef5c5f55970346633f92-Paper.pdf.
Sundblad, Göran, Härmä, Meri, Lappalainen, Antti, Urho, Lauri, Bergström, Ulf, 2009. Transferability of predictive fish distribution models in two coastal systems. Estuar. Coast. Shelf Sci. 83 (1), 90–96. https://doi.org/10.1016/j.ecss.2009.03.025.
Tessarolo, Geiziane, Lobo, Jorge M., Rangel, Thiago Fernando, Hortal, Joaquín, 2021. High uncertainty in the effects of data characteristics on the performance of species distribution models. Ecol. Ind. 121, 107147 https://doi.org/10.1016/j.ecolind.2020.107147.
Torres, Leigh G., Sutton, Philip J.H., Thompson, David R., Delord, Karine, Weimerskirch, Henri, Sagar, Paul M., Sommer, Erica, Dilley, Ben J., Ryan, Peter G., Phillips, Richard A., 2015. Poor transferability of species distribution models for a pelagic predator, the grey petrel, indicates contrasting habitat preferences across ocean basins. PLoS One 10 (3), e0120014. https://doi.org/10.1371/journal.pone.0120014.
Townsend Peterson, A., Papeş, Monica, Eaton, Muir, 2007. Transferability and model evaluation in ecological niche modeling: a comparison of garp and maxent. Ecography 30 (4), 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x.
Travers, Henry, Selinske, Matthew, Nuno, Ana, Serban, Anca, Mancini, Francesca, Barychka, Tatsiana, Bush, Emma, Rasolofoson, Ranaivo A., Watson, James E.M., Milner-Gulland, E.J., 2019. A manifesto for predictive conservation. Biol. Conserv. 237, 12–18. https://doi.org/10.1016/j.biocon.2019.05.059.
Vanreusel, Wouter, Maes, Dirk, Van Dyck, Hans, 2007. Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. Conserv. Biol. 21 (1), 201–212. https://doi.org/10.1111/j.1523-1739.2006.00577.x.
Warton, David, Aarts, Geert, 2013. Advancing our thinking in presence-only and used-available analysis. J. Anim. Ecol. 82 (6), 1125–1134. https://doi.org/10.1111/1365-2656.12071.
Warton, David I., Shepherd, Leah C., 2010. Poisson point process models solve the pseudo-absence problem for presence-only data in ecology. Ann. Appl. Stat. 1383–1402. https://doi.org/10.1214/10-AOAS331.
Wenger, Seth J., Olden, Julian D., 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol. Evol. 3 (2), 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x.
Wenger, Seth J., Isaak, Daniel J., Dunham, Jason B., Fausch, Kurt D., Luce, Charles H., Neville, Helen M., Rieman, Bruce E., Young, Michael K., Nagel, David E., Horan, Dona L., et al., 2011. Role of climate and invasive species in structuring trout distributions in the interior Columbia River Basin, USA. Can. J. Fish. Aquat. Sci. 68 (6), 988–1008. https://doi.org/10.1139/f2011-034.
Wogan, Guinevere O.U., 2016. Life history traits and niche instability impact accuracy and temporal transferability for historically calibrated distribution models of North American birds. PLoS One 11 (3), e0151024. https://doi.org/10.1371/journal.pone.0151024.
Yates, Katherine L., Bouchet, Phil J., Julian Caley, M., Mengersen, Kerrie, Randin, Christophe F., Parnell, Stephen, Fielding, Alan H., Bamford, Andrew J., Ban, Stephen, Márcia Barbosa, A., et al., 2018. Outstanding challenges in the transferability of ecological models. Trends Ecol. Evol. 33 (10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001.
Zurell, Damaris, Jeltsch, Florian, Dormann, Carsten F, Schröder, Boris, 2009. Static species distribution models in dynamically changing systems: how good can predictions really be? Ecography 32 (5), 733–744. https://doi.org/10.1111/j.1600-0587.2009.05810.x.