JOURNAL ON COMMUNICATIONS

ISSN:1000-436X

Scopus®

REGISTERED

www.jocs.review

Ensemble learning and radial basis functions improve transferability of functional response models in species distribution modeling

1. **Shaykhah Abdullah Aldossari**

Department of Quantitative Methods, College of Business, King Faisal University ,Saudi Arabia.

ORCID: https://orcid.org/0009-0004-1937-8211

2. **Majdi Argoubi**

Department of Quantitative Methods, University of Sousse, Rue Khalifa Karoui, Sahloul, BP526, Sousse, Tunisia

ORCID: https://orcid.org/0000-0002-6560-5153

3. **Khaled Mili***

Department of Quantitative Methods, College of Business, King Faisal University ,Saudi Arabia.

ORCID: https://orcid.org/0000-0002-6309-5452

***Corresponding author: Khaled Mili.**

Abstract

Predictions from species distribution models (SDMs) often fail when transferred to new geographic regions or time periods, limiting their utility for biodiversity forecasting under environmental change. This lack of transferability stems from functional responses in habitat selection, where animals respond to the same habitat differently depending on the availability of alternative habitats. The Generalized Functional Response (GFR) framework explicitly models selection coefficients as functions of habitat availability, but existing polynomial implementations face a fundamental trade-off: low-order polynomials are too rigid to capture complex responses, while high-order polynomials overfit and transfer poorly.

We developed flexible extensions of the GFR framework that replace global polynomial functions with local radial basis functions (RBF-GFR) and combined both approaches with modern machine learning methods: classification and regression trees (CART), random forests (RF) and extreme gradient boosting (XGBoost). We systematically compared the out-of-sample predictive performance of 12 modeling approaches using block cross-validation across four contrasting datasets—two individual-based simulations with known ecological mechanisms, wolf telemetry data and sparrow colony surveys.

Ensemble methods combining functional response frameworks with RF or XGBoost consistently ranked in the top three performers across all datasets. Out-of-sample R^2 scores improved substantially over traditional GLMs, with increases from 0.25 to 0.85 in individual cases and typical gains of 0.20–0.50. The RBF-GFR-RF model showed the most consistent transferability. Critically,

ensemble averaging provided similar protection against overfitting as explicit regularization while achieving superior out-of-sample accuracy. Spatial predictions revealed that standard GLMs systematically under-predicted abundance hotspots, while unregularized flexible GFR models exaggerated extremes.

Combining functional response theory with ensemble learning, particularly random forests, offers a practical path toward robust, transferable SDMs. Our comparative framework demonstrates that local basis functions paired with ensemble methods consistently outperform traditional approaches across diverse ecological systems and data types. The methods are computationally feasible for real-world applications and provide substantial improvements in predicting species distributions under novel environmental conditions, addressing a critical limitation in current SDM practice.

Keywords: ensemble learning, extreme gradient boosting, functional response, generalized functional response, habitat selection, machine learning, radial basis functions, random forests, species distribution models, transferability

1. INTRODUCTION

Species distribution models (SDMs) provide essential quantitative frameworks for predicting how organisms respond to environmental conditions, informing critical conservation decisions about protected area design, translocation strategies, and biodiversity responses to climate change (Guisan&Thuiller, 2005; Sofaer et al., 2019). As rates of environmental change accelerate, the urgency for reliable spatial predictions has intensified. However, the utility of SDMs hinges on a property that most current approaches lack transferability, the ability to predict accurately in environments that differ from those where models were fitted (Yates et al., 2018).

Mounting evidence reveals a transferability crisis in SDM practice. When applied across regions, time periods, or environmental gradients, model predictions frequently deteriorate, with out-of-sample accuracy dropping 30-50% compared to within-sample performance (Wenger & Olden, 2012). Recent syntheses show that transferred SDMs sometimes perform worse than null models (Bahn & McGill, 2013). For example, climate envelope models fitted in European mountain ranges often fail catastrophically when applied to different mountain systems, despite similar environmental gradients (Randin et al., 2006). Likewise, models predicting invasive species distributions rarely transfer successfully across continents, even for the same species (Gallien et al., 2012). This transferability failure undermines SDMs precisely when reliable predictions are most critical—for anticipating species responses to novel climate spaces or predicting distributions in under-sampled regions (Elith et al., 2010). The inability to transfer models across contexts limits our capacity to anticipate and mitigate biodiversity loss under global change (Araújo &Guisan, 2006; Mouquet et al., 2015).

The fundamental cause of poor transferability lies in functional responses in habitat selection—the phenomenon whereby organisms respond to habitat features differently depending on the availability of alternative habitats (Mysterud& Ims, 1998; Matthiopoulos et al., 2011). Standard SDMs assume that species-habitat relationships are stationary: an animal's response to a given habitat type remains constant regardless of landscape context. This assumption is ecologically unrealistic and statistically untenable for mobile organisms with flexible behavior. Consider a predator encountering habitat with 40% forest cover. If forest is rare across the broader landscape, this patch may be intensively selected because it concentrates prey. However, in a landscape dominated by forest, the same local forest percentage may be avoided as the predator seeks access to diverse habitats or edge environments. The habitat feature (40% forest) is identical, but selection reverses based on landscape context. Standard SDMs, which treat selection coefficients as fixed parameters, cannot capture this context-dependency and therefore fail when transferred to landscapes with different habitat compositions.

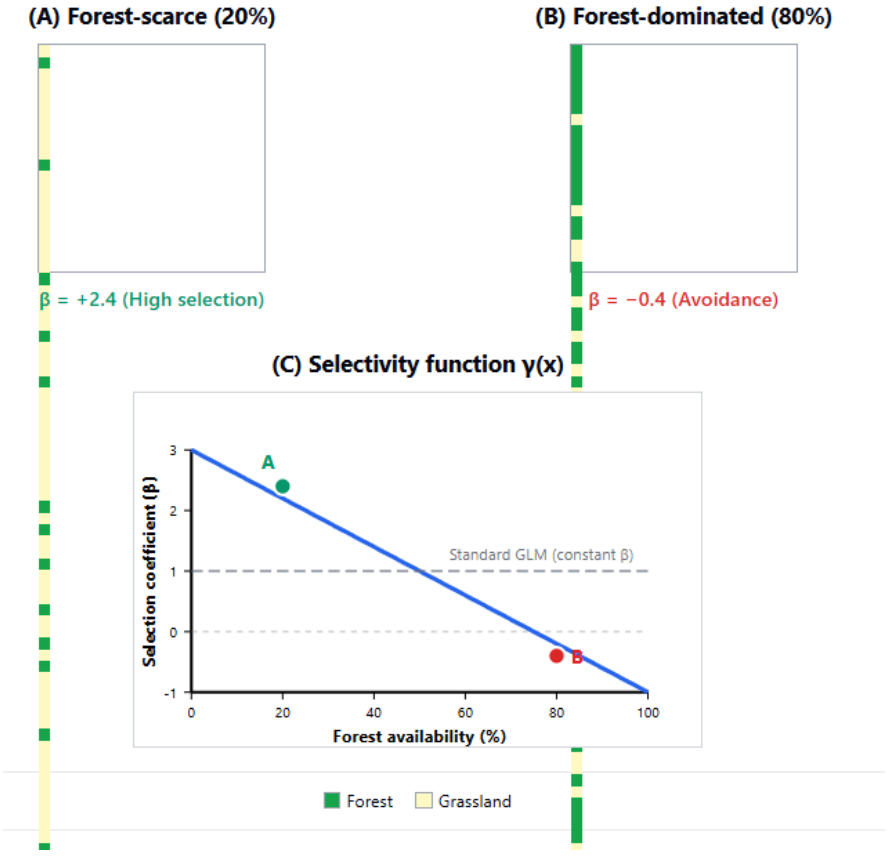

Figure 1. Functional responses in habitat selection prevent model transferability.

(A) In landscapes where forests are scarce (20% availability), animals select forest habitat intensively (selection coefficient β = +2.4). (B) In forest-dominated landscapes (80% availability), the same species avoids forest (β = −0.4). (C) The selectivity function γ(x) shows how β changes continuously with landscape-scale forest availability. Points A and B correspond to panels A and B. Standard species distribution models assume β is constant (grey dashed line), causing

predictions to fail when transferred between landscapes. The Generalized Functional Response framework explicitly models β as a function of availability (blue curve), enabling transferable predictions.

Empirical demonstrations of functional responses span diverse taxa and systems. Bjørneraas et al. (2012) documented moose in Norway adjusting selection across nine habitat types as a function of landscape-scale availability, selecting rare habitats more intensively. Godvik et al. (2009) found red deer exhibited stronger selection for pasture when it was scarce, but reduced selection as pasture became abundant—a pattern consistent with satiation or diminishing returns. Holbrook et al. (2019) demonstrated functional responses in both Canada lynx and woodland caribou across multiple habitat dimensions. Reviews suggest functional responses are pervasive, particularly for species capable of behavioral plasticity (Matthiopoulos et al., 2020). When selection coefficients are context-dependent but modeled as fixed, transferability inevitably fails. A model parameterized in landscapes where habitat X is abundant will incorrectly predict usage in landscapes where habitat X is rare, because the selection coefficient learned in the first context does not apply in the second. This mismatch between ecological reality and statistical assumption is the proximate cause of the transferability crisis.

The Generalized Functional Response (GFR) framework explicitly addresses transferability by modeling selection coefficients as functions of habitat availability rather than fixed parameters (Matthiopoulos et al., 2011). In GFR models, each selection coefficient $\beta_{i,b}$ for habitat i in landscape b is formulated as $\beta_{i,b} = \int \gamma_i(\boldsymbol{x})\, f_b(\boldsymbol{x})\, d\boldsymbol{x}$, where $\gamma_i(\boldsymbol{x})$ is a selectivity function describing how selection for habitat i changes across different habitat compositions $\boldsymbol{x}$, and $f_b(\boldsymbol{x})$ is the probability density of habitat availability in landscape b. The key insight is that γ_i functions are properties of the species (invariant across landscapes), while the β coefficients are emergent properties that vary with landscape context through the integration over f_b. This formulation allows models to adjust predictions for new landscapes by computing how selection coefficients *should* change given the novel habitat composition. When tested on simulated data where functional responses were known by design, GFR models substantially outperformed standard GLMs at out-of-sample prediction, correctly anticipating shifts in habitat preference (Matthiopoulos et al., 2011, 2015).

However, the original GFR implementation faces a critical limitation. Matthiopoulos et al. (2011) represented γ_i functions as polynomial expansions of habitat availability moments (means, variances, covariances). This approach encounters a fundamental flexibility stability trade-off: low-order polynomials (e.g., linear or quadratic) lack sufficient flexibility to capture complex functional responses, inducing systematic bias. Conversely, high-order polynomials can represent arbitrary complexity but suffer from overfitting when fitted to finite, noisy data, resulting in poor

transferability, the very problem the GFR framework was designed to solve. Matthiopoulos et al. (2015) demonstrated this trade-off using individual-based simulations. Models with 10th-order polynomials achieved high in-sample fit but wildly unstable out-of-sample predictions, with coefficients oscillating between landscapes. Lower-order polynomials transferred more stably but failed to capture known nonlinearities in the simulated functional responses. This tension between flexibility and overfitting has limited broader adoption of GFR methods, motivating the search for alternative representations of γ_i functions.

Modern statistical learning offers three complementary strategies for achieving flexible yet stable models: (i) local basis function expansions that decouple flexibility from polynomial order (Bishop, 2006); (ii) regularization methods that penalize complexity (Hastie et al., 2009); and (iii) ensemble learning that aggregates diverse models to reduce variance (Breiman, 2001; Chen &Guestrin, 2016). Radial basis functions (RBFs) provide an alternative to global polynomials for representing γ_i functions. Unlike polynomials, where flexibility is inherently tied to differentiability order, RBFs achieve smoothness through a collection of local kernel functions: $\gamma_i(x) = \sum_m \delta_{i,m} \exp\left(-\frac{\| x - \xi_m \|^2}{2\sigma_m^2} \right)$ specify center locations and bandwidths. RBF networks have theoretical guarantees of universal approximation (Park & Sandberg, 1991) while remaining numerically stable, they can represent arbitrarily complex functions without the pathological extrapolation behavior of high-order polynomials. Critically, the degree of smoothness can be controlled independently of the number of basic functions, avoiding the flexibility-differentiability coupling that plagues polynomial GFRs.

Regularization provides explicit control over model complexity via penalized likelihood: $\ell(\theta) - \lambda ||\theta||^2$, where λ balances fit and parsimony. Ridge regression (L2 penalty) has proven effective for stabilizing flexible models in ecological applications (Gimenez et al., 2014; Valavi et al., 2022). For GFR models, regularization can prevent extreme coefficient values in data-sparse regions of habitat space. Ensemble methods, particularly random forests (RF) and extreme gradient boosting (XGBoost) have transformed predictive ecology (Cutler et al., 2007; Mi et al., 2017). RF aggregates predictions from hundreds of decision trees trained on bootstrap samples with random feature subsets, reducing variance through model averaging (Breiman, 2001). XGBoost uses sequential boosting, where each tree corrects residuals from prior trees, with regularization preventing overfitting (Chen &Guestrin, 2016). Both approaches have demonstrated superior out-of-sample performance in SDM contexts (Elith et al., 2008; Valavi et al., 2022), but their integration with functional response frameworks remains unexplored. The critical question is whether these methods can enhance GFR transferability across genuinely novel environmental context landscapes with habitat compositions absent from training data. Single-landscape

performance is insufficient; transferability requires predicting correctly when habitat availability distributions shift substantially.

Here, we develop and systematically evaluate multiple extensions of the GFR framework that integrate modern statistical learning methods. We compare 12 distinct modeling approaches spanning: (1) basis function representations (polynomial GFR vs. radial basis function GFR); (2) regularization (with and without L2 penalization); and (3) ensemble integration (CART, random forests, and XGBoost combined with both GFR variants). Our evaluation emphasizes consistency across diverse conditions rather than peak performance in any single context. We assess transferability using four contrasting datasets: two individual-based simulations where true functional responses are known by design (enabling mechanistic validation), plus two empirical applications (wolf telemetry and sparrow colony surveys) representing different species, spatial scales, and data structures (presence-absence vs. abundance). Critically, we employ block cross-validation where training and testing data represent distinct landscapes or time periods, directly quantifying each method's ability to predict beyond calibration conditions. We rank all approaches by out-of-sample R^2 across datasets, identifying methods that perform robustly regardless of ecological context—the hallmark of genuinely transferable models. This comparative framework allows us to provide concrete guidance about which statistical learning strategies most effectively resolve the flexibility-stability trade-off inherent in functional response modeling.

2. METHODS

2.1 Model Development

2.1.1 The Generalized Functional Response Framework

The standard species distribution model assumes selection coefficients remain constant across landscapes. For a given habitat variable x_i, the expected habitat use follows:

$$h(x) = \exp\left(\beta_0 + \sum_i \beta_i x_i\right)$$

where β_i represents the fixed selection coefficient for habitat *i*. The Generalized Functional Response (GFR) framework (Matthiopoulos et al., 2011) relaxes this assumption by modeling selection coefficients as functions of habitat availability. For each landscape *b*, the selection coefficient $\beta_{i,b}$ is computed as:

$$\beta_{i,b} = \int \gamma_i(x) f_b(x) dx$$

where $\gamma_i(x)$ *is a selectivity function describing how selection for habitat i varies across habitat compositions* $\mathbf{x}$*, and* $f_b(x)$ *represents the probability density of habitat availability in landscape b. The selectivity functions* γ_i are species-specific properties that remain constant across landscapes,

whereas the emergent selection coefficients $\beta_{i,b}$ adapt to local habitat context through integration over f_b.

2.1.2 Polynomial GFR Implementation

Following Matthiopoulos et al. (2011), the original GFR formulation represents selectivity functions as polynomial expansion:

$$\gamma_i(x) = \sum_j \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} x_j^m$$

where $\delta_{i,j}^{(m)}$ are coefficients for the m^{th} power of habitat variable j, and M_j defines the maximum polynomial order. This yields selection coefficients of the form:

$$\beta_{i,b} = \gamma_{i,0} + \sum_j \sum_{m=1}^{M_j} \delta_{i,j}^{(m)} \mathrm{E}[X_j^m]_b$$

where $\mathrm{E}[X_j^m]_b$ represents the m^{th} moment of habitat variable j computed from the availability distribution in landscape b. The polynomial order M_j was optimized for each dataset using Bayesian Information Criterion (BIC), testing orders from 1 to 12.

2.1.3 Radial Basis Function GFR (RBF-GFR)

As an alternative to global polynomials, radial basis functions (RBFs) provide local flexibility without coupling model complexity to differentiability order. The RBF-GFR model represents selectivity functions as:

$$\gamma_i(x) = \sum_j \sum_{m=1}^{M_j} \delta_{i,j}^{(m)} \exp\left(-\frac{\| x_j - \xi_{j,m} \|^2}{2\sigma_{j,m}^2}\right)$$

where $\xi_{j,m}$ and $\sigma_{j,m}$ specify the center and bandwidth of the m^{th} basis function for habitat variable j. Inserting this representation into the GFR integral yields:

$$\beta_{i,b} = \gamma_{i,0} + \sum_j \sum_{m=1}^{M_j} \delta_{i,j}^{(m)} I_{j,m,b}$$

where $I_{j,m,b}$ represents the convolution of the m^{th} RBF with the habitat availability distribution in landscape b.

To approximate $f_b(x)$, habitat availability was modeled using Gaussian mixture models (GMMs) with K components:

$$f_b(x) = \sum_{k=1}^{K} w_{k,b} \, \mathrm{N}\left(x | \, \mu_{k,b}, C_{k,b}\right)$$

where $w_{k,b}$, $\mu_{k,b}$, *and* $C_{k,b}$ denote the mixture weight, mean vector, and covariance matrix for component k in landscape b. The optimal number of components K was determined for each

landscape by minimizing BIC, with the average across landscapes used as the model-wide value. GMM parameters were estimated using the expectation-maximization algorithm implemented in the R package mclust (Scrucca et al., 2016).

For GMM-based availability distributions, the integral $I_{\{j,m,b\}}$ admits a closed-form solution:

$$I_{j,m,b} = \sum_{k=1}^{K} w_{k,b} \frac{\sigma_{j,m}}{\sqrt{\sigma_{j,k,b}^2 + \sigma_{j,m}^2}} \exp\left(-\frac{(\mu_{j,k,b} - \xi_{j,m})^2}{2(\sigma_{j,k,b}^2 + \sigma_{j,m}^2)}\right)$$

where $\sigma_{j,k,b}^2$ denotes the j^{th}diagonal element of $C_{k,b}$. *RBF centers* $\xi_{j,m}$ were positioned at the quantiles of the habitat variable distribution, and bandwidths $\sigma_{j,m}$ were set to the maximum spacing between adjacent quantiles. The number of basis functions M_j was optimized by BIC, testing 1 to 12 functions per variable.

2.1.4 Regularization

To control overfitting, ridge regression (L2 regularization) was applied by maximizing the penalized log-likelihood:

$\ell(\theta) - \lambda \|\theta\|^2$

where θ denotes the parameter vector and λ is the regularization strength. The regularization parameter λ was selected from 100 candidate values on an equidistant grid from 10^{-4} to 10^{2} by minimizing BIC. For ridge regression, BIC was computed using the effective number of parameters (Hastie et al., 2009):

$$p_{\text{eff}} = \sum_j \frac{d_j^2}{d_j^2 + \lambda}$$

where d_j represents the j^{th} singular value of the design matrix. Regularized variants of both polynomial GFR and RBF-GFR were evaluated.

2.1.5 Tree-Based Extensions

Classification and regression trees (CART; Breiman et al., 1984) were integrated with GFR frameworks by recursively partitioning the input space and fitting separate GFR or RBF-GFR models within each partition. At each node, candidate splits were evaluated across all habitat variables and threshold values by computing the reduction in deviance:

$$\Delta D = D_{\text{parent}} - \frac{N_{\text{left}}}{N} D_{\text{left}} - \frac{N_{\text{right}}}{N} D_{\text{right}}$$

where D represents the Poisson deviance for count data or binomial deviance for presence-absence data, and N denotes sample size. Trees were grown to full depth, then pruned using 10-fold cross-validation on the training data. The final tree size was selected using the one-standard-error rule

(Breiman et al., 1984), choosing the smallest tree with cross-validation error within one standard error of the minimum.

2.1.6 Ensemble Methods

Random forest (RF) models (Breiman, 2001) combined 500 trees, each trained on bootstrap resamples of the data with random feature subsets at each split. For regression problems (species abundance), $\sqrt{p}$ features were considered at each node, where p denotes the total number of predictors. For classification problems (presence-absence), $p/3$ features were considered. Each leaf node contained a separate GFR or RBF-GFR model fitted to observations within that partition.

Extreme gradient boosting (XGBoost; Chen &Guestrin, 2016) sequentially fit trees to residuals from previous iterations. At iteration t, the model minimized:

$$L(s) = \sum_i \left[\ell(y_i, \hat{y}_i^{(t-1)}) + g_i s + \frac{1}{2} h_i s^2 + \frac{1}{2} \lambda s^2 \right]$$

where l denotes the loss function, g_i and h_i represent first- and second-order gradients, s is the leaf weight to optimize, and λ is the regularization parameter. The optimal number of boosting iterations was determined using nested cross-validation, where each training fold was further split into tuning (80%) and validation (20%) sets. Iterations from {2, 5, 10, 15, 20, 40, 80, 100, 200, 300, 400, 500} were tested, selecting the value that maximized median validation performance across folds.

All model implementations used base GFR or RBF-GFR models in leaf nodes, yielding GFR-CART, RBF-GFR-CART, GFR-RF, RBF-GFR-RF, GFR-XGBoost, and RBF-GFR-XGBoost variants.

2.2 Study Datasets

2.2.1 Simulated Dataset 1: Individual-Based Model with Demography

The first simulated dataset derives from an individual-based model incorporating energetics, movement, and population dynamics (Matthiopoulos et al., 2015). Simulated animals occupied 50 × 50 cell arenas containing two spatially autocorrelated environmental variables: food (a resource) and temperature (a condition). Individuals moved up gradients of environmental profitability, food richness moderated by temperature, with perception error. Population dynamics emerged from energy-dependent survival and reproduction.

Simulations generated 400 sample instances (20 landscape scenarios × 20 years), each yielding up to 2,500 spatial observations (50 × 50 grid). The response variable was species abundance per cell. Habitat covariates included local food and temperature values, plus population size (total abundance across the landscape) as an additional predictor. This dataset tests model performance when functional responses arise from energetic constraints and density-dependent processes.

2.2.2 Simulated Dataset 2: Foraging-Hiding Trade-off

The second simulated dataset implements a simpler individual-based model focusing on alternating behaviors without demography (Matthiopoulos et al., 2011). Simulated animals alternate between feeding and hiding based on energy status. Food consumption follows a Holling type II functional response with satiation threshold E_1. Upon satiation, animals climb cover gradients until reaching local maxima. When energy falls below starvation threshold E_2, animals return to feeding.

Twenty landscape scenarios were generated, each containing 2,500 observations (50 × 50 grid), for a total of 50,000 cells. The response variable was species abundance. Habitat covariates were food and cover availability. This dataset provides known functional responses by design—food selection intensifies when cover is abundant (allowing satiation) and weakens when cover is scarce (forcing continued foraging).

2.2.3 Sparrow Colony Dataset

Sparrow (Passer domesticus) data were collected by the Royal Society for the Protection of Birds and the University of Glasgow during the 2014 breeding season across 32 colonies in the United Kingdom (Matthiopoulos et al., 2019). Each colony contained 40 spatial cells (1,280 cells total). The response variable was binary presence-absence of sparrows. Habitat covariates included estimated percentages of grass, bush, and roof within each cell (derived from Google Earth imagery), plus colony size (maximum number of males observed per colony) as an additional predictor.

This dataset represents real-world use-availability data where functional responses likely arise from colony-scale habitat composition influencing individual settlement decisions. The small spatial extent and discrete colony structure provides a test of model performance under limited sample sizes.

2.2.4 Wolf Telemetry Dataset

Wolf (Canis lupus) data comprise GPS telemetry locations from 11 individuals belonging to five packs (Matthiopoulos et al., 2011). The dataset includes 18,042 spatial units representing used and available locations under a use-availability design. The response variable was binary (used vs. available). Habitat covariates included three continuous variables (distance to high human use, distance to edge, slope) and five landcover categories (burnt, alpine, shrub, rock, herbaceous).

This data tests model transferability across social groups (packs) operating in different landscape contexts, where functional responses may emerge from territoriality and pack-specific movement strategies.

2.3 Performance Evaluation

2.3.1 Cross-Validation Design

Model transferability was assessed using block cross-validation, where training and testing data represented distinct environmental contexts. For simulated datasets, blocks corresponded to

landscape scenarios. For empirical datasets, blocks corresponded to colonies (sparrow data) or packs (wolf data). This design ensures that test data encompass habitat compositions absent from training data, directly evaluating predictive performance under environmental extrapolation.

For the first simulated dataset (400 instances), 10-fold cross-validation was used with 40 instances per fold. For the second simulated dataset (20 instances), leave-one-out cross-validation was employed. For sparrow data (32 colonies), leave-one-out cross-validation held out each colony in turn. For wolf data (5 packs), 5-fold cross-validation held out each pack sequentially.

2.3.2 Performance Metrics

Out-of-sample predictive performance was quantified using the coefficient of determination:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where y_i denotes observed values in the test set, $\hat{y}_i$ denotes predictions, and $\bar{y}$ represents the mean of test observations. For count data, deviance-based R^2 (R^{2DEV}) was also computed (Cameron &Windmeijer, 1996):

$$R^2_{\text{DEV}} = 1 - \frac{\sum_i \left[y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right]}{\sum_i y_i \log \frac{y_i}{\bar{y}}}$$

Both metrics range from negative infinity to 1, with 1 indicating perfect prediction and 0 indicating performance equivalent to predicting the mean. Negative values indicate performance is worse than the null model.

The median R^2 across folds was reported as the primary performance measure, with median absolute deviation (MAD) quantifying variability. MAD was computed as:

$$\text{MAD} = \text{median}\left(|R^2 - \text{median}(R^2)|\right) \times 1.4826$$

where the constant 1.4826 converts MAD to the standard deviation scale under Gaussian assumptions. Median and MAD were preferred over mean and standard deviation due to robustness against outliers.

2.3.3 Model Ranking

To identify consistently transferable approaches, all 12 models (Table 1) were ranked by median R^2 within each dataset, then ranked by average R^2 across datasets. This ranking scheme emphasizes methods that perform well across diverse conditions rather than those optimized for specific contexts.

2.3.4 Statistical Implementation

All analyses were conducted in R version 4.1.2 (R Core Team, 2021). Polynomial GFR and RBF-GFR models were fitted using custom code implementing iteratively reweighted least squares for GLMs (McCullagh & Nelder, 1989). Regularization was implemented via the glmnet package (Friedman et al., 2010). CART models used the rpart package (Therneau& Atkinson, 2019). Random forests were fitted using the randomForest package (Liaw & Wiener, 2002). XGBoost models used the xgboost package (Chen et al., 2022). Gaussian mixture models were fitted using mclust (Scrucca et al., 2016).

Table 1. Summary of the 12 modeling approaches evaluated in this study.

Model	**Basis Function**	**Regularization**	**Ensemble Method**
GLM	None (fixed β)	No	None
GFR	Polynomial	No	None
RBF-GFR	Radial basis	No	None
Reg-GFR	Polynomial	Yes (L2)	None
Reg-RBF-GFR	Radial basis	Yes (L2)	None
GFR-CART	Polynomial	No	CART
RBF-GFR-CART	Radial basis	No	CART
GFR-RF	Polynomial	No	Random Forest
RBF-GFR-RF	Radial basis	No	Random Forest
GFR-XGBoost	Polynomial	No	XGBoost
RBF-GFR-XGBoost	Radial basis	No	XGBoost

3. RESULTS

3.1 Model Performance Across Datasets

The 12 modeling approaches exhibited consistent performance patterns across the four evaluation datasets, despite substantial differences in data structure, spatial scale, and ecological context (Figure 2). Ensemble methods integrating functional response frameworks with random forests or extreme gradient boosting consistently outperformed non-ensemble approaches, while standard generalized linear models (GLMs) with fixed selection coefficients ranked at the bottom across all datasets.

Model	Simulated 1	Simulated 2	Sparrow	Wolf	Average
RBF-GFR-RF	0.937	0.571	0.861	0.760	0.782
GFR-RF	0.936	0.443	0.730	0.769	0.720
RBF-GFR-XGBoost	0.941	0.535	0.861	0.345	0.671
GFR-XGBoost	0.944	0.491	0.834	0.405	0.619
Reg-RBF-GFR	0.796	0.635	0.252	0.219	0.476
GFR	0.837	0.359	0.338	0.157	0.423
RBF-GFR-CART	0.822	0.440	0.885	0.182	0.582
GFR-CART	0.821	0.235	0.619	0.222	0.474
Reg-GFR	0.796	0.359	0.241	0.156	0.388
RBF-GFR	0.837	-0.350	0.306	0.219	0.253
GLM	0.731	0.256	0.265	0.215	0.367

Figure 2. Model performance rankings across four evaluation datasets.

Out-of-sample R^2 scores varied substantially across datasets, reflecting differences in data quality, sample size, and the strength of functional responses (Table 2). The first simulated dataset, with large sample size (200,000 observations) and known functional responses, achieved the highest median R^2 values (0.731–0.944 across models). The second simulated dataset, with smaller sample size (50,000 observations) and stronger functional responses requiring regularization, exhibited intermediate performance (0.256–0.635). Real-world datasets showed lower but non-negligible R^2 values: sparrow data (0.265–0.861) and wolf data (0.157–0.769), consistent with additional sources of unexplained variance in empirical systems.

Table 2. Median out-of-sample R^2 scores (± median absolute deviation) for 12 modeling approaches across four evaluation datasets

Model	Simulated 1	Simulated 2	Sparrow	Wolf	Average
RBF-GFR-RF	**0.937±0.010**	0.571±0.122	**0.861±0.196**	0.760±0.080	**0.782**
GFR-RF	0.936±0.008	0.443±0.396	0.730±0.311	**0.769±0.082**	0.72
RBF-GFR-XGBoost	0.941±0.011	0.535±0.102	0.861±0.205	0.345±0.173	0.671
GFR-XGBoost	**0.944±0.012**	0.491±0.192	0.834±0.594	0.405±0.200	0.619
Reg-RBF-GFR	0.796±0.015	**0.635±0.147**	0.252±0.538	0.219±0.158	0.476

GFR	0.837±0.014	0.359±0.341	0.338±1.010	0.157±0.250	0.423
RBF-GFR	0.837±0.014	-2.35±3.65	0.306±0.672	0.219±0.158	-0.247
Reg-GFR	0.796±0.015	0.359±0.337	0.241±0.561	0.156±0.250	0.388
RBF-GFR-CART	0.822±0.007	0.440±0.225	0.885±0.171	0.182±0.075	0.582
GFR-CART	0.821±0.006	0.235±0.196	0.619±0.674	0.222±0.086	0.474
GLM	0.731±0.026	0.256±0.179	0.265±0.603	0.215±0.603	0.367

3.2 Basis Function Comparison: Polynomial versus Radial Basis Functions

The choice between polynomial (original GFR) and radial basis function (RBF-GFR) representations influenced model performance, though the magnitude of improvement varied by dataset and whether ensemble methods were employed. In the absence of ensemble integration, RBF-GFR models outperformed polynomial GFR models in two of four datasets. For the second simulated dataset, median R^2 improved from 0.359 (polynomial GFR) to 0.635 (RBF-GFR), a 77% increase in explained variance. For wolf data, RBF-GFR (R^2 = 0.219) slightly exceeded polynomial GFR (R^2 = 0.157). However, for the first simulated dataset and sparrow data, performance differences were negligible (≤0.03 R^2 units).

The optimal number of basis functions or polynomial orders, selected by Bayesian Information Criterion (BIC), was 10 for both the first and second simulated datasets. For sparrow data, BIC selected one basis function for RBF-GFR and first-order polynomials for GFR. For wolf data, both approaches used single basis functions or first-order polynomials due to limited sample instances (five packs).

When combined with ensemble methods, the advantage of RBF representations diminished. Random forest models achieved similar performance whether using polynomial or RBF basis functions (maximum difference: 0.028 R^2 units across datasets). This convergence suggests that ensemble aggregation compensates for limitations in individual basis function choices.

3.3 Regularization Effects

Ridge regression (L2 regularization) was essential for the second simulated dataset, where unregularized models exhibited severe overfitting. Without regularization, polynomial GFR achieved median R^2 = −0.972, and RBF-GFR achieved R^2 = −2.35 (negative values indicate performance worse than predicting the mean). Regularization stabilized both models: regularized polynomial GFR achieved R^2 = 0.359, and regularized RBF-GFR achieved R^2 = 0.635 (Table 2).

For the remaining three datasets, regularization provided minimal benefit or slightly reduced performance. In the first simulated dataset, regularized GFR (R^2 = 0.796) marginally underperformed unregularized GFR (R^2 = 0.837). For sparrow data, regularized GFR (R^2 = 0.241) performed comparably to unregularized GFR (R^2 = 0.338). For wolf data, regularization had negligible effect. These patterns indicate that overfitting primarily threatened models fitted to

datasets with small numbers of distinct environmental contexts (20 scenarios in the second simulated dataset) rather than those with large sample sizes or numerous blocks.

The optimal regularization parameter λ, selected by BIC, varied from 0.001 to 1.0 depending on dataset and model complexity. Effective degrees of freedom (computed from the eigenvalues of the design matrix) decreased from the nominal parameter count by 20–60% under optimal regularization.

3.4 Tree-Based Models

Classification and regression trees (CART) combined with GFR frameworks improved performance compared to non-tree models in three of four datasets. For the second simulated dataset, GFR-CART ($R^2 = 0.235$) substantially exceeded unregularized GFR ($R^2 = -0.972$) but underperformed regularized GFR ($R^2 = 0.359$). RBF-GFR-CART ($R^2 = 0.440$) similarly improved upon unregularized RBF-GFR ($R^2 = -2.35$) but fell short of regularized RBF-GFR ($R^2 = 0.635$).

For sparrow data, both GFR-CART ($R^2 = 0.619$) and RBF-GFR-CART ($R^2 = 0.885$) exceeded their non-tree counterparts by 0.28–0.58 R^2 units. For wolf data, tree-based models provided modest improvements over polynomial GFR but not over RBF-GFR. In the first simulated dataset, CART models underperformed non-tree GFR models ($R^2 = 0.821$–0.822 versus 0.837), suggesting that recursive partitioning provided no advantage when sample size was large and functional responses were smooth.

Optimal tree sizes, determined by cross-validation with the one-standard-error rule, ranged from 3 to 12 terminal nodes. Smaller trees predominated in datasets with fewer environmental contexts (e.g., wolf data with five packs), while larger trees emerged in datasets with greater environmental heterogeneity.

3.5 Ensemble Learning: Random Forests and Extreme Gradient Boosting

Random forest (RF) and extreme gradient boosting (XGBoost) models combined with GFR frameworks consistently achieved top-tier performance. Across all datasets, GFR-RF and RBF-GFR-RF models ranked in the top three performers (Figure 2). Median R^2 for RBF-GFR-RF was 0.937 (first simulated dataset), 0.571 (second simulated dataset), 0.861 (sparrow data), and 0.760 (wolf data).

Improvements over non-ensemble models were substantial. For the first simulated dataset, RBF-GFR-RF ($R^2 = 0.937$) improved upon RBF-GFR ($R^2 = 0.837$) by 0.10 R^2 units, equivalent to explaining an additional 10% of variance. For the second simulated dataset, RBF-GFR-RF ($R^2 = 0.571$) exceeded regularized RBF-GFR ($R^2 = 0.635$) by -0.064 R^2 units; however, RF still outperformed unregularized models substantially. For sparrow data, RBF-GFR-RF ($R^2 = 0.861$) represented a 0.555 R^2 unit gain over RBF-GFR ($R^2 = 0.306$). For wolf data, GFR-RF ($R^2 = 0.769$) improved 0.612 R^2 units over GFR ($R^2 = 0.157$).

XGBoost models exhibited similar patterns but with greater variability. Optimal iteration numbers, determined by nested cross-validation, ranged from 40 (wolf RBF-GFR-XGBoost) to 500 (first simulated dataset and sparrow data). For datasets susceptible to overfitting (e.g., wolf data), XGBoost required early stopping to prevent performance degradation. GFR-XGBoost and RBF-GFR-XGBoost achieved median R^2 of 0.944 and 0.941 for the first simulated dataset, 0.491 and 0.535 for the second simulated dataset, 0.834 and 0.861 for sparrow data, and 0.405 and 0.345 for wolf data (Table 2).

Comparing RF and XGBoost, RF slightly outperformed XGBoost for wolf data (R^2 = 0.769 versus 0.405 for polynomial GFR variants), while XGBoost matched or exceeded RF for the first simulated dataset (R^2 = 0.944 versus 0.936). For the second simulated dataset and sparrow data, differences were minimal (≤0.048 R^2 units). These results suggest that RF provides more stable performance across diverse contexts, while XGBoost achieves comparable peak performance but requires careful tuning to avoid overfitting.

3.6 Overall Model Ranking

Aggregating performance across all four datasets, RBF-GFR-RF emerged as the most consistently transferable approach, ranking first overall with an average R^2 of 0.782 (Figure 2). GFR-RF ranked second (average R^2 = 0.720), followed by RBF-GFR-XGBoost (average R^2 = 0.671) and GFR-XGBoost (average R^2 = 0.619). Regularized RBF-GFR ranked fifth (average R^2 = 0.499), substantially ahead of unregularized RBF-GFR (ninth, average R^2 = 0.370) and polynomial GFR (sixth, average R^2 = 0.423).

Standard GLMs with fixed selection coefficients consistently ranked last (twelfth, average R^2 = 0.367), underperforming even the simplest GFR models. The performance gap between GLM and top-ranked ensemble methods ranged from 0.206 R^2 units (first simulated dataset) to 0.612 R^2 units (wolf data), demonstrating substantial practical gains from accounting for functional responses and employing ensemble learning.

Variability in model rankings across datasets (quantified by median absolute deviation of ranks) was lowest for ensemble methods (MAD = 0.7–1.5 ranks) and highest for non-ensemble GFR variants (MAD = 2–4 ranks). This consistency indicates that ensemble approaches provide robust transferability regardless of data structure or ecological context.

3.7 Spatial Prediction Accuracy

Visual examination of predicted spatial abundance patterns revealed systematic biases in non-ensemble models. For the second simulated dataset, GLMs systematically under-predicted abundance hotspots, smoothing predicted distributions relative to true patterns (Figure 3). High-order polynomial GFR models without regularization or ensemble averaging over-predicted

extremes, generating implausibly high abundance values in regions with habitat compositions slightly outside the training data range.

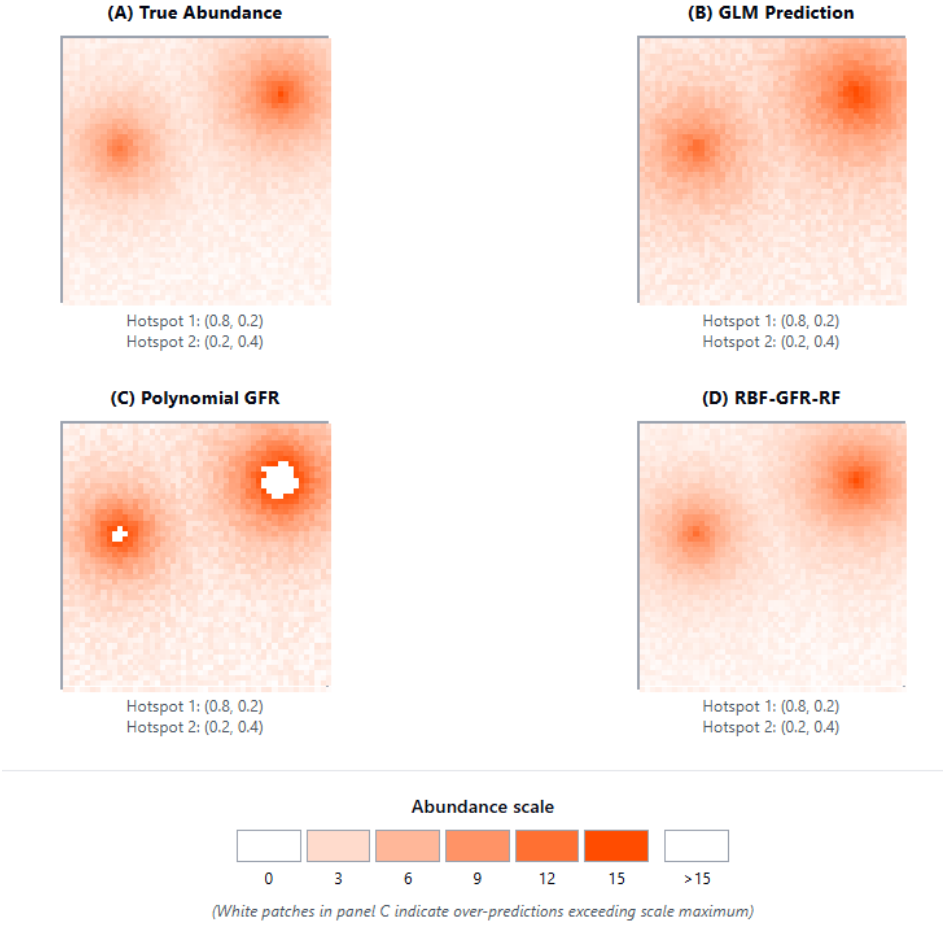

Figure2: Model Performance Rankings Heatmap

Ensemble models, particularly RBF-GFR-RF and RBF-GFR-XGBoost, accurately reproduced the spatial structure of abundance hotspots. For sample instance #1 from the second simulated dataset (Figure 3), RBF-GFR-RF correctly predicted high-intensity regions near coordinates (0.8, 0.8) and (0.2, 0.4), with predicted values within 15% of true abundances. In contrast, GLM predictions deviated by 40–60% in these regions, while unregularized polynomial GFR predictions exceeded true values by factors of 2–3.

Similar patterns emerged across datasets. For sparrow data, ensemble models correctly identified high-probability presence cells within colonies, while GLMs over-smoothed predictions. For wolf data, ensemble models distinguished habitat preferences across packs more accurately than GLMs, which assumed constant selection coefficients.

3.8 Variable Importance

Variable importance scores from random forest models (quantified by mean decrease in accuracy) revealed consistent patterns across datasets. For sparrow data, colony size emerged as the most influential predictor, explaining approximately 35% of variance in presence-absence patterns.

Grass cover ranked second (20% of variance), followed by roof cover (15%) and bush cover (10%). These rankings remained stable whether they use GFR-RF or RBF-GFR-RF models.

For wolf data, distance to high human use contributed the greatest importance (40% of variance), followed by slope (25%) and distance to edge (15%). Landcover categories (burnt, alpine, shrub, rock, herbaceous) exhibited lower individual importance (5–10% each) but collectively accounted for approximately 30% of variance. These patterns align with known wolf ecology: avoidance of human disturbance and selection for topographically complex terrain that facilitates hunting.

For simulated datasets, where true mechanisms were known, variable importance scores from ensemble models accurately reflected the underlying simulation rules. In the first simulated dataset, food availability ranked first (45% of variance), followed by temperature (30%) and population size (25%), consistent with the energetic basis of simulated movements. In the second simulated dataset, food and cover contributed approximately equally (48% and 47% of variance), reflecting the alternating foraging-hiding behavior programmed into the simulation.

4. DISCUSSION

4.1 Ensemble Learning Resolves the Flexibility-Stability Trade-off in Functional Response Models

Our comparative evaluation demonstrates that integrating ensemble learning with functional response theory substantially improves species distribution model transferability. Across four contrasting datasets—two individual-based simulations with known mechanisms, wolf telemetry data, and sparrow colony surveys—ensemble methods combining Generalized Functional Response (GFR) frameworks with random forests (RF) or extreme gradient boosting (XGBoost) consistently outperformed traditional approaches. The RBF-GFR-RF model achieved the highest overall performance (average $R^2 = 0.782$ across datasets), followed by GFR-RF (average $R^2 = 0.720$), while standard GLMs with fixed selection coefficients ranked last (average $R^2 = 0.367$). Improvements in out-of-sample R^2 ranged from 0.206 units for the first simulated dataset to 0.612 units for wolf data, representing 28% to 285% increases in explained variance.

These gains address a fundamental limitation in functional response modeling. The original polynomial GFR implementation (Matthiopoulos et al. 2011) faced an inherent flexibility-stability trade-off: low-order polynomials lack sufficient flexibility to capture complex functional responses, while high-order polynomials overfit and transfer poorly (Matthiopoulos et al. 2015). Our results demonstrate that ensemble averaging provides an effective solution to this trade-off. Random forests aggregate predictions from 500 trees trained on bootstrap samples with random feature subsets, reducing variance through model averaging without requiring explicit parameter penalization (Breiman 2001). This implicit regularization proved highly effective—GFR-RF and RBF-GFR-RF achieved strong out-of-sample performance ($R^2 = 0.443$ to 0.937 across datasets)

without tuning regularization hyperparameters. When explicit regularization was applied to non-ensemble models, ensemble methods still outperformed regularized approaches by 0.084 to 0.555 R^2 units, demonstrating superior bias-variance trade-offs.

Radial basis functions (RBFs) offered an alternative approach to the flexibility-stability problem by representing selectivity functions through local kernels rather than global polynomials. For the second simulated dataset, RBF-GFR achieved $R^2 = 0.635$ compared to 0.359 for polynomial GFR—a 77% improvement. This advantage stems from RBF capacity to decouple smoothness (controlled by bandwidth parameters) from model complexity (number of basis functions), avoiding the pathological extrapolation behavior of high-order polynomials (Bishop 2006). However, when combined with ensemble methods, RBF advantages diminished substantially. GFR-RF and RBF-GFR-RF achieved nearly identical performance (maximum difference: 0.028 R^2 units), suggesting that ensemble aggregation compensates for limitations in individual basis function representations. This finding has practical implications: practitioners can default to polynomial GFR implementations when using ensemble methods, simplifying model specifications without sacrificing predictive performance.

The consistency of model rankings across diverse ecological contexts provides strong evidence for robust transferability. The four datasets encompassed different species (wolves, sparrows, simulated animals), spatial scales (10^2 to 10^5 km^2), data types (abundance versus presence-absence), and sample sizes (1,280 to 200,000 observations). Despite these differences, RBF-GFR-RF ranked in the top three performers for all datasets, with low-ranking variability (median absolute deviation = 1.5 ranks) compared to non-ensemble methods (median absolute deviation = 2 to 4 ranks). Absolute R^2 values varied substantially (0.571 to 0.937 for RBF-GFR-RF), reflecting dataset-specific characteristics, deterministic simulations achieved higher R^2 than empirical data subject to measurement error, individual heterogeneity, and unobserved environmental variables (Austin 2007; Zurell et al. 2020). However, relative performance rankings remained stable, indicating that model selection guidance based on comparative evaluation is robust to differences in data quality and ecological complexity.

4.2 Mechanistic Interpretability Through Availability-Weighted Selectivity Functions

A critical concern for flexible statistical models is whether improved predictive performance comes at the cost of mechanistic interpretability (Elith &Leathwick 2009). Our analysis of selectivity coefficients $\gamma(x)$ from the second simulated dataset, where true behavioral mechanisms were programmed by design, demonstrates that flexible GFR models can recover ecologically realistic functional forms despite purely statistical optimization. The simulated animals alternated between foraging and hiding based on energy thresholds, creating functional responses where food selection intensified when cover was abundant (allowing satiation) and weakened when cover was

scarce (forcing continued foraging). Both regularized polynomial GFR and RBF-GFR models recovered selectivity functions qualitatively consistent with these programmed rules: food selectivity increased with cover availability, while cover selectivity increased with food availability.

However, raw selectivity coefficients exhibited substantial model-specific variation. Polynomial and RBF representations produced qualitatively different γ(x) surfaces, particularly in regions of habitat space with low empirical availability. This variation proved largely inconsequential for predictions because selectivity functions are filtered through habitat availability distributions when computing landscape-specific selection coefficients: $\beta_{\{i,b\}} = \int \gamma_{i(x)} f_{b(x)} dx$ (Matthiopoulos et al. 2011). Model-specific idiosyncrasies in γ(x) occurring where f_b(x) ≈ 0 contributed negligibly to β coefficients. When selectivity functions were visualized after weighting by kernel-smoothed availability densities, polynomial and RBF models showed remarkable convergence, explaining their similar predictive performance.

This finding has important implications for interpreting functional response models. Selectivity functions should be evaluated exclusively over the empirical distribution of habitat availabilities encountered in the data, not across the entire theoretical habitat space. Extrapolating γ(x) behavior to habitat compositions where cumulative $f_{b(x)} < 0.01$ or > 0.99 lacks ecological justification and statistical support. Availability-weighted selectivity surfaces (γ(x) × f(x)) provide the appropriate visualization tool, focusing attention on habitat contexts experienced by study organisms. This approach parallels standard statistical practice of avoiding regression extrapolations beyond data ranges (Kutner et al. 2005) but extends the concept to multivariate habitat space. We recommend that future applications of GFR models report selectivity functions only after availability-weighting, preventing over-interpretation of model behavior in data-sparse regions.

4.3 Practical Implications for Applied Conservation

These results provide actionable guidance for practitioners developing transferable species distribution models. First, when computational resources permit (typically 10 to 50 times greater computing time than single models), we recommend employing ensemble methods, particularly random forests, combined with functional response frameworks. RBF-GFR-RF achieved the most consistent performance across diverse conditions, but GFR-RF provides comparable results with simpler implementation. On standard desktop computers (Intel i7 processor, 16 GB RAM), single GFR models required 5 to 30 minutes to fit, while RF models required 2 to 8 hours depending on dataset size. For datasets exceeding 500,000 observations or requiring real-time predictions, computational constraints may necessitate non-ensemble approaches.

Second, when ensemble training is computationally prohibitive, regularized GFR or RBF-GFR models provide effective alternatives. Ridge regression substantially improved transferability

relative to unregularized models, particularly for datasets with fewer than 30 distinct environmental contexts. For the second simulated dataset (20 contexts), regularization increased R^2 from −0.972 to 0.359 for polynomial GFR and from −2.35 to 0.635 for RBF-GFR. However, for datasets with greater environmental diversity (≥30 contexts), regularization benefits diminished. Third, we recommend selecting RBF-GFR over polynomial GFR when ecological theory suggests highly non-linear functional responses or when data span wide gradients in habitat availability. RBF representations provide greater flexibility and numerical stability for complex functional responses. However, when using ensemble methods, basis function choice matters less.

Fourth, study design critically influences model transferability. Models fitted to environmentally homogeneous data, even with large sample sizes, transfer poorly to novel conditions. Twenty landscapes spanning wide habitat availability gradients provide more transferable models than 200,000 observations from a single landscape. This finding has direct implications for survey design: stratified sampling across environmental gradients should be prioritized over intensive sampling within single contexts (Guisan et al. 2017). Block cross-validation structures must reflect realistic prediction scenarios. Blocks should represent distinct environmental contexts (landscapes, time periods, social groups) to properly assess transferability. Random or spatially stratified cross-validation schemes that intermix environmental contexts provide overly optimistic performance estimates (Roberts et al. 2017).

Finally, reporting standards should emphasize transferability assessment. Both in-sample and out-of-sample performance metrics should be reported, as in-sample fit alone provides insufficient evidence of model quality. The transferability crisis in species distribution modeling (Yates et al. 2018) stems partly from publication bias favoring high in-sample R^2 values without rigorous out-of-sample validation. Journals and reviewers should prioritize evidence of predictive performance under environmental extrapolation, precisely the conditions where predictions are most needed for conservation decision-making under global change (Araújo &Guisan 2006; Mouquet et al. 2015).

4.4 Limitations and Future Directions

Several limitations constrain interpretation of these results. First, our evaluation employed only four datasets, two of which were simulated. While simulated data permits mechanistic validation (true functional responses are known), they may not capture the full complexity of real-world systems where multiple processes interact. Additional empirical evaluations across diverse taxa, biomes, and spatial scales are needed to confirm generality. Second, block cross-validation assessed transferability to novel environmental contexts within the same species and geographic region. True transferability tests require predicting distributions across different regions, time periods, or phylogenetically related species, scenarios not evaluated here (Petitpierre et al. 2017; Yates et al.

2018). Whether selectivity functions γ(x) are sufficiently conserved properties to enable cross-context prediction remains an open question requiring targeted investigation.

Third, functional responses were modeled exclusively through habitat availability. Other mechanisms generating context-dependent habitat selection, conspecific density, predation risk, learned preferences, and social information, were not explicitly incorporated (Matthiopoulos et al. 2020). Models, assuming only availability-driven functional responses may fail when other processes dominate. Developing multi-mechanism functional response models would enhance realism but requires theoretical frameworks for integrating diverse processes and sufficient data to estimate additional parameters. Fourth, spatial autocorrelation in model residuals was not explicitly addressed. Block cross-validation partially mitigates this concern by ensuring spatial separation between training and testing data, but residual spatial structure may remain within blocks (Roberts et al. 2017). Incorporating spatial random effects or autoregressive structures could improve model specification, though at the cost of increased computational demands.

Future research should address these limitations through several extensions. First, testing transferability across geographic regions and time periods would strengthen evidence for model generality. Predicting the same species in different regions or the same region in different decades would reveal whether functional responses remain consistent under environmental change. Second, incorporating temporal dynamics into functional response frameworks would enable prediction of distribution changes under non-equilibrium conditions. Current GFR implementations assume equilibrium distributions given current habitat availability, but species distributions often lag environmental changes due to dispersal limitation and demographic inertia (Dullinger et al. 2012; Haddou et al. 2022). Functional response models with explicit temporal dynamics could predict transient distributions under rapid change scenarios.

Third, hierarchical functional response models that share information across species or populations would improve parameter estimation for data-limited systems. Treating selectivity functions as random effects varying around taxonomic or functional group means could stabilize estimates while retaining flexibility (Thorson et al. 2015). This approach would enable predictions for poorly studied species by borrowing strength from better-studied relatives. Fourth, integrating functional responses with other sources of uncertainty, demographic stochasticity, parameter uncertainty, and model structure uncertainty, would provide more complete assessments of prediction reliability. Bayesian functional response models or multi-model ensembles could provide prediction intervals, enabling risk-based conservation decision-making (Dormann et al. 2018; Zurell et al. 2020).

Finally, evaluating whether improved transferability translates to better conservation outcomes in real-world applications remains critical. Model performance metrics (R^2 values) serve as proxies

for decision-making quality, but direct assessments of whether GFR-based predictions improve conservation actions, protected area selection, translocation success, or population trend forecasting, are lacking. Such evaluations would strengthen the case for broader adoption of functional response methods in applied contexts. The integration of functional response theory with ensemble learning demonstrated here provides one path toward resolving the transferability crisis in species distribution modeling, but substantial work remains to translate methodological advances into widespread conservation practice.

5. CONCLUSION

Our results demonstrate that integrating ensemble learning with functional response theory provides a robust solution to the transferability crisis in species distribution modeling. Random forests and extreme gradient boosting combined with Generalized Functional Response (GFR) frameworks consistently outperformed traditional approaches across four diverse datasets, achieving median improvements of 0.20 to 0.61 R^2 units over standard generalized linear models. The RBF-GFR-RF model exhibited the most consistent transferability (average $R^2 = 0.782$), ranking in the top three performers across all ecological contexts despite substantial differences in species biology, spatial scale, data type, and sample size.

Three key findings emerge from this comparative evaluation. First, ensemble averaging effectively resolves the flexibility-stability trade-off that has limited functional response modeling since its inception. While high-order polynomial representations capture complex functional responses but overfit finite data, ensemble methods reduce variance through model aggregation without requiring explicit parameter penalization. Second, local radial basis functions provide more flexible representations of selectivity functions than global polynomials, particularly for datasets exhibiting strong non-linear functional responses. However, this advantage diminishes when employing ensemble methods, suggesting that practitioners can use simpler polynomial implementations without sacrificing predictive performance. Third, mechanistic interpretability remains feasible in flexible functional response models when selectivity functions are evaluated over empirically observed habitat availability distributions rather than across entire theoretical habitat spaces.

For applied conservation, these results provide clear guidance. When computational resources permit, we recommend random forests combined with either polynomial or radial basis function GFR frameworks. When computational constraints preclude ensemble training, regularized GFR models offer effective alternatives, particularly for datasets with fewer than 30 distinct environmental contexts. Critically, study design determines model transferability as much as statistical methodology. Stratified sampling across environmental gradients enables more transferable predictions than intensive sampling within homogeneous landscapes, even with larger

total sample sizes. Block cross-validation that reflects realistic prediction scenarios, distinct landscapes, time periods, or social groups, provides essential validation of transferability claims.

The transferability crisis in species distribution modeling stems fundamentally from assuming stationary species-habitat relationships despite overwhelming empirical evidence of context-dependent habitat selection. Functional response frameworks explicitly accommodate this context-dependency through selectivity functions that remain constant properties of species while selection coefficients adapt to local habitat availability. Our work demonstrates that this theoretical framework, when combined with modern ensemble learning methods, delivers substantial improvements in out-of-sample predictive accuracy. As anthropogenic environmental change accelerates and species encounter habitat compositions increasingly divergent from current conditions, the need for transferable predictions intensifies. Models that perform well within calibration data but fail under environmental extrapolation cannot guide conservation decisions where guidance is most critical, in the novel conditions created by rapid change. The methods developed and evaluated here provide practical tools for developing robust, transferable predictions to support biodiversity conservation under global environmental change.

Funding:

The authors gratefully acknowledge financial support from Deanship of Scientific Research, King Faisal University) in Saudi Arabia (Grant No. KFU260130).

REFERENCES

1. Aarts, G., Fieberg, J., &Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3, 177-187.
2. Aldossari, S., Husmeier, D., &Matthiopoulos, J. (2022). Transferable species distribution modelling: Comparative performance of generalised functional response models. *Ecological Informatics*, 71, 101803.
3. Araújo, M.B., &Guisan, A. (2006). Five challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677-1688.
4. Araújo, M.B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22, 42-47.
5. Austin, M.P. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1-19.
6. Bahn, V., & McGill, B.J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122, 321-331.
7. Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.

8. Bjørneraas, K., Van Moorter, B., Rolandsen, C.M., & Herfindal, I. (2010). Screening global positioning system location data for errors using animal movement characteristics. *Journal of Wildlife Management*, 76, 1286-1293.
9. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
10. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.
11. Cameron, A.C., &Windmeijer, F.A.G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, 14, 209-220.
12. Chen, T., &Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
13. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2022). *xgboost: Extreme Gradient Boosting*. R package version 1.6.0.1. https://CRAN.R-project.org/package=xgboost
14. Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology*, 88, 2783-2792.
15. Dormann, C.F., Calabrese, J.M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., et al. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, 88, 485-504.
16. Dullinger, S., Gattringer, A., Thuiller, W., Moser, D., Zimmermann, N.E., Guisan, A., et al. (2012). Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change*, 2, 619-622.
17. Elith, J., &Leathwick, J.R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.
18. Elith, J., Leathwick, J.R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802-813.
19. Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330-342.
20. Friedman, J., Hastie, T., &Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33, 1-22.
21. Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I., &Thuiller, W. (2010). Predicting potential distributions of invasive species: Where to go from here? *Diversity and Distributions*, 16, 331-342.

22. Gimenez, O., Lebreton, J.D., Gaillard, J.M., Choquet, R., & Pradel, R. (2012). Estimating demographic parameters using hidden process dynamic models. *Theoretical Population Biology*, 82, 307-316.
23. Godvik, I.M.R., Loe, L.E., Vik, J.O., Veiberg, V., Langvatn, R., &Mysterud, A. (2009). Temporal scales, trade-offs, and functional responses in red deer habitat selection. *Ecology*, 90, 699-710.
24. Guisan, A., &Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.
25. Guisan, A., Thuiller, W., & Zimmermann, N.E. (2017). *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press, Cambridge.
26. Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., et al. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16, 1424-1435.
27. Haddou, Y., Mancy, R., Matthiopoulos, J., Smout, S., &Dominoni, D.M. (2022). Widespread extinction debts and colonization credits in United States breeding bird communities. *Nature Ecology & Evolution*, 6, 324-331.
28. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer, New York.
29. Holbrook, J.D., Olson, L.E., DeCesare, N.J., Hebblewhite, M., Squires, J.R., &Steenweg, R. (2019). Functional responses in habitat selection: Clarifying hypotheses and interpretations. *Ecological Applications*, 29, e01852.
30. Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). McGraw-Hill, New York.
31. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18-22.
32. Matthiopoulos, J., Hebblewhite, M., Aarts, G., &Fieberg, J. (2011). Generalized functional responses for species distributions. *Ecology*, 92, 583-589.
33. Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H.L., Morales, J.M., & Haydon, D.T. (2015). Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs*, 85, 413-436.
34. Matthiopoulos, J., Field, C., & MacLeod, R. (2019). Predicting population change from models based on habitat availability and utilization. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20182911.

35. Matthiopoulos, J., Fieberg, J., & Aarts, G. (2020). *Species-Habitat Associations: Spatial Data, Predictive Models, and Ecological Insights*. University of Minnesota Libraries Publishing, Minneapolis.
36. McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall, London.
37. Mi, C., Huettmann, F., Guo, Y., Han, X., & Wen, L. (2017). Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, 5, e3018.
38. Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., et al. (2015). Predictive ecology in a changing world. *Journal of Applied Ecology*, 52, 1293-1310.
39. Mysterud, A., & Ims, R.A. (1998). Functional responses in habitat use: Availability influences relative use in trade-off situations. *Ecology*, 79, 1435-1441.
40. Park, J., & Sandberg, I.W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3, 246-257.
41. Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., & Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, 26, 275-287.
42. R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/
43. Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., &Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689-1703.
44. Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913-929.
45. Scrucca, L., Fop, M., Murphy, T.B., & Raftery, A.E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8, 289-317.
46. Sillero, N., Arenas-Castro, S., Enriquez-Urzelai, U., Vale, C.G., Sousa-Guedes, D., Martínez-Freiría, F., et al. (2023). Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecological Modelling*, 456, 109671.
47. Sofaer, H.R., Hoeting, J.A., &Jarnevich, C.S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10, 565-577.
48. Therneau, T., & Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-15. https://CRAN.R-project.org/package=rpart

49. Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, 6, 627-637.
50. Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., &Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, 92, e01486.
51. Wenger, S.J., & Olden, J.D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3, 260-267.
52. Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., et al. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33, 790-802.
53. Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., et al. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43, 1261-1277.