

# Generalized functional responses for species distributions

JASON MATTHIOPOULOS,<sup>1,2,7</sup> MARK HEBBLEWHITE,<sup>3</sup> GEERT AARTS,<sup>4,5</sup> AND JOHN FIEBERG<sup>6</sup>

<sup>1</sup>Scottish Oceans Institute, School of Biology, University of St. Andrews, East Sands, St. Andrews, Fife KY16 8LB Scotland, United Kingdom

<sup>2</sup>Centre for Research into Environmental and Ecological Modeling, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LZ Scotland, United Kingdom

<sup>3</sup>Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, College of Forestry and Conservation, University of Montana, Missoula, Montana 59812 USA

<sup>4</sup>Royal Netherlands Institute for Sea Research (NIOZ), P.O. Box 59, 1790 AB Den Burg, The Netherlands

<sup>5</sup>IMARES Wageningen UR, Institute for Marine Resources and Ecosystem Studies, P.O. Box 167, 1790 AD Den Burg, The Netherlands

<sup>6</sup>Biometrics Unit, Minnesota Department of Natural Resources (Minnesota DNR), 5463-C West Broadway, Forest Lake, Minnesota 55025 USA

**Abstract.** Researchers employing resource selection functions (RSFs) and other related methods aim to detect correlates of space-use and mitigate against detrimental environmental change. However, an empirical model fit to data from one place or time is unlikely to capture species responses under different conditions because organisms respond nonlinearly to changes in habitat availability. This phenomenon, known as a functional response in resource selection, has been debated extensively in the RSF literature but continues to be ignored by practitioners for lack of a practical treatment. We therefore extend the RSF approach to enable it to estimate generalized functional responses (GFRs) from spatial data. GFRs employ data from several sampling instances characterized by diverse profiles of habitat availability. By modeling the regression coefficients of the underlying RSF as functions of availability, GFRs can account for environmental change and thus predict population distributions in new environments. We formulate the approach as a mixed-effects model so that it is estimable by readily available statistical software. We illustrate its application using (1) simulation and (2) wolf home-range telemetry. Our results indicate that GFRs can offer considerable improvements in estimation speed and predictive ability over existing mixed-effects approaches.

**Key words:** *Canis lupis*; climate change; generalized linear mixed model; habitat preference; home range; predictive models; simulation study; space-use; spatial ecology; species distributions; utilization distribution; wolf.

## INTRODUCTION

Empirical models of space use by individuals, populations and species, aim to bolster their predictions with environmental covariates. This works well for spatial interpolation and, also, spatial extrapolation when the availability of habitat types remains approximately the same (Mladenoff et al. 1999, Aarts et al. 2008). Nevertheless, in most scenarios of extrapolation, habitat availabilities will also change, implying that models estimated from single or pooled instances of data collection may fail to capture the response of species to changing environments. This phenomenon, known as a functional response in resource selection (Myrsterud and Ims 1998), is particularly influential when the study organisms respond nonlinearly to changes in the availability of different environments. Fig. 1 illustrates the problem using a simulated experiment: An animal

whose priorities alternate between feeding and hiding is observed in a particular environment (Fig. 1a and b). A generalized linear model (GLM) provides a good fit to these data (compare Fig. 1c and d). The same animal is then placed in a new environment (Fig. 1e and f) but the previously fitted model yields poor predictions of space use (compare Fig. 1g and h).

The problem was discussed by Boyce and McDonald (1999), Myrsterud and Ims (1999), and Boyce et al. (1999) and partially tackled for a few discrete habitats (Arthur et al. 1996, Mauritzen et al. 2003). Recent studies have adopted mixed-effects models to detect functional responses (Gillies et al. 2006, Hebblewhite and Merrill 2008, Godvik et al. 2009, Duchesne et al. 2010) and extract the commonalities between animals exposed to different environments. Notably, Duchesne et al. (2010) have used a discrete choice modeling framework with random coefficients, because this framework allows the relative probabilities of selection to depend on the set of choices available to the animal. Here, we show how mixed-effects models can naturally arise by explicitly

Manuscript received 15 April 2010; revised 10 September 2010; accepted 27 September 2010. Corresponding Editor: A. M. de Roos.

<sup>7</sup> E-mail: jm37@st-andrews.ac.uk

modeling the dependence of RSF coefficients on the availability of environmental resources.

We distinguish between environmental and geographical space (Hirzel and LeLay 2008, Elith and Leathwick 2009). The dimensions of environmental space are resources or environmental conditions (e.g., Fig. 1b, f). A point  $\mathbf{x} = (x_1, \dots, x_I)$  in  $I$ -dimensional environmental space completely specifies a particular environment. A point  $\mathbf{s}$  in geographical space is completely defined in terms of dimensions such as latitude, longitude, and altitude/depth (e.g., Fig. 1a, e). We consider arbitrarily small cells (lengths, areas, volumes, or hypervolumes),  $d\mathbf{x}$  and  $d\mathbf{s}$  in environmental and geographical space respectively. A cell  $d\mathbf{x}$  comprises the environmental neighborhood of the center point  $\mathbf{x}$  and a cell  $d\mathbf{s}$  comprises the geographical neighborhood of the center point  $\mathbf{s}$ .

Terminology for the units ( $d\mathbf{x}$ ) of environmental space is loaded with historical debate. An  $I$ -dimensional hypervolume in environmental space is similar to Hutchinson's (1959) definition of a species' niche. However, a unit of environmental space need not coincide with the niche of any one species and several species may use the same unit to different extents. Some authors (Arthur et al. 1996, Mauritzen et al. 2003, Aarts et al. 2008) use the term "habitat" for  $d\mathbf{x}$  but this suffers from conflicting definitions (Hall et al. 1997). Being pragmatic, and to avoid confusion, we will call  $d\mathbf{x}$  an "environmental unit."

For free-ranging animals with equal access to the entire study region, the availability of an environmental unit is the proportion of area occupied by that unit within the region. More generally, in the case of unequal accessibility, availability can be defined as the proportion of time that animals would spend in an environmental unit, in the absence of preference (Matthiopoulos 2003, Mauritzen et al. 2003).

As defined by Johnson (1980), "preference"  $w(\mathbf{x})$  is the ratio of the usage  $g(\mathbf{x})$  over the availability  $f(\mathbf{x})$  of an environmental unit  $d\mathbf{x}$  centered at  $\mathbf{x}$ . Typically,  $w(\mathbf{x})$ , known as a resource selection function (RSF), is estimated as a generalized linear model (Manly et al. 2002). The exact approach depends on how availability is measured and whether environmental space is discretized. For example, if the variables in environmental space are continuous, then the data are either 1 (presence) or 0 (absence) and can be modeled as a Bernoulli process (Aarts et al. 2008). Alternatively, a discretization of environmental space may allow several occurrences within each environmental unit. The resulting counts are often modeled as a Poisson process. In either case, the expectation  $E(Y|\mathbf{x})$  of the response variable  $Y$  is a linear function,  $\tau(\mathbf{x})$ , of covariates,  $\mathbf{x}$ , and regression parameters,  $\beta$ , on a transformed scale:

$$E(Y|\mathbf{x}) = \tau(\mathbf{x}) = h^{-1}\left(\beta_0 + \sum_{i=1}^I \beta_i x_i\right). \quad (1)$$

The link function,  $h$ , is typically the log transformation

(for count data) or logit transformation (for binary data). The modeling objective is to estimate the coefficients  $\beta$  and draw inferences about the importance and direction of the relationship between preference and environmental variables. When modeling count data arising from a regular discretization of space, it is often reasonable to assume  $w(\mathbf{x}) = \exp(\Sigma \beta_i x_i)$ . Estimating  $w(\mathbf{x})$  from binary response models is more complicated, because the interpretation of  $\tau(\mathbf{x})$  depends on the sampling design (Keating and Cherry 2004, Aarts et al. 2008). However, in logistic regression  $\exp(\Sigma \beta_i x_i)$  can approximate  $w(\mathbf{x})$  well (Johnson et al. 2006).

Although widely used, the term "resource selection" is perhaps inappropriate since the dimensions of environmental space can be nondepletable conditions (e.g., temperature) as well as resources (e.g., forage) and because organisms select combinations of values of environmental variables rather than single resources or conditions (e.g., it makes little sense to say that a particular species selects temperature. It is more likely that it selects a particular range of temperatures combined with types of vegetation, and ranges of moisture, slope, etc.).

RSFs are usually fit in environmental space and then used for geographical predictions (e.g., Fig. 1). Underlying this process, is the implicit (and incorrect) assumption that if an empirical model of preference is not anchored to particular geographical locations, then it will automatically capture the essence of the behavior of the animals and will therefore be portable across space or time. This fallacious assumption has been made by several mainstream approaches to niche modeling, despite the fact that many studies (Johnson 1980, Boyce and McDonald 1999, Mysterud and Ims 1999, Mauritzen et al. 2003, Osko et al. 2004, Aarts et al. 2008, Godvik et al. 2009, Beyer et al. 2010) have emphasized that parameter estimates of species distribution models are conditional on the availability of *all* environmental units to the study animals. Therefore, predictions of these models are valid only for the spatiotemporal frame of the data on which they were fit (Hirzel and LeLay 2008) and are furthermore completely reliant on the ad hoc definition of availability imposed by the data collection or analysis protocols (Beyer et al. 2010).

#### Generalized functional responses

A solution to this problem, alluded to by Boyce et al. (1999), is to write the coefficients  $\beta_i$  of Eq. 1 as linear functions of the availability of all environmental units and then estimate the parameters of these new functions from a wide a range of environmental scenarios. We call this approach a generalized functional response (GFR).

We define the availability function,  $f$ , which takes values  $\mathbf{x}$  from  $I$ -dimensional environmental space and satisfies the requirement  $\int_{R^I} f(\mathbf{x}) d\mathbf{x} = 1$ . For an arbitrary discretization of environmental space comprising  $N$  environmental units,  $f(\mathbf{x}_m)$  gives the relative availability

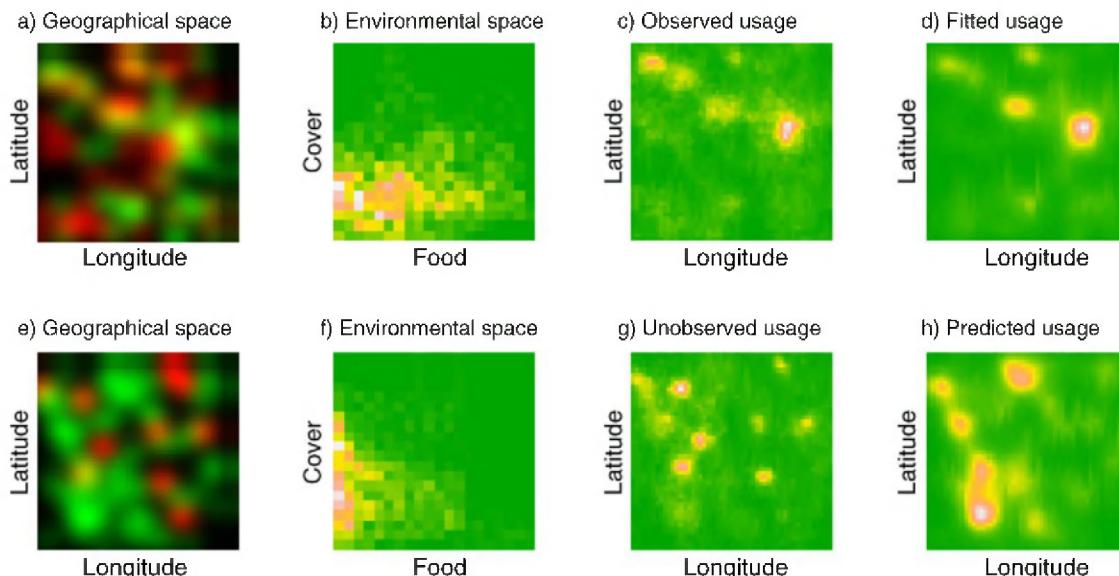


FIG. 1. Models that have been fit to data from one type of environment may predict usage poorly in new situations, as can be demonstrated by this simulated example (details on the simulation are provided in Appendix A and Supplement 1). The two rows of this composite plot represent two different environmental scenarios. The results in the top row (a–d) are based on equal overall availability of two resources. The bottom row (e–h) was produced by assuming a 1:9 split between the two resources. The first column (a, e) shows geographical space, so local densities of the two resources are represented by the intensity of the two colors (red and green). The second column (b, f) shows the environmental spaces corresponding to panels (a) and (e). The colors in these plots, going from green (low) to white (high), represent the prevalence of a particular combination of values for the two resources. When using data on the observed usage (c) and its covariates (a) to estimate a GLM, the fit is quite good (d). The same animal responds to the new regime with a new distribution of usage (g). Using the model as estimated from the previous scenario to predict the new distribution of usage (h), gives particularly poor results [compare panel (h) with the true usage in panel (g)].

of the  $n$ th environmental unit centered at the point  $\mathbf{x}_n$ . The  $i$ th coefficient of Eq. 1 may be written as an empirical function of all these availabilities:

$$\beta_i = \gamma_{i,0} + \sum_{n=1}^N \gamma_{i,n} f(\mathbf{x}_n) + \varepsilon_i \quad (i = 0, \dots, I). \quad (2)$$

Here,  $\varepsilon_i \sim N(0, \sigma_i^2)$  and the identity link function relating  $\beta_i$  to its linear predictor is the default choice, given that the regression coefficients ( $\beta$ ) are unconstrained. The  $\beta$ 's are assumed uncorrelated and are therefore modeled independently of each other. The intercept  $\gamma_{i,0}$  is the part of  $\beta_i$  that does not depend on changes in availability ( $f$ ). Therefore, Eq. 2 describes how changes in the availability of any environmental unit will make the slope ( $\beta_i$ ) of the animals' response to the  $i$ th environmental variable deviate from the baseline value  $\gamma_{i,0}$ . Note that, because  $f$  integrates to 1, changes in the availability of one environmental unit, have an impact on the availability of all units (i.e., within a given area, one environmental unit is made more abundant at the expense of others).

Eq. 2 is over-specified because it requires one  $\gamma$  coefficient for each environmental unit. Neighboring environmental units are likely to have similar effects on  $\beta_i$  so, considerable economies in the number of parameters can be achieved by replacing the individual  $\gamma$ 's by an interpolating function in environmental space

$$(\gamma_i: \mathbb{R}^I \rightarrow \mathbb{R})$$

$$\beta_i = \gamma_{i,0} + \sum_{n=1}^N \gamma_i(\mathbf{x}_n) f(\mathbf{x}_n) + \varepsilon_i. \quad (3)$$

Such economies are crucial for cases where it is difficult or undesirable to artificially discretize environmental space. For a continuous environmental space, Eq. 3 is written

$$\beta_i = \gamma_{i,0} + \int_{\mathbb{R}^I} \gamma_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + \varepsilon_i. \quad (4)$$

The case of an environmental space comprising both discrete and continuous dimensions can be treated by specifying the  $\beta_i$  as a nested combination of summation (Eq. 3) and integration (Eq. 4). Without loss of generality and to simplify notation we focus on Eq. 4. To retain some of the original flexibility of Eq. 2, the functional form of  $\gamma(\mathbf{x})$  must be allowed to be arbitrarily elastic. This may be achieved by using a polynomial of order  $M_j$  for each covariate:

$$\gamma_i(\mathbf{x}) = \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} x_j^m \quad (5)$$

where  $\delta_{i,j}^{(m)}$  is the coefficient used for the  $m$ th power of the  $j$ th environmental covariate. The intercepts are generated by allowing  $m$  to start from 0.

We now consider  $K$  sampling instances, each characterized by a different availability scenario (e.g., sampling the distribution of the same population in different years or sampling geographically distinct subpopulations). The availability of environmental units in the  $k$ th sampling instance is fully described by an instance-specific function  $f_k(\mathbf{x})$ . Although the parameters  $\beta$  (describing the response to environmental variables) are expected to differ in different sampling instances, the parameters  $\gamma$  (describing the response to changing availability) will not. Eqs. 4 and 5 give the following model for the  $\beta$ 's:

$$\begin{aligned}\beta_{i,k} &= \gamma_{i,0} + \int_{\mathbf{R}'} \left\{ \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} x_j^m \right\} f_k(\mathbf{x}) d\mathbf{x} + \varepsilon_{i,k} \\ &= \gamma_{i,0} + \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} \int_{\mathbf{R}'} x_j^m f_k(\mathbf{x}) d\mathbf{x} + \varepsilon_{i,k} \\ &= \gamma_{i,0} + \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} E[X_j^m]_k + \varepsilon_{i,k}\end{aligned}\quad (6)$$

where  $E[X_j^m]_k$  is the  $m$ th moment of the  $j$ th environmental variable calculated for the conditions prevailing in the  $k$ th sampling instance. Therefore, the coefficient of the  $i$ th covariate on the  $k$ th sampling instance can be expressed in terms of the moments of the availabilities of all environmental covariates on that sampling instance, i.e., the moments of the marginals of  $f_k(\mathbf{x})$  are used as cluster-level predictors, that remain constant for all observations within a sampling instance.

The full model from Eq. 1 can now be expanded with the aid of Eq. 6:

$$\begin{aligned}\tau(\mathbf{x}) &= h^{-1} \left\{ (\gamma_{0,0} + \varepsilon_{0,k}) + \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{0,j}^{(m)} E[X_j^m]_k \right. \\ &\quad \left. + \sum_{i=1}^I \left( (\gamma_{i,0} + \varepsilon_{i,k}) x_i + x_i \sum_{j=1}^I \sum_{m=0}^{M_j} \delta_{i,j}^{(m)} E[X_j^m]_k \right) \right\}\end{aligned}\quad (7)$$

where  $\mathbf{x}$  refers to the environmental conditions associated with a particular observation made in the  $k$ th sampling instance. The linear predictor therefore comprises (1) a random intercept of the form  $(\gamma + \varepsilon)$ , (2) mixed-effects terms of the form  $(\gamma + \varepsilon)X$ , (3) fixed effects involving the expectations  $\delta E[X^m]$  of each environmental variable in each sampling instance, and (4) all pairwise interactions  $\delta X E[X^m]$  between environmental variables and their moments.

#### More complicated formulations for the linear predictor

In accordance with the general methodology of GLMs, the formulation of the RSF in Eq. 1 can be extended by including nonlinear terms such as interactions between environmental variables (e.g.,  $\beta x_i x_j$ ) or powers of single environmental variables (e.g.,  $\beta x_i^2$ ).

These additions can be readily accommodated by the GFR framework: the nonlinear terms ( $\beta x_i x_j$ ,  $\beta x_i^2$ ) enter Eq. 7 as additional variables but the expectation terms  $E[X_i^m]_k$  refer only to the nonlinear terms of the model.

#### Mixed-effects implementation

Recent years have seen an increase in the use of mixed-effects models to take account of individual/group variation (Gillies et al. 2006, Aarts et al. 2008, Hebblewhite and Merrill 2008). Hebblewhite and Merrill (2008) suggested the use of mixed-effects models to capture variations due to differences in resource availability. The mixed-effects estimation framework is suitable for the model in Eq. 7 because it caters for random coefficients and quantifies the variance in each sampling instance. We therefore generalize on the approach of Hebblewhite and Merrill (2008) in the following sense: Like that paper, we detect the existence of a functional response and estimate a mixed model that refers to each and all of the sampling instances in the data. This is achieved by the terms of type 1 and 2 in Eq. 7. In addition, we introduce the terms of type 3 and 4, which help predict usage in any new scenario of availability.

#### Application to simulated data on home range use

We constructed a simple individual-based model of the trade-off between food ( $u$ ) and cover ( $v$ , the converse of predation risk). A similar real-life scenario with mutually exclusive resources was considered by Mauritzen et al. (2003) but in our simulation, the two resources were independently distributed. For simulation details, see Appendix A and Supplement 1.

To generate different availabilities for most environmental units, we randomly manipulated the overall amount of food ( $u$ ) and cover ( $v$ ), within a range of 1–100 arbitrary units. We obtained space use data from 10 such “training” scenarios and combined them in the fitting data set. We fit log-linear GLMs to the rate of occurrence of observations per grid cell and standardized the spatial predictions to sum to 1. We used four linear predictors corresponding to different approaches: (1) random coefficients (RE), (2) random intercept with interactions involving first-order expectations (O1), (3) random intercept with interactions involving second-order expectations (O2), and (4) random coefficients, with interactions involving first-order expectations (REO1):

$$\begin{aligned}\text{RE} & a_0 + a_1 u + a_3 v \\ \text{O1} & a_0 + \delta_1 u + \delta_3 v + \delta_4 \bar{u} + \delta_5 \bar{v} + \delta_6 u \bar{u} + \delta_7 v \bar{v} + \delta_8 v \bar{u} \\ & + \delta_9 u \bar{v} \\ \text{O2} & a_0 + \delta_1 u + \delta_3 v + \delta_4 \bar{u} + \delta_5 \bar{v} + \delta_6 u \bar{u}^2 + \delta_7 v \bar{v}^2 + \delta_8 u \bar{u} \\ & + \delta_9 v \bar{v} + \delta_{10} v \bar{u} + \delta_{11} u \bar{v} + \delta_{12} u \bar{u} \bar{u}^2 + \delta_{13} v \bar{v} \bar{v}^2 \\ & + \delta_{14} v \bar{u} \bar{u}^2 + \delta_{15} u \bar{v} \bar{v}^2 \\ \text{REO1} & a_0 + a_1 u + a_3 v + \delta_4 \bar{u} + \delta_5 \bar{v} + \delta_6 u \bar{u} + \delta_7 v \bar{v} + \delta_8 v \bar{u} \\ & + \delta_9 u \bar{v}\end{aligned}\quad (8)$$

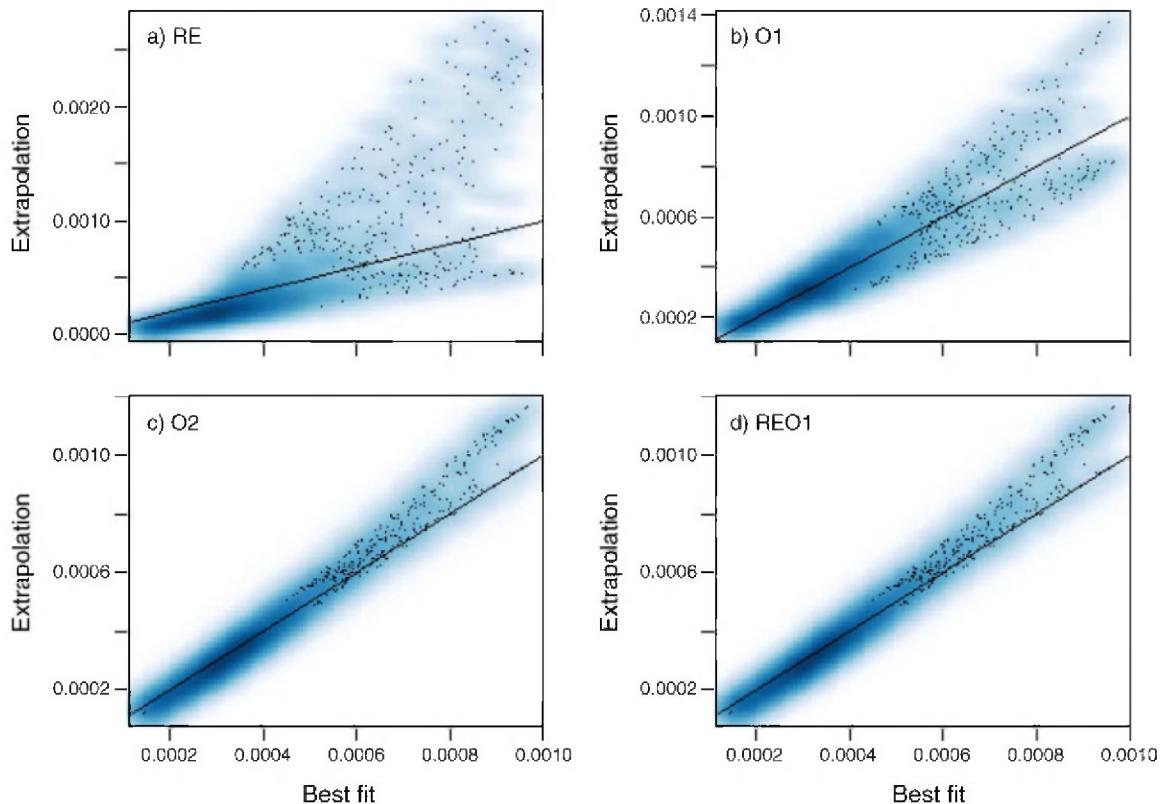


FIG. 2. Quality of extrapolation for simulated scenario 2. Each plate refers to one of the four models (RE, O1, O2, and REO1) and compares that model's predictions (on the  $y$ -axis) with the best possible estimate obtainable by a GLM fit directly to the new data (on the  $x$ -axis). The scatter plot is smoothed to indicate high concentrations of points (in dark blue), but the 200 most extreme points are also shown as black dots. The 45° line is shown for easier reference.

where  $\bar{u}$ ,  $\bar{v}$ ,  $\bar{u}^2$ ,  $\bar{v}^2$  are the first- and second-order expectations of food and cover in each environmental scenario, the  $a$ 's are random coefficients of the form  $a = \gamma + \varepsilon$  whose random components ( $\varepsilon$ ) are grouped by scenario and the  $\delta$ 's are fixed effects throughout. The values of  $u$ ,  $v$  in these models vary by observation but the values of  $\bar{u}$ ,  $\bar{v}$ ,  $\bar{u}^2$ ,  $\bar{v}^2$  vary only by sampling instance. Of these four models, RE corresponds to the approach of Hebblewhite and Merrill (2008) and the others are GFRs as in Eq. 7.

We asked each model to extrapolate usage in three new scenarios of availability. Scenario 1 ( $u = 50$ ,  $v = 50$ ), was in the middle of the ranges (0 to 100) used for the 10 training scenarios. Scenario 2 ( $u = 80$ ,  $v = 60$ ) was away from the middle but still within the ranges used for fitting. Scenario 3 ( $u = 120$ ,  $v = 120$ ) was outside the range of the availabilities used for fitting. Predictions for each scenario were made from the fixed effects of each model, based on the two environmental layers.

To evaluate the models, a GLM was fit to the usage data from each of the three prediction scenarios. These fitted responses represented a smooth surface approximating the animal's true usage and were used to calculate the precision of the predictions from the four models:  $\Sigma(\text{GLM} - \text{predicted})^2$ .

Under scenario 1, all models performed equally well giving predictions that captured the unknown underlying distribution (Appendix B: Fig. B1). This result indicated that the interaction terms of models O1 and O2 were able to perform the same role as the random coefficients of model RE. Under scenario 2, the GFR models (O1, O2, REO1), performed considerably better than RE, giving better spatial predictions (Appendix B: Fig. B2) and considerably higher precision (Fig. 2). By comparison, all models performed poorly under scenario 3. Models RE and O1 mis-predicted many of the spatial features of the underlying distribution (Appendix B: Fig. B3). However, models O2 and REO1 were more robust to such environmental extrapolation.

#### *Application to wolf telemetry data*

To illustrate the GFR model alongside the mixed modeling approach of Hebblewhite and Merrill (2008), we used the same data as that paper. The authors found that nesting the data by pack and individual improved model fit, but most of the variance in the data was explained at the level of the individual. We used summertime data for 11 wolves with the same explanatory variables selected by Hebblewhite and Merrill (2008). To simplify this illustration, we only clustered

the data by individual. We followed a use-availability design, in which the response data took the value 1 (telemetry point) or 0 (point randomly selected from the home range of each wolf). We used one random point for each telemetry observation and calculated covariate expectations for each wolf from the random points.

We compared two models. The first, (RE) was similar to the model of Hebblewhite and Merrill (2008) containing random coefficients for continuous variables (e.g., human activity) and fixed coefficients for factor levels (e.g., vegetation type). The second model (O1), had a random intercept, interaction terms using 1st order expectations for the continuous variables and fixed coefficients for factor levels. Given that the 11 wolves belonged to only five packs, there was not sufficient diversity in the environmental scenarios to support more complex models (like O2 or REO1 examined in the simulation study).

The performance of these two models was evaluated as follows. Models RE and O1 were fit to the data from 10 wolves and their estimated fixed effects were used to generate predictions for the 11th. These predictions were compared to the best estimates generated from a GLM that was fit to the data from the missing wolf. We compared the predictive precision of RE and O1 repeatedly by omitting all 11 wolves in turn.

We found that O1 gave better predictions than RE for 8 out of 11 wolves. Graphical comparisons between the two models for all wolves can be found in Appendix C. The R code used for the analysis is listed in Supplement 2.

#### DISCUSSION

Processes such as climate change and habitat fragmentation are occurring at increasing rates on a global scale, implying that most species will need to adapt to rapidly changing environmental conditions. Since mitigation often happens through spatially explicit conservation measures, it is important to anticipate change in spatial distributions. This is easier said than done because observed large-scale population distributions arise from complex interactions between physiological, demographic and behavioral responses at the level of the individual (Guisan and Thuiller 2005).

We have presented an addition to the RSF and species distribution literature that increases the predictive reach of these widely used models. The main advantages of the method are: (1) It removes the bias imposed on the fixed effects by unbalanced sampling effort across different environmental scenarios. It therefore decouples the quality of the predictions from the vagaries of the sampling regime. (2) It potentially replaces random coefficients by interaction terms hence speeding up fitting and allowing the estimation of models with more covariates. (3) It can help make better use of a fixed amount of sampling effort. For example, if data are collected from the extremes of a species' range, GFRs may be used to predict species distribution in its interior. (4) It is easy to implement with available software. We used the lme4 library in R (Bates and Maechler 2010). (5)

It has intuitive appeal. Interaction terms are frequently employed in empirical models to capture changes in the response to one covariate brought about by another. Here, they describe changes in an organism's response to environmental attributes as a nonadditive function of the statistical characteristics of its entire environment.

The components of the method could be further extended. For example, the linear formulation for the  $\beta_i$  in Eq. 2 may be reconsidered in the light of more mechanistic arguments, involving animal behavior and life history priorities (Buckley et al. 2010). It is likely that this will further increase the model's predictive power. The flexibility of the method might also be increased by exploring other possibilities for the function  $\gamma_i(\mathbf{x})$  such as kernel and spline smoothers.

We illustrated the method using both simulated and real data. In both cases, use of a GFR brought improvements in predictive ability but these were more pronounced for the simulation. There are several reasons why the wolf data might have proved more challenging. (1) The simulation assumed that the animal was observed until convergence of the home range had been achieved but this cannot be guaranteed for any of the wolves in the sample. (2) The simulation used two covariates and 10 environmental scenarios. The wolf analysis used three continuous covariates, several factor levels for habitat, and individuals from five environmental scenarios (wolves in the same pack experience similar conditions, even if they do not respond in the same way). (3) Unlike the simulation that assumed exactly the same rules of behavior, real animals are likely to behave intrinsically differently from each other, even when exposed to the same environments. (4) The wolf territories were neighboring, so it could be argued that even the limited number of five scenarios were not too dissimilar from each other. If the above explanations are valid, they suggest that the GFR model passed a rather strenuous validation test by extracting a predictive trend in the coefficients of the RSF based only on five contiguous environmental scenarios.

The difficult problem of predicting use in new environments will nearly always require extrapolation of some form. Spatial and temporal extrapolation are unavoidable for every applied objective. Environmental extrapolation will also be required whenever the multivariate distribution of environmental variables does not fall within the range of conditions experienced in the data set used to fit the model. Here, we have accepted this challenge and suggested a possible way through it. Our solution works well in cases of spatiotemporal extrapolation and may also prove more robust than other empirical models for environmental extrapolation.

#### ACKNOWLEDGMENTS

We thank the organizers and participants (particularly Francesca Cagnacci, E. Merrill, and Mark Boyce) of the workshop "GPS Telemetry Data: Challenges and Opportunities," hosted by the E. Mach Foundation in September 2008 at Viote del Monte Bondone, Trento, Italy. We acknowledge

insightful comments from D. Haydon, J. Pots, M. Loneragan, N. Yoccoz, H. Beyer, and an anonymous referee.

#### LITERATURE CITED

- Aarts, G., M. L. MacKenzie, B. J. McConnell, M. A. Fedak, and J. Matthiopoulos. 2008. Estimating space use and habitat preference from wildlife telemetry data. *Ecography* 31:140–160.
- Arthur, S. M., B. F. J. Manly, L. L. McDonald, and G. W. Garner. 1996. Assessing habitat selection when availability changes. *Ecology* 77:215–227.
- Bates, D., and M. Maechler. 2010. lme4: linear mixed-effects models using Eigen and Eigen. R package version 0.999375-31. (<http://CRAN.R-project.org/package=lme4>)
- Beyer, H., D. Haydon, J. Morales, J. L. Frair, M. Hebblewhite, M. Mitchell, and J. Matthiopoulos. 2010. Habitat preference: understanding use versus availability designs. *Philosophical Transactions of the Royal Society B* 365:2245–2254.
- Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. *Trends in Ecology and Evolution* 14:268–272.
- Boyce, M. S., L. L. McDonald, and B. F. J. Manly. 1999. Relating populations to habitats: reply to Myrsetud and Ims. *Trends in Ecology and Evolution* 14:490.
- Buckley, L. B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears. 2010. Can mechanism inform species distribution models? *Ecology Letters* 13:1041–1054.
- Duchesne, T., D. Fortin, and N. Courbin. 2010. Mixed conditional logistic regression for habitat selection studies. *Journal of Animal Ecology* 79:548–555.
- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697.
- Gillies, C. S., M. Hebblewhite, S. E. Nielsen, M. A. Krawchuk, C. L. Aldridge, J. L. Frair, D. J. Saher, C. E. Stevens, and C. L. Jerde. 2006. Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology* 75:887–898.
- Godvik, I. M. R., L. E. Loe, J. O. Vik, V. Veiberg, R. Langvatn, and A. Myrsetud. 2009. Temporal scales, trade-offs, and functional responses in red deer habitat selection. *Ecology* 90:699–710.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993–1009.
- Hall, L. S., P. R. Krausman, and M. L. Morrison. 1997. The habitat concept and a plea for standard terminology. *Wildlife Society Bulletin* 25:173–182.
- Hebblewhite, M., and E. Merrill. 2008. Modeling wildlife–human relationships for social species with mixed-effects resource selection models. *Journal of Applied Ecology* 45:834–844.
- Hirzel, A. H., and G. LeLay. 2008. Habitat suitability modeling and niche theory. *Journal of Applied Ecology* 45:1372–1381.
- Hutchinson, G. E. 1959. Homage to Santa-Rosalia or Why are there so many kinds of animals? *American Naturalist* 93:145–159.
- Johnson, C. J., S. E. Nielsen, E. H. Merrill, T. L. McDonald, and M. S. Boyce. 2006. Resource selection functions based on use–availability data: theoretical motivation and evaluation methods. *Journal of Wildlife Management* 70:347–357.
- Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 6:65–71.
- Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management* 68:774–789.
- Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2002. Resource selection by animals: statistical design and analysis for field studies. Second edition. Chapman and Hall, London, UK.
- Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. *Ecological Modelling* 159:239–268.
- Mauritzen, M., S. E. Belikov, A. N. Boltunov, A. E. Derocher, E. Hansen, R. A. Ims, Ø. Wiig, and N. Yoccoz. 2003. Functional responses in polar bear habitat selection. *Oikos* 100:112–124.
- Mladenoff, D. J., T. A. Sickley, and A. P. Wydeven. 1999. Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. *Ecological Applications* 9:37–44.
- Myrsetud, A., and R. A. Ims. 1998. Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecology* 79:1435–1441.
- Myrsetud, A., and R. A. Ims. 1999. Relating populations to habitats. *Trends in Ecology and Evolution* 14:489–490.
- Osko, T. J., M. N. Hiltz, R. J. Hudson, and S. M. Wasel. 2004. Moose habitat preferences in response to changing availability. *Journal of Wildlife Management* 68:576–584.

#### APPENDIX A

Details of simulation used to generate the validation data (*Ecological Archives* E092-051-A1).

#### APPENDIX B

Spatial output from simulated study (*Ecological Archives* E092-051-A2).

#### APPENDIX C

Model comparisons for individual wolves (*Ecological Archives* E092-051-A3).

#### SUPPLEMENT 1

R code used for simulation (*Ecological Archives* E092-051-S1).

#### SUPPLEMENT 2

R code used for wolf analysis (*Ecological Archives* E092-051-S2).