Received: 30 December 2022 | Accepted: 27 July 2023

DOI: 10.1111/2041-210X.14208

Methods in Ecology and Evolution

RESEARCH ARTICLE

Accounting for unobserved spatial variation in step selection analyses of animal movement via spatial random effects

Rafael Arce Guillen[1] | Finn Lindgren[2] | Stefanie Muff[3] | Thomas W. Glass[4,5] | Greg A. Breed[6] | Ulrike E. Schlägel[1]

[1]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany; [2]Chair of Statistics, The University of Edinburgh, Edinburgh, UK; [3]Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway; [4]Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, Alaska, USA; [5]W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, Montana, USA and [6]Institute of Arctic Biology and Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, Alaska, USA

Correspondence
Rafael Arce Guillen
Email: arceguillen@uni-potsdam.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: SCHL 2259/1-1; Deutsche Forschungsgemeinschaft, Grant/Award Number: 491466077

Handling Editor: Aldo Compagnoni

Abstract

1. Step selection analysis (SSA) is a common framework for understanding animal movement and resource selection using telemetry data. Such data are, however, inherently autocorrelated in space, a complication that could impact SSA-based inference if left unaddressed. Accounting for spatial correlation is standard statistical practice when analysing spatial data, and its importance is increasingly recognized in ecological models (e.g. species distribution models). Nonetheless, no framework yet exists to account for such correlation when analysing animal movement using SSA.
2. Here, we extend the popular method *integrated step selection analysis* (iSSA) by including a *Gaussian field* (GF) in the linear predictor to account for spatial correlation. For this, we use the Bayesian framework R-INLA and the *stochastic partial differential equations* (SPDE) technique.
3. We show through a simulation study that our method provides accurate fixed effects estimates, quantifies their uncertainty well and improves the predictions. In addition, we demonstrate the practical utility of our method by applying it to three wolverine (*Gulo gulo*) tracks.
4. Our method solves the problems of assuming spatially independent residuals in the SSA framework. In addition, it offers new possibilities for making long-term predictions of habitat usage.

KEYWORDS
animal movement, habitat selection, inlabru, spatial statistics, step selection analysis, telemetry data

1 | INTRODUCTION

Fine-scale animal tracking has become a popular tool in ecology and conservation research (Kays et al., 2015; Nathan et al., 2008). An important application of animal movement data is to make inference about the effects of environmental resources on movement decisions of animals (Thurfjell et al., 2014). For instance, Marshall et al. (2020) analysed telemetry data of King Cobras in Northeast Thailand based on different landscape types. They showed evidence that these animals tend to move less on agricultural landscapes. Similarly, Prokopenko et al. (2017), studied the influence of roads on animal movement behaviour. Their analysis suggests that crossing roads was avoided by elk, even when traffic was low. Understanding these mechanisms is essential, as it allows humans to identify how animals react to disturbance elements or identify important habitat features, crucial for effective management and conservation. From a basic science perspective, these methods allow ecologists to understand crucial processes such as species distributions, home range formation, and assessing the intensity of species interactions (Fortin et al., 2005; Matthews et al., 2020; Thurfjell et al., 2014).

Among the most commonly used statistical tools for analysis of animals' habitat selection using telemetry data are *resource selection analyses* (RSA; Boyce et al., 2002; Manly et al., 2007) and *step selection analysis* (SSA; Forester et al., 2009, Fortin et al., 2005). The RSA framework is based on *resource selection functions* (RSF) to understand habitat selection. The SSA approach uses RSFs as well but they are weighted by a function called the movement kernel, which accounts for movement constraints between the consecutive locations. Such weighted RSF is also known as a *step selection function* (SSF) (Avgar et al., 2016). The main aim of both approaches is to understand which landscape features or different habitat types influence animal space use and to quantify the intensity of affinity or aversion to these explanatory variables. Methodologically, the original idea is to compare observed animal locations with a sample of locations that were available to the animals (Lele et al., 2013). For RSA, available locations are typically sampled uniformly over an area that is, in principle, deemed accessible to the studied animals, and the common practice is to use a logistic regression to make statistical inference on the parameters of the RSF. An inherent assumption of this approach is that there are no movement constraints for animals, that is the animals could reach any location of the study area between consecutive observation times (Fieberg et al., 2021). However, this is not plausible if the time interval between consecutively observed locations is not large enough to assume their independence, which is almost always the case when RSFs are used to analyse movement data collected by modern biotelemerty devices (Nathan et al., 2022). To better account for this, SSFs instead make inference about consecutive steps/locations that connect sequential locations rather than treating each location as independent. This is implemented through a conditional logistic regression (Fieberg et al., 2021; Forester et al., 2009). Over the years, the use of SSFs has been refined, including different approaches to sample available steps (Forester et al., 2009), a correction for methodological approximations (iSSA; Avgar et al., 2016), and a reformulation of the conditional logistic regression approach as Poisson GLM to account for individual-level effects on habitat selection strength (Muff et al., 2020). Nowadays, many datasets are modelled with help of SSFs rather than RSF approaches. A more detailed summary of the development of SSA can be found in Northrup et al. (2022). In addition, an extensive description of RSA, SSA and many other animal movement models is presented in Hooten et al. (2017).

Despite SSA being a well-established inference framework for telemetry data, there are some weaknesses. In SSA approaches, the predictive quality of the model depends exclusively on the covariates provided by the user. Examples of missing covariates that could strongly influence movement include home ranges of species or unobserved individuals that interact with tracked individuals, which are difficult to identify a priori and can have a considerable effect on space-use decisions (Börger et al., 2008; Noonan et al., 2019). Another example of unexplained spatial variation is the concept of the landscape of fear in the context of predator–prey relationships. The risk of predation may cause prey to avoid certain regions and influence their landscape-feature selection. However, the *landscape of fear* is rarely included in analyses since it is nontrivial to quantify (Gallagher et al., 2017; Gaynor et al., 2019; Gehr et al., 2017). Neither SSA nor RSA in their current form have the flexibility to explain and compensate for spatial correlation in the residuals caused by omitting relevant explanatory spatial variables. Although the SSA framework accounts for spatiotemporal autocorrelation of animal locations due to the movement process through a first-order Markov model to some extent (Potts et al., 2014), it assumes residuals to be spatially uncorrelated. As such, it does not fully account for the spatial nature of the data generating process, as the true process is governed by complex behavioural decisions of animals in a heterogeneous landscape with potentially many influencing environmental factors. Since rarely all factors influencing movement are measured, the unexplained spatial variation leads to residual spatial correlation beyond what is accounted for in the SSA model. While omitted covariates should not generate a strong estimation bias for the fixed effects (Clarke, 2005), ignoring this can cause an underestimation of the uncertainty of the parameters of interest (Fieberg et al., 2021), and underestimation of uncertainly can lead to incorrectly inferring statistical significance.

In a variety of modelling contexts, a common approach to account for spatial, temporal or spatiotemporal correlation is to incorporate a continuous-space Gaussian field (GF) into the linear predictor. A spatial GF is a random effect which follows a multivariate Gaussian distribution with mean zero and spatial dependent covariance matrix, which is typically dominated by two hyperparameters, the spatial range and the standard deviation. Based on a flexible combination of these hyperparameters, the GF can take different shapes, and thus capture various patterns of spatial dependence in the data (Lindgren et al., 2011). Including a GF in statistical models plays a fundamental role in present spatial statistics (Gelfand & Schliep, 2016; Lindgren et al., 2011, 2022) and is becoming popular in species distribution modelling (Engel et al., 2022; Lezama-Ochoa et al., 2020; Renner

et al., 2015; Ward et al., 2015). Although such species distribution models are somewhat related to SSA models for telemetry data, the latter need to account for the additional sequential nature of the data. Perhaps due to this added difficulty, GFs have not yet been implemented in SSA models.

Incorporating random effects into SSA models has long been restrictive due to the lack of existing software to fit the resulting mixed conditional logistic regression model. Fortunately, as Muff et al. (2020) point out, the conditional logistic regression model is a special case of a multinomial model and thus, it is likelihood-equivalent to a conditional Poisson model (conditioned on the number of observed events being fixed). In addition, Aarts et al. (2012) showed in the context of RSA that the maximum likelihood estimators of the slope parameters of a conditional *non-homogeneous Poisson process* (NHPP) are equivalent to the ones from the unconditional NHPP. Thus, a conditional NHPP can be fitted as an unconditional Poisson model with strata-specific intercepts (Aarts et al., 2012). Using the Poisson specification puts us in the *Generalized Linear Models* framework and therefore, random effects can be naturally incorporated (Muff et al., 2020). However, Muff et al. (2020) did not include the movement kernel in their implementation. Here, we combine the contributions of Aarts et al. (2012) and Muff et al. (2019) to show analytically that the movement kernel can be also included in the Poisson implementation. In addition, we extend this to make use of the spatial nature of telemetry data and to account for missing spatial variation, by adding a spatial GF to the SSA framework. We take advantage of the fact that incorporating a GF is nothing else than including spatial random effects in the model. Although conceptually straightforward, the implementation of this is more complex, since we are fitting a hierarchical spatial model.

A popular inference method in the Bayesian framework is using the *integrated nested Laplace approximation* (INLA). In general, when fitting hierarchical models, INLA offers a faster alternative to *Markov chain Monte Carlo* (MCMC) approaches. Unlike MCMC sampling methods, INLA is based on deterministic approximations of the marginal posterior distributions of fixed- and random effects as well of hyperparameters (Rue et al., 2009). Here, we use the R package inlabru, which is based on INLA (Bachl et al., 2019). This package is specialized in dealing with spatially structured data and is particularly convenient when fitting point processes. Simpson et al. (2016) illustrate the benefits of using a mesh-based approach known as *stochastic partial differential equations* (SPDE) when fitting an unconditional NHPP with random intensity. For this reason, we use the SPDE approach from Lindgren et al. (2011). An unconditional NHPP with random intensity is known as a *log Gaussian Cox process* (LGCP) (Diggle et al., 2013). We show analytically that we can use an LGCP to incorporate spatial random effects into SSA models. Note that this use of SPDEs is not to be confused with stochastic differential equation models for individual animal movement from e.g. Brillinger et al. (2002), Preisler et al. (2004) or Hanks et al. (2017).

We name our model *Gaussian field integrated step selection analysis* (GF-iSSA). With our method, we model unexplained spatial variation in the data, allowing users to account for missing spatial covariates, to make more reliable inference about the fixed effects and to improve model predictions. We demonstrate the utility of our approach by applying it to simulated data and real data from three female wolverines (*Gulo gulo*).

2 | MATERIALS AND METHODS

2.1 | Model description

Telemetry data consist of time series of animal locations s_t. At each time point, we specify the probability to observe an animal at a location, given where we observed it at the previous two times. In the classical step-selection model (Forester et al., 2009), this is done based on two aspects:

1. The general movement tendency of an animal in the absence of habitat selection, modelled by a movement kernel ϕ, also termed selection-free movement kernel (Fieberg et al., 2021).
2. Selection behaviour of the animal with respect to environmental variables, modelled by a RSF ω.

The spatial density of observing an animal at location s_t at time point t, given the last two observed locations s_{t-1}, s_{t-2} and a landscape $\boldsymbol{X}$, is then modelled as (Forester et al., 2009; Hooten et al., 2017):

$$f(s_t \mid s_{t-2}, s_{t-1}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{\overbrace{\phi(s_{t-2}, s_{t-1}, s_t; \boldsymbol{\theta})}^{\text{Movement kernel}}\overbrace{\omega(\boldsymbol{X}(s_t); \boldsymbol{\beta})}^{\text{RSF}}}{\underbrace{\int_{q_t \in S} \phi(s_{t-2}, s_{t-1}, q_t; \boldsymbol{\theta})\omega(\boldsymbol{X}(q_t); \boldsymbol{\beta})\, \mathrm{d}q_t}_{\text{Normalizing constant}}}, \quad (1)$$

where S represents the spatial domain over which the animal may possibly move.

The selection function ω is modelled analogously to a RSF as

$$\omega(\boldsymbol{X}(s_t); \boldsymbol{\beta}) = \exp(\beta_1 X_1(s_t) + \cdots + \beta_p X_p(s_t) + u(s_t)) = \exp(\eta(s_t) + u(s_t)). \quad (2)$$

Here, X_i, $i = 1, \ldots, n$, are spatial covariates, and we are interested in making inference about their effects on movement decisions. The term u represents a GF, which accounts for the spatial variation not explained by the fixed effects $\eta(s_t)$. Thus, u follows a multivariate Gaussian distribution with covariance matrix C:

$$u \sim \mathrm{MVN}(0, C(r_s, \sigma)), \quad (3)$$

where the hyperparameters r_s and σ represent the spatial range and the standard deviation, respectively.

The movement kernel ϕ is commonly modelled via a product of probability density functions for step lengths (SL) and turning angles (TA), and it depends on the parameter vector $\boldsymbol{\theta}$. This requires a transformation from polar to Euclidean coordinates through the change-of-variable technique, which results in an extra factor corresponding to the reciprocal of the step length (Schlägel & Lewis, 2016). Similar

to Avgar et al. (2016), we assume a gamma and a von Mises distribution for step lengths and turning angles, respectively. However, any distribution from the exponential family would work in our approach. Note that the movement kernel ϕ depends on s_{t-1} for calculating the step lengths and additionally on s_{t-2} for the turning angles. Due to the special form of densities belonging to the exponential family, we can express the movement kernel as (Avgar et al., 2016; Munden et al., 2021):

$$\begin{aligned}\boldsymbol{\phi}(s_{t-2}, s_{t-1}, s_t; \theta) &= \phi_{\mathrm{SL}}(s_{t-1}, s_t; \theta_{\mathrm{SL}})\phi_{\mathrm{TA}}(s_{t-2}, s_{t-1}, s_t; \theta_{\mathrm{TA}}) \\ &\propto \exp(\beta_{p+1}\log(\| s_t - s_{t-1} \|) + \beta_{p+2} \| s_t - s_{t-1} \| + \beta_{p+3}\cos(\psi_{s_{t-2}, s_{t-1}, s_t})) \\ &= \exp(\zeta(s_{t-2}, s_{t-1}, s_t)).\end{aligned} \tag{4}$$

Here β_{p+1} and β_{p+2} are linked to the shape and rate parameters α and δ of the gamma distribution, respectively. In addition, β_{p+3} represents the concentration parameter κ of the zero-mean von Mises distribution. A detailed derivation of Equation (4) can be found in the supplementary material section A.

Combining the movement kernel (Equation 2) and selection function (Equation 4), we can express the likelihood (Equation 1) as follows:

$$\begin{aligned}L(\boldsymbol{\beta}, \theta | s_{t-2}, s_{t-1}, s_t) &= \frac{\exp(\eta(s_t) + u(s_t) + \zeta(s_{t-2}, s_{t-1}, s_t))}{\int_S \exp(\eta(q_t) + u(q_t) + \zeta(s_{t-2}, s_{t-1}, q_t))\, dq_t} \\ &:= \frac{\Lambda(s_t | s_{t-2}, s_{t-1}, \mathbf{X}(s_t))}{\int_S \Lambda(q_t | s_{t-2}, s_{t-1}, \mathbf{X}(q_t))\, dq_t}.\end{aligned} \tag{5}$$

This is the likelihood function of a *conditional* NHPP (Aarts et al., 2012). Classic approaches to SSA first approximate the integral from Equation (1) by sampling integration points from an empirically parameterized movement kernel ϕ^* based on the observed step lengths and turning angles (Fieberg et al., 2021; Forester et al., 2009). The resulting equation has consequently the form of a conditional logistic likelihood with discrete space. In contrast to this, we interpret telemetry data at each time point directly as an observation from a conditional NHPP. From a practical perspective, however, it is more convenient to implement an *unconditional* NHPP. This can be achieved by using at each time point an unconditional NHPP with an additional intercept. It has been shown that for such a model, the maximum likelihood estimator of the slope parameters are equivalent to the ones from the conditional NHPP (Aarts et al., 2012; Muff et al., 2020). This equivalence can also be shown in a Bayesian framework for the posterior distribution of our parameters (Supplementary material section B). Therefore, the joint log-likelihood of our model for a total of $T-2$ time points (The first two locations are used to calculate the first TA and SL of the first step) results in the following:

$$\begin{aligned}\sum_{t=3}^{T} l(\boldsymbol{\beta}, \theta | s_{t-2}, s_{t-1}, s_t) = &\sum_{t=3}^{T} (\log(\Lambda(s_t | s_{t-2}, s_{t-1}, \mathbf{X}(s_t)) \exp(\beta_{0t})) \\ &- \int_S \Lambda(q_t | s_{t-2}, s_{t-1}, \mathbf{X}(q_t)) \exp(\beta_{0t})\, dq_t),\end{aligned} \tag{6}$$

where $\beta_{0t}, t = 3, \ldots, T$ represent the time dependent intercepts which allow us to use the unconditional NHPP at each time point. Note that the intensity function $\Lambda(s_t | s_{t-1}, s_{t-2}, \mathbf{X}(s_t))$ is stochastic due to the presence of spatial random effects $u(s_t)$, making it a hierarchical model. This flexible model is known as a LGCP. Given a realization of $\Lambda(s_t | s_{t-1}, s_{t-2}, \mathbf{X}(s_t))$, at each time point the model is a NHPP (Diggle et al., 2013).

The interpretation of the parameters $\boldsymbol{\beta}$ of the selection function is the same as the one presented by Fieberg et al. (2021). For example, given two locations s_u and s_v that are equally available (based on the movement kernel) and have the same spatial resources but only vary by one unit of X_1, then the relative use of these two locations is equal to $\exp(\beta_1)$ (Fieberg et al., 2021).

2.2 | Model fitting

We performed Bayesian inference with the inlabru package (Bachl et al., 2019). To fit the model specified in Equation (6) to data, the integral from Equation (6) was computed numerically over discretized space. Using the SPDE method, the model needs an approximation of the GF. This is achieved via linear basis functions applied at integration points defined by all the nodes of the mesh (Simpson et al., 2016). The mesh is a triangulation of the domain (Lindgren et al., 2011). The purpose of this is to capture all the spatial effects not included in the fixed effects. We therefore approximated Equation (6) by

$$\begin{aligned}\sum_{t=3}^{T} l(\boldsymbol{\beta}, \theta | s_{t-2}, s_{t-1}, s_t) \approx &\sum_{t=3}^{T} (\log(\Lambda(s_t | s_{t-2}, s_{t-1}, \mathbf{X}(s_t)) \exp(\beta_{0t})) \\ &- \sum_{t=3}^{T} \sum_{j=1}^{n_t} w_{t,j} \Lambda(q_{t,j} | s_{t-2}, s_{t-1}, \mathbf{X}(q_{t,j})) \exp(\beta_{0t})\end{aligned}, \tag{7}$$

for suitable integration points $q_{t,j}$ and weights $w_{t,j}$. For this, we defined a uniform mesh over the study area (Figure 1) and used the nodes of the mesh both as integration points and to approximate the GF by a Gaussian Markov random field (Simpson et al., 2016). Thus, the model can be reformulated to a single Poisson model with time-specific intercepts. Since we are not interested in making inference about them and for numerical reasons, we apply the solution of Muff et al. (2020) and specify them as random intercepts with a large fixed variance in order to avoid shrinkage.

The use of a deterministic integration scheme instead of a random *Monte Carlo* integration approach has two motivations. First, the integration scheme needs to capture the full range of freedom of the random field components in order to give a valid approximation of the likelihood. Second, for spatial dimensions lower than four, the numerical bias in a basic deterministic integration scheme decreases faster than the standard deviation of a Monte Carlo integration scheme, for an increasing number of integration points. By placing

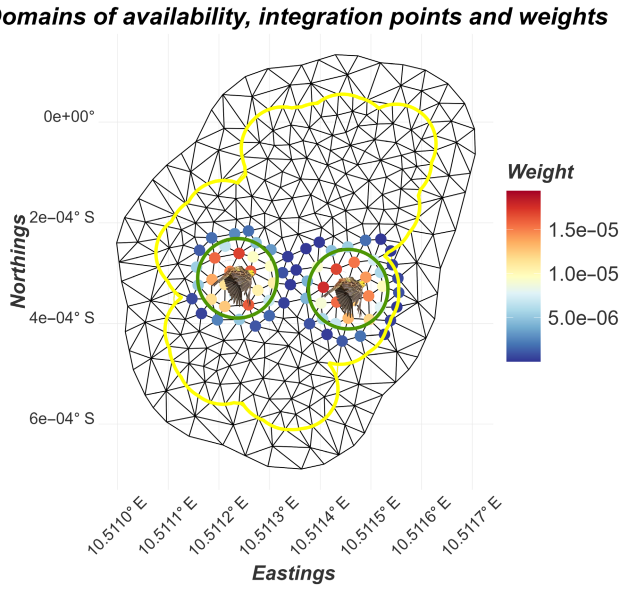

FIGURE 1 Domains of availability for two time points. The birds represent the observed animal locations. The nodes of the mesh are used as integration points with their corresponding weights based on linear basis function integrals. The yellow boundary represents the domain of interest and the nodes outside it are used to avoid boundary effects.

an integration point at each mesh node, both aspects are taken into account.

To reduce computational burden, we restricted the domain of availability S and hence integration points at each time point to a disk around the observed location s_t with a radius at least equal to the maximum observed step length over the entire dataset (Figure 1). This means that we essentially truncated the kernel ϕ at a distance for which the probability of observing a step became negligible. The integration points serve a similar purpose as the available steps in the classical SSA framework, i.e. computing numerically an approximation to the integral from Equation (1). The weights were calculated based on the integrals of the piecewise linear basis functions used to represent the spatially discretized GF model. This generated non-zero weights for all triangle vertices of triangles that intersect the corresponding availability disk (Figure 1), computed via dense deterministic sampling within each triangle (Jullum, 2020). This step was automatically performed with help of the inlabru package. A code example of our approach can be found in the supplementary material.

For all parameters of the selection function and movement kernel, we used the default R-INLA reference priors, that is a Gaussian distribution with mean equal to zero and a large variance. Furthermore, we used the penalized complexity priors for our hyperparameters as explained in Gómez-Rubio (2020). Consequently, the penalized complexity priors are defined as probability beliefs:

1. $P(r_s < r_0) = p_{r_s}$.

2. $P(\sigma > \sigma_0) = p_\sigma$.

(Fuglstad et al., 2019). Thus, the users can assign their prior beliefs through r_0, p_{r_s}, σ_0 and p_σ, respectively. Notice that r_s represents the practical range, which is the distance at which the spatial correlation is around 0.139 (Krainski et al., 2018). When setting the probabilities p_{r_s} and p_σ equal to 0.05, Fuglstad et al. (2019) suggest to assign a value σ_0 which is 2.5 to 40 times the true value of the standard deviation. In addition r_0 can be set to a value which should be between $\frac{1}{10}$ and $\frac{2}{5}$ of the true range. This specification leads to stable inference results of the marginal posterior distributions (Fuglstad et al., 2019).

2.3 | Simulation study

We simulated animal tracks based on the step-selection model defined by the conceptional likelihood from Forester et al. (2009) (Equation 1). For this, we discretized space, using a fine grid over the whole study area represented by a raster object of resolution 1000×1000 and assigning at each time point a likelihood value for each grid cell.

For the movement kernel, we specified a gamma distribution for the step lengths with shape α and rate δ parameters equal to 4 and 2, respectively. In addition, we specified for the turning angles a von Mises distribution centred at 0 with concentration parameter κ set to 1. Thus, these animals had a slight tendency to move in a straight direction.

For the selection function, we used three fixed effects: A continuous variable x1, a discrete covariate x2 and the distance to home range centre cen. Their corresponding selection coefficients were set to 1.5, 1 and − 0.04. The variable cen implements a centralizing tendency that accounts for home ranging behaviour of animals (Figure 2). In real applications x1 could for example represent elevation and x2 could describe a variable representing different landscape types. In addition, we included as a raster layer a sample of a GF with a Matérn covariance function with help of the geoR package (Ribeiro Jr et al., 2020) and incorporated it in the linear predictor. The GF represents the spatial variation not explained by the fixed effects. We simulated in total nine scenarios using different combinations of spatial ranges r_{true} (30, 40, 50) and standard deviations σ_{true} (1, 2, 3) hyperparameters for the GF. The geoR package was also used to generate the environmental covariates. For each scenario, we simulated 25 animals leading to 225 individual tracks. Each track consists of 1000 observed locations. We made sure that there is no high correlation ($|\rho| < 0.10$) between the GF and the fixed effects to avoid confounding results. In order to test the performance of the GF-iSSA on a larger parameter space, we included two additional simulation settings in the supplementary material section C. These results support the findings of this manuscript.

We fitted separate models to each individual track. In order to specify the models, the radius of the domain of availability at each time point was set to 1.25 times the maximum observed step length. Since the theoretical mean of the step length distribution is equal to 2 spatial units, we preferred to use a mesh resolution lower than this. Thus, we defined our meshes to have a maximum edge length of 1

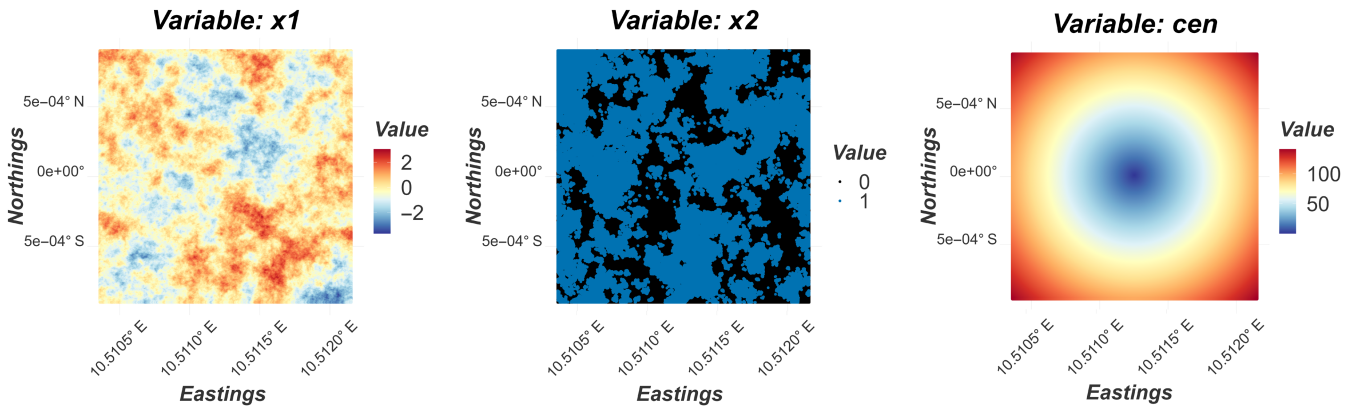

FIGURE 2 Environmental covariates of the simulation. Continuous covariate x1 (left), discrete covariate x2 (middle) and distance to home range centre cen (right).

spatial unit. This mesh is surrounded by a coarser mesh (maximum edge length equal to 5 spatial units) to avoid boundary effects. For the hyperparameters of the GF, we used the penalized complexity priors as follows:

1. $P\left(r_s < \frac{r_{\text{true}}}{4}\right) = 0.05$.

2. $P\left(\sigma > 4\sigma_{\text{true}}\right) = 0.05$.

The choice of these priors was based on Fuglstad et al. (2019) and thus, other prior choices would work properly as well.

To compare our new approach to the most commonly used one, we also fitted iSSA models (Avgar et al., 2016) to all simulated tracks, using three approaches. First, we fitted the model with the sampled GF used in the simulations as a known covariate, i.e. we treated the GF as a fixed effect (Full-iSSA). This scenario is, however, utopian for real data applications. Second, we fitted the iSSA sampling integration points from an initial movement kernel (iSSA). Third, we fitted a model using the same deterministic integration points and weights as the GF-iSSA but without including a GF in the linear predictor (NHPP). This can be understood as the iSSA model from Muff et al. (2020) but with deterministic integration points instead of random integration points. These models were chosen in order to observe how they react to missing spatial covariates. We applied the iSSA and NHPP methods excluding the GF and compared the results with our GF-iSSA approach. Both the Full-iSSA and the iSSA were fitted using a conditional logistic regression. For this, we sampled 500 available locations per used location from the initial movement kernel parametrized based on the observed step lengths and turning angles using the R package amt (Signer et al., 2019). The NHPP and the GF-iSSA were fitted as a Poisson GLM. We compared the mean estimates of the fixed effects of all four models. In addition, since the general purpose of making inference is quantifying uncertainty, we calculated the coverage properties of the GF-iSSA in comparison to the NHPP and iSSA approaches. Thus, for all simulations, we calculated for each fixed effect the rate for which the true value was covered by the corresponding 95 % credible intervals and confidence intervals, respectively. Finally, we investigated how the contribution of the GF can increase the predictive quality of the model by comparing the GF-iSSA with the NHPP. This comparison is most suitable since both approaches use exactly the same integration points. For both models, we computed squared errors (SE-Score) and Dawid-Sebastiani predictive scores (DS-Score; Gneiting & Raftery, 2007) using as predictive target the normalized RSF based on the 'true' parameter values from the simulation, arguing that for prediction purposes in habitat selection studies the RSF part of the model is typically the main interest. We calculated the predictive scores at the centroids of the grid cells used to simulate the data. Then, for each track we computed the average of the scores over all grid cells. The scores were calculated by sampling 100 realizations of the estimated posterior distribution of the RSF. Lower scores indicate a higher predictive quality.

2.4 | Case study

As exemplary case study, we applied our method to GPS collar data from three female wolverines in Arctic Alaska (Glass et al., 2021). The data were collected between March 2017 and February 2019 in the vicinity of Toolik Field Station (68.63° N, 149.60° W). Wolverines were equipped with GPS collars, programmed to obtain location coordinates every 40 min. For more details, see Glass et al. (2021). All wolverine capture and handling procedures were approved by University of Alaska Fairbanks Institutional Animal Care and Use Committee protocol 847,738 and Alaska Department of Fish and Game scientific permit 18-085.

We here used as spatial covariates two of the environmental variables from the habitat selection analysis of Glass et al. (2021): (i) distance to streams and rivers and (ii) distance to lake edges. Glass et al. (2021) also included a terrain ruggedness index as a linear and quadratic effect, as well as snow depth, density and melt as spatiotemporal explanatory covariates. We omitted these variables in our analysis and instead tested the GF's ability to adjust for them and any other missing information. We fitted first both the

GF-iSSA and the NHPP, keeping in mind that the NHPP does not account for missing spatial covariates. After this, we fitted again the NHPP but including the linear and quadratic effects of the terrain ruggedness index ($NHPP_{terrain}$) to see how the NHPP model reacts when including this covariate. Similar to Glass et al. (2021), all covariates were standardized. As interest often centres around inference at the population rather than the individual level, we illustrate how our model extension can be used to account for multiple individuals, similar to Muff et al. (2020). To this end, we included individual random slope effects for the distance to streams and rivers, while the distance to lake edges was only modelled as a population-level effect. Note that the GF summarizes in this case all the missing spatial variation at a population-level scale. In total, we had 3439 observed animal locations in our analysis.

Although we here analysed only three individuals ('F6', 'F11', 'F12'), the procedure is the same as when applying it to a greater number of individuals. The Full-iSSA could not be applied here since the true data generating process is unknown. Given that we observed some large outliers in the observed step lengths, we defined the radius for the domain of availability as the 0.99-quantile of the observed step lengths (5125 m). Observed steps that fell outside the domain of availability were removed from the analysis. This had no significant impact on the likelihood estimation and the model assumptions were not violated. The resulting gaps in the step time series were handled analogously to other missing observations and in the same way as in other implementations of SSA (see e.g. Signer et al., 2019). However, users who wish to include these outliers in the model could extend the corresponding disks of these particular time steps to have a radius equal to the respective outliers step lengths while keeping the radius of the remaining disks of availability at a smaller value. This way users can ensure that the outliers remain in the integration domain without having to extend the disks of availability for all steps to an unnecessarily large area. In addition, since the SL distribution was right-skewed with many short step lengths, we assumed an exponential distribution with rate parameter δ for the SL kernel. All other model fitting specifications were analogous to the simulation study.

3 | RESULTS

3.1 | Simulation study

Overall, the GF-iSSA reliably estimated the posterior means of the fixed effects. The estimates of the selection coefficients were on average estimated without any noticeable bias in any direction (Figure 3). The estimated posterior mean selection coefficients of our method deviated absolutely on average by 0.10 from the simulated values. In addition, our approach seemed not to be overconfident given that the box-plots cover in general the true values. This was also true for the movement kernel parameters, for which our model performed well despite the presence of missing spatial variation (Figure 3). Here, the average absolute deviation from the true movement kernel parameters was 0.08. In one out of the 225 animal tracks our method failed or crashed numerically, presumably due to unsuitable prior specification of the GF. Therefore, the results include 224 animal tracks. We calculated the running times for the GF-iSSA approach. Our models had a median running time of around 24 min per track. The 25%-quantile and the 75%-quantile were equal to around 17 and 36 min. In addition, the maximum running time was 1.5 h. We used a Debian GNU/Linux 11 cluster specifying one core for each model.

When comparing the GF-iSSA to both the NHPP and iSSA, we found that in general all methods returned similar mean estimates in case of the environmental parameters. However, the GF-iSSA showed in general wider box-plots suggesting that the GF-iSSA is more conservative in presence of spatial autocorrelation (Figure 3). When the standard deviation of the GF was low (SD = 1), all methods estimated the coefficients stably. However, when the standard deviation hyperparameter increased, all approaches seemed to show slightly more variability across the estimated selection coefficients compared to the Full-iSSA. A difference between our method and both the NHPP and iSSA methods was that the GF-iSSA tended to sometimes overestimate the selection strength while both iSSA models occasionally underestimated it. In general, all methods estimated the mean estimates of the selection coefficients decently under missing spatial variation. The marginal effect of the integration points was negligible. Both the NHPP and the iSSA returned very similar estimates and bias directions.

The mean estimates of the movement kernel parameters were, however, noticeably less biased in case of the GF-iSSA. Here, we could also clearly observe that the estimates of the iSSA and NHPP approaches were systematically underestimated with an increasing standard deviation of the GF while the ones from the GF-iSSA remained more centred to the corresponding true values (Figure 3). This was true for the shape parameter α and the concentration parameter κ. Despite being biased, the box-plots did not show a large variation suggesting that the iSSA is overconfident at estimating these parameters. As expected, the utopian Full-iSSA remained stable and did not show any noticeable bias for all the fixed effects.

On average, for all the simulations of the nine scenarios, the iSSA, NHPP and GF-iSSA returned decent mean estimates of the fixed effects, which did not deviate significantly from the true values (Table 1). However, the posterior mean estimates of the GF-iSSA showed slightly larger variation compared to the iSSA and NHPP estimates based on the 2.5%- and 97.5%-quantile of the mean estimates. In addition, the coverage results of the GF-iSSA were better than the ones from the iSSA and NHPP methods. For all the fixed effects with the exception of the coefficient for δ, our method had a higher coverage than the models with unexplained spatial variation. In addition, our method had a mean coverage of around 92% while the NHPP and iSSA had a mean coverage of 80% and 81%, respectively (Table 1), meaning that in presence of missing spatial variation the uncertainty of the iSSA and NHPP parameters was underestimated compared to the uncertainties

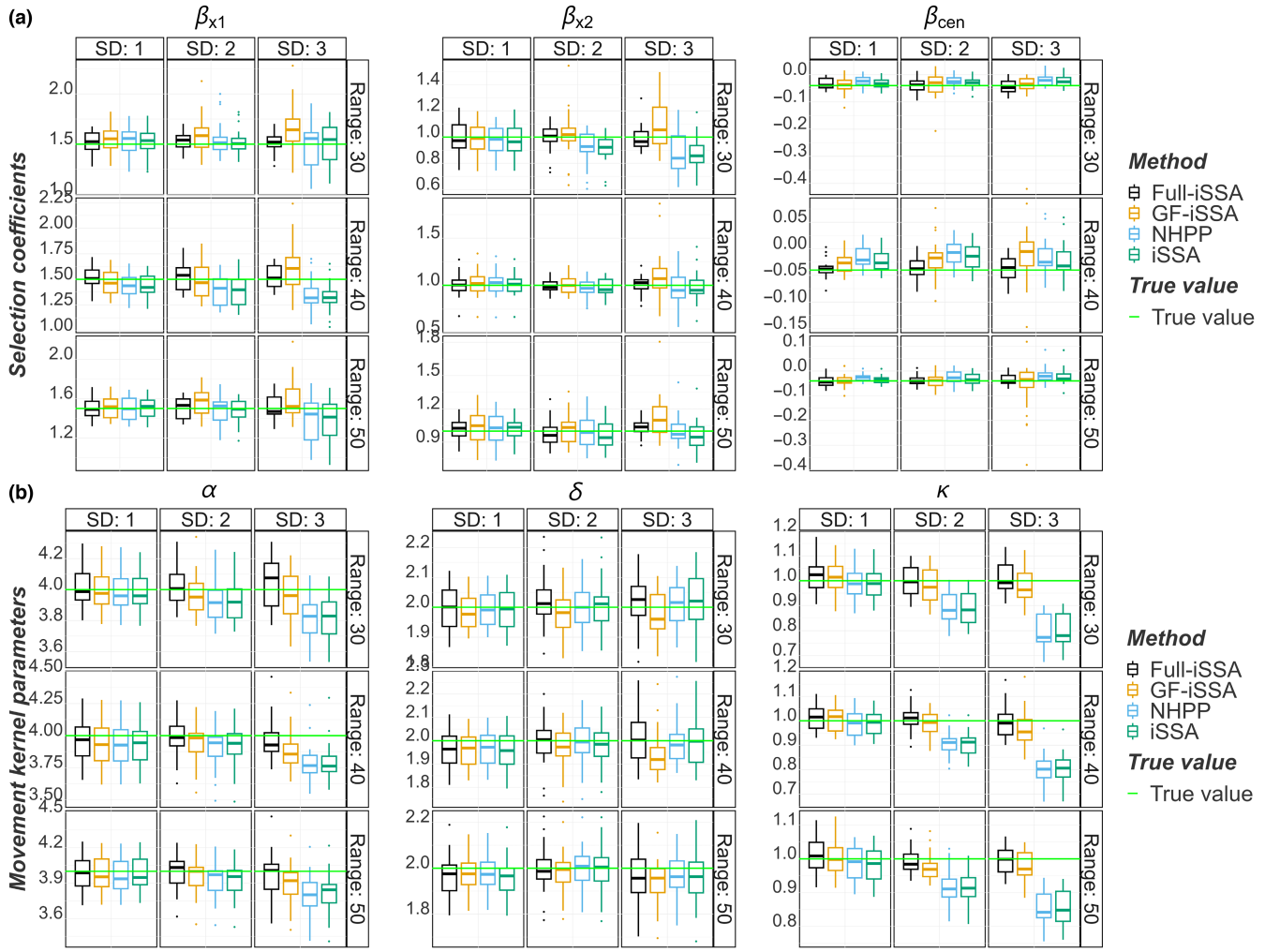

FIGURE 3 Mean estimates of fixed effects in the simulation study. Comparison of (a) selection coefficients for the three covariates x1, x2 and cen, and (b) movement kernel parameters (shape α, rate δ, concentration κ), for various combinations of the hyperparameters (standard deviation and spatial range) of the simulated Gaussian field. The non-homogeneous Poisson process (NHPP) and the integrated step selection analysis (iSSA) do not account for the missing spatial variation. The green horizontal lines represent the true values of the parameters.

estimated from the GF-iSSA, which was closer to the expected 95% coverage.

The posterior mean estimates of the hyperparameters (parameters of the GF) were slightly underestimated in case of the standard deviation. This occurred for two out of the three true values. They were near, but not surrounded by the lower- and upper quartiles (Figure S3). The three spatial range hyperparameters did not show any large bias. However, we simulated the GF for the whole study area and not for the effective study area, which was the one after joining the domains of availability with help of the aforementioned disks. Thus, is it natural to have a slightly lower variability of values of the GF within this smaller study area. This is reflected by a slightly underestimated posterior mean of the standard deviation.

The inlabru package can easily provide a map of the GF. The GF-iSSA approach estimated the GF well. It was able to detect the high and low impact areas and therefore effectively account for missing spatial covariates. We show illustratively the true GF, the estimated posterior mean distribution of the GF and a sample of it for one of our animal tracks (Figure 4). Here we can observe that for the effective study area, the SD is well estimated and explains the missing spatial variation of the study area decently. In addition, the GF can have a large contribution to the linear predictor where missing covariates have a large effect on animal movement. When displaying the contribution to the linear predictor of the spatial covariates without the GF, the estimated selection function did not match the observed animal locations. However, when additionally including the GF this mismatch became much lower (Figure 4).

The improved predictive ability of GF-iSSA was also supported by the SE-Score and DS-Score results. For all nine scenarios, the SE-Score suggested a better predictive performance of the GF-iSSA over the NHPP. This was supported by the DS-Scores, which indicated again a better predictive performance of the GF-iSSA over the NHPP for all nine scenarios (Table 2).

TABLE 1 Mean estimates of fixed effects.

Parameter	Method	Mean	True value	0.025-quantile	0.975-quantile	Coverage (%)
β_{x1}	Full-iSSA	1.515	1.500	1.319	1.702	0.973
β_{x1}	GF-iSSA	1.558	1.500	1.262	2.016	0.871
β_{x1}	NHPP	1.456	1.500	1.127	1.776	0.739
β_{x1}	iSSA	1.450	1.500	1.132	1.774	0.741
β_{x2}	Full-iSSA	1.001	1.000	0.792	1.195	0.960
β_{x2}	GF-iSSA	1.045	1.000	0.730	1.395	0.853
β_{x2}	NHPP	0.972	1.000	0.695	1.247	0.838
β_{x2}	iSSA	0.965	1.000	0.724	1.207	0.848
β_{cen}	Full-iSSA	−0.038	−0.040	−0.080	0.004	0.969
β_{cen}	GF-iSSA	−0.036	−0.040	−0.162	0.037	0.973
β_{cen}	NHPP	−0.018	−0.040	−0.055	0.030	0.829
β_{cen}	iSSA	−0.026	−0.040	−0.064	0.023	0.893
α	Full-iSSA	3.995	4.000	3.706	4.287	0.964
α	GF-iSSA	3.946	4.000	3.636	4.235	0.951
α	NHPP	3.895	4.000	3.552	4.222	0.905
α	iSSA	3.899	4.000	3.580	4.240	0.911
δ	Full-iSSA	1.991	2.000	1.795	2.194	0.924
δ	GF-iSSA	1.968	2.000	1.814	2.124	0.938
δ	NHPP	1.988	2.000	1.825	2.151	0.955
δ	iSSA	1.986	2.000	1.789	2.182	0.920
κ	Full-iSSA	1.005	1.000	0.909	1.115	0.951
κ	GF-iSSA	0.990	1.000	0.882	1.103	0.933
κ	NHPP	0.902	1.000	0.715	1.057	0.545
κ	iSSA	0.901	1.000	0.718	1.053	0.554

Note: The results of each parameter are based on 224 tracks, summarized from all nine simulation scenarios. The quantiles are those of the mean estimates. The coverage represents the percentage for which the corresponding true values were covered by the 95% credible intervals (GF-iSSA) and Wald confidence intervals (iSSA), respectively.

Abbreviations: GF-iSSA, Gaussian field integrated step selection analysis; iSSA, integrated step selection analysis; NHPP, non-homogeneous Poisson process.

3.2 | Case study

The results of the GF-iSSA applied jointly to the three wolverine tracks demonstrate the utility of the GF-iSSA. Despite our analyses not including the terrain ruggedness index and snow layers as covariates, which were known from Glass et al. (2021) to influence movement of wolverines, our model was able to estimate selection coefficients of the two effects related to water sources that were consistent with the findings of Glass et al. (2021) (Table 3). Both the GF-iSSA and the NHPP indicated strong evidence for an effect of the distance to steams and rivers based on the 95% credible intervals. However, the GF-iSSA suggested a slightly larger overall effect of this covariate compared to the NHPP. The difference between the models becomes more apparent in the estimated effect for distance to lakes. This was larger for the GF-iSSA than for the NHPP, where the latter did not suggest any evidence for this effect, while our model indicated moderate evidence for it. Thus, while the GF-iSSA model suggested an affinity for both water sources, the NHPP only revealed evidence that distance to rivers matters. In addition, the random slope's precision parameter τ was smaller for the GF-iSSA compared to the NHPP, indicating that some of its variation was absorbed by the GF (Table 3). In addition, the uncertainty of this parameter was larger for the NHPP. Estimates of the step length parameter revealed a further difference between the models. The GF-iSSA and the NHPP indicated a rate parameter equal to 2.113 and 2.944, resulting in a corresponding average SL of around 469 and 340 meters, respectively. Finally, the concentration parameter mean estimates were negative, which is outside of the theoretical domain of this parameter. However, the observed turning angles for these animals indicate that they were more likely to turn abruptly rather than travelling in a straight line (Figure S4). Thus, returning a negative concentration parameter indicates that the mean of the von Mises distribution was not necessarily 0, as expected for data with relatively large time intervals between consecutive locations (here every 40 min).

When including the terrain variable in the analysis, the $\text{NHPP}_{\text{terrain}}$ model returned very similar estimates as the NHPP model. However it still returned a positive coefficient for the distance to lakes with

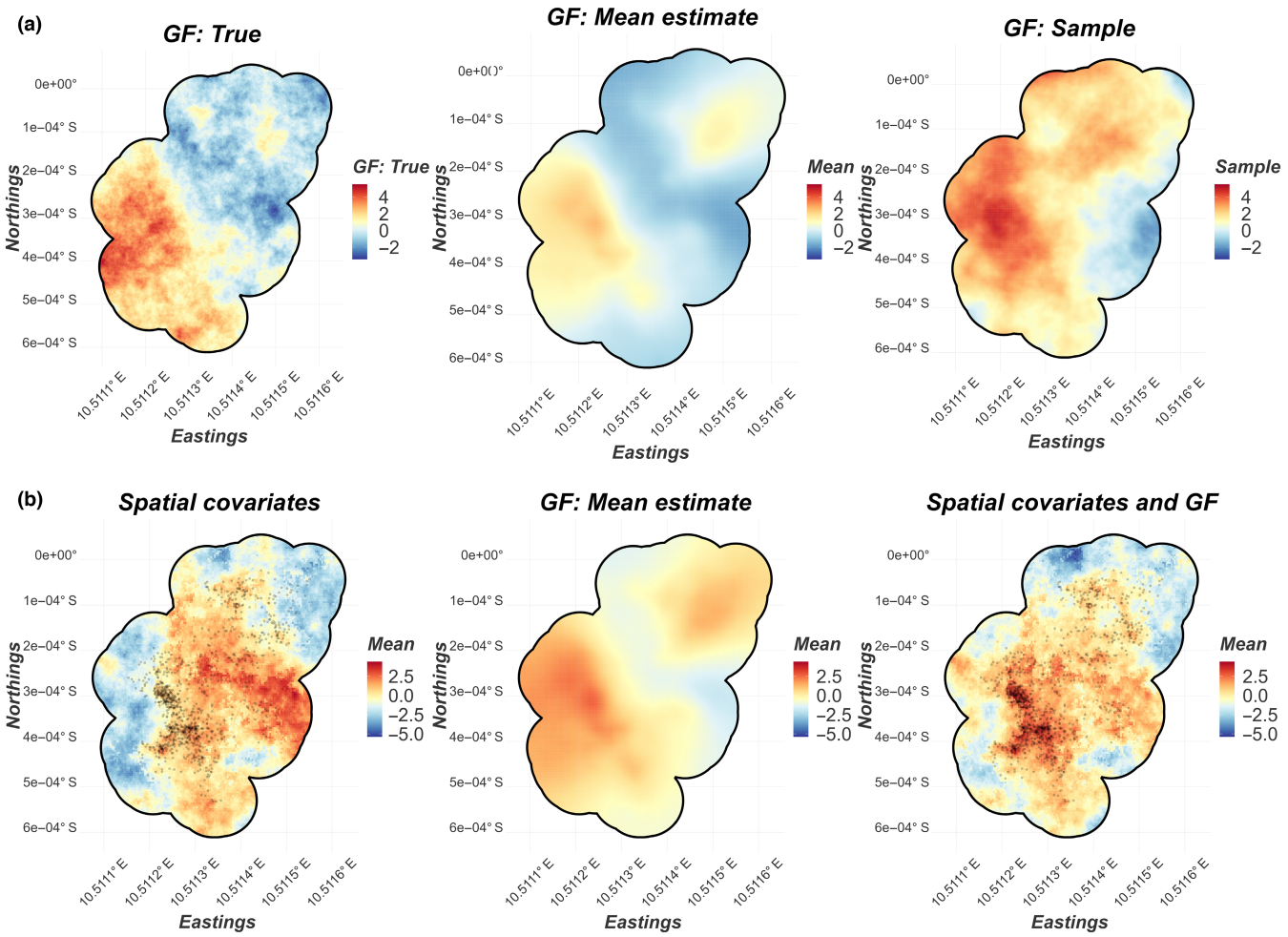

FIGURE 4 Visualization of the Gaussian field (GF) for one track and its predictive contribution. (a) Comparison between the true GF (left), the mean estimates of the GF (middle) and a sample of it (right). (b) Contribution to the linear predictor of the spatial covariates (left), the estimated GF (middle) and the combined effect (right). The black dots represent the observed animal locations.

very little evidence for an effect (Table 3). In their original analysis, Glass et al. (2021) reported negative effects for both the distance to rivers and lakes with the absolute effect larger for the distance to rivers (acknowledging that our analysis was limited to three of the original 21 individuals). This is more in line with the reported results from the GF-iSSA model.

Given the estimated standard deviation of the GF, the population-level effect of the unexplained spatial variation was rather large for the wolverine data (Figure S5). For our three tracks, the GF-iSSA estimated on average a spatial range of 2130 m and a standard deviation of 1.47. This result was expected since the model did not explicitly include the terrain ruggedness index, which had a large effect based on Glass et al. (2021).

4 | DISCUSSION

Here, we introduced an approach for incorporating and estimating spatial random effects in step selection analyses. This method, which we term GF-iSSA, helps to resolve a perennial problem in animal movement analyses by accounting for residual spatial autocorrelation arising from unobserved spatial effects in habitat selection analyses via a continuous-space GF. Rather than being an alternative model, the GF-iSSA extends iSSA to account for missing spatial covariates. Incorporating random effects leads to more conservative estimates of fixed effects parameters, which in current SSA frameworks may be contaminated by unaccounted spatial covariates. In addition, the estimate of the GF itself can be biologically interpreted and help to identify hitherto unknown causes of spatial autocorrelation in observed movement (e.g. landscape features). We have shown that our approach works on both simulated and empirical data. Our model estimated fixed effects means and uncertainties reliably. Including spatial random effects in analyses of spatially structured data has become a common practice in the last decade; analyses of animal movement data should not be an exception.

Based on a simulation study, we have shown that the GF-iSSA reliably estimates fixed-effects for various simulation settings. Overall, our method produced accurate estimates for both the selection coefficients and the movement kernel parameters. In particular, GF-iSSA estimates of the movement-kernel parameters were

consistently close to the underlying simulated parameter values, in contrast to estimates from the iSSA and NHPP, which showed a tendency to underestimate the shape and concentration parameters. This may be due to these models trying to adjust for the unexplained small step lengths by underestimating the shape parameter α. Large step lengths are not likely to occur since the tracks are also controlled by the centralizing tendency cen. In addition, the GF that was added to the simulated data leads to more variability in the observed turning angles, inducing the model to detect that the animal is less likely to travel in a straight line and, ultimately, to the underestimation of the concentration parameter κ. Regarding the habitat selection parameters, the difference in mean estimates from GF-iSSA and iSSA was less pronounced, showing some robustness of the iSSA under missing spatial variation. In general, for both the GF-iSSA and the iSSA, variability in the mean estimates increased as the standard deviation of the GF increased. However, not accounting for spatial correlation (i.e. using iSSA) resulted in an underestimated uncertainty of these parameters. Thus, based on the coverage results, the GF-iSSA quantifies uncertainty more reliably and is therefore more likely to prevent underestimating the uncertainty and resultant increased risk of spurious statistical support. In addition, our method estimated the GF accurately and is therefore able to represent the underlying spatial process in its entirety.

TABLE 2 Predictive performance of the GF-iSSA and NHPP models.

Range	SD	Method	Median (DS-score)	Median (SE-score)
30	1.00	GF-iSSA	−0.98	1.06
30	1.00	NHPP	3.84	1.19
30	2.00	GF-iSSA	0.33	2.83
30	2.00	NHPP	37.68	5.69
30	3.00	GF-iSSA	0.11	5.63
30	3.00	NHPP	111.49	15.31
40	1.00	GF-iSSA	−0.66	1.18
40	1.00	NHPP	16.36	1.54
40	2.00	GF-iSSA	0.32	1.81
40	2.00	NHPP	54.48	3.66
40	3.00	GF-iSSA	0.71	4.04
40	3.00	NHPP	349.60	9.45
50	1.00	GF-iSSA	−1.52	0.87
50	1.00	NHPP	1.53	1.02
50	2.00	GF-iSSA	−0.90	2.16
50	2.00	NHPP	13.99	2.84
50	3.00	GF-iSSA	−0.04	2.74
50	3.00	NHPP	37.55	3.90

Note: The first two columns represent the nine hyperparameter combinations of the GF. The scores represent the median of the corresponding scores of 25 animal tracks.

Abbreviations: GF-iSSA, Gaussian field integrated step selection analysis; NHPP, non-homogeneous Poisson process.

The SE and DS scores show that our method estimates the RSF-part of the model more accurately than the iSSA approach. However, these scores can only be calculated when the true RSF is known. Therefore, for real data, we recommend using the deviance information criterion (DIC) as a measure of goodness of fit. Similar to AIC, a model with a lower DIC should be preferred (Gómez-Rubio, 2020; Spiegelhalter et al., 2002). R-INLA provides these scores in the model output but they are based on the unconditional NHPPs likelihood that contain the time-dependent intercepts. Although this model yields the correct estimates for our model as defined through Equation (1), it may not return the exact corresponding value of the joint likelihood function. As an alternative, the DIC can be calculated manually for the series of conditional NHPPs, defined through the original likelihoods (Equation 1) for all steps, using a numerical approximation to the integral. For computational reasons, this is out of the scope of this paper but we show a minimal example code of the DIC calculation in the supplemental material.

In the wolverine case study, the GF-iSSA recovered the parameters of the static habitat features from the original analysis of Glass et al. (2021), even though we omitted known influential covariates (snow and terrain ruggedness). This indicates that the GF-iSSA could explain most of the spatial variation of the missing

TABLE 3 Estimates of fixed effects for movement tracks of the three female wolverines simultaneously.

	GF-iSSA			NHPP			$NHPP_{terrain}$		
Parameter	Estimate	CI 0.025	CI 0.975	Estimate	CI 0.025	CI 0.975	Estimate	CI 0.025	CI 0.975
β_{rivers}	−1.153	−2.195	−0.111	−1.056	−1.918	−0.193	−1.110	−1.894	−0.326
β_{lakes}	−0.256	−0.538	0.026	−0.003	−0.152	0.147	0.026	−0.136	0.188
$\beta_{terrain}$	—	—	—	—	—	—	1.086	0.984	1.188
$\beta_{terrain^2}$	—	—	—	—	—	—	−0.027	−0.046	−0.008
τ random slope	1.182	0.941	1.457	1.925	0.262	6.274	2.379	0.305	7.936
δ	2.133	2.032	2.234	2.944	2.841	3.048	2.985	2.874	3.096
κ	−0.139	−0.191	−0.087	−0.155	−0.206	−0.105	−0.169	−0.222	−0.116

Note: Selection was estimated with respect to distance to rivers and streams (rivers), and distance to lakes (lakes) for of the GF-iSSA and the NHPP. The $NHPP_{terrain}$ includes additionally linear and quadratic effects of the terrain ruggedness index (terrain). CI 0.025 and CI 0.975 represent the 2.5% and 97.5% credible intervals. The δ parameter represents the rate parameter of the assumed exponential distribution for the SL kernel.

covariates, allowing for reliable inference of the fixed effects that were included. Additionally, from the GF-iSSA model fit, we can derive suggestions for potentially missing effects in the analysis. The estimated selection-free movement tendency (i.e. independent of habitat or other preferences) of the wolverines was estimated to be larger by the GF-iSSA compared to the models without GF. In other words, the GF-iSSA indicates a stronger constraint of movement due to animals' preferences compared to the models without GF, suggesting that the inclusion of the GF in the model allowed us to capture habitat-selection (in a wide sense) effects on movement beyond the included covariates. One possibility may be home-ranging behaviour and individuals preferring more familiar locations in the environment (Spencer, 2012), restricting their observed movements compared to their a priori movement abilities. Notice that for simplicity and computational reasons, we preferred to omit the snow layers from the analysis since they were spatiotemporal covariates.

From a methodological perspective, the case study revealed that using deterministic integration points via a mesh and therefore estimating the parameters of the movement kernel directly (GF-iSSA, NHPP) could lead to estimates that are outside their theoretical domains (e.g. negative values for the concentration parameter of the von Mises distribution). Nonetheless, this would also be the case if the user fits the iSSA with deterministic integration points. Thus, this problem does not arise from the added GF, but rather from the integration strategy. In addition, although less likely, the same problem may also occur even when sampling integration points from an initial movement kernel (traditional iSSA) and is therefore inherent to the model formulation. Accounting for this model behaviour with restrictive priors led to numerical issues that prevented model fitting. Thus, we opted to allow the movement kernel parameters to be estimated without constraint. Enforcing parameter space constraints, in addition to introducing numerical issues, may also result in incorrect inference in case of the concentration parameter κ since there is no unique parametrization,

$$\kappa\cos(\mathrm{TA} - 0) = -\kappa\cos(\mathrm{TA} - \pi). \quad (8)$$

Thus, a negative concentration parameter κ is an indicator that the underlying von Mises distribution was not centred at zero, as usually assumed for simplicity in iSSA. However, this has no impact on the inference of the other model parameters. For this reason, restricting the concentration parameter κ without additionally estimating the mean parameter μ, may not represent well the underlying Von Mises distribution. In case of the SL kernel parameters, with inlabru, users could restrict these to their theoretical domains by specifying non-Gaussian priors. Nonetheless, this may not be numerically as stable as not constraining these parameters. In addition, with this package users could specify movement kernel distributions that do not belong to the exponential family. However, this is out of the scope of this paper.

Despite all presented advantages of the GF-iSSA, it comes with a higher computational cost than some conventional alternatives, depending upon the desired resolution of the mesh. A finer mesh requires more integration points and commensurately longer computation times for model fitting. Fitting LGCPs is generally computationally expensive but the aforementioned advantages, in our opinion, justify the relatively modest numerical expenditures here (median fitting time is about 24 min).

For our simulation study, we assumed that the GF is independent of the fixed effects. This is a strong assumption, since in real situations the spatial effects represented by the GF are likely to affect both the dependent variable and the observed spatial covariates simultaneously. If this assumption is not met, this could lead to biased inference results given the potential high correlation between the GF and some spatial covariates (Thaden & Kneib, 2018). This problem, commonly known as spatial confounding, may however be present in any real data, irrespective of whether we use iSSA or GF-iSSA and does not arise with the inclusion of the GF. A future avenue for development that might resolve this issue could be to use flexible spatial confounding approaches like *Spatial+* from Dupont et al. (2020) or the *Geoadditive Structural Equation Model* approach from Thaden and Kneib (2018). These approaches require first accounting for missing spatial variation in the linear predictor. Thus, the GF-iSSA opens new horizons for addressing these confounding issues. Nonetheless, it is not yet clear how to adapt and formally implement these approaches in a Bayesian framework. Spatial confounding in species distribution modelling has however, been already discussed by van Ee et al. (2022).

We emphasize that our model, as well as other SSA framework models, do not make inference about the true movement process of the animals since the data are discrete in time. Rather, it describes the observational process of animals at constant, discrete time intervals. However, methods using non-constant time intervals like the *time-varying iSSA* (tiSSA) from Munden et al. (2021) could be extended to include a GF as a spatial random effect in the linear predictor. For this, the domains of availability could be defined as disks where the radius equals the time difference multiplied by the maximum observed speed. In addition, our method can also be applied for making inference about tracks of multiple animals simultaneously using random effects as proposed by Muff et al. (2020).

For the future, it would be interesting to account for missing spatiotemporal covariates in the covariance matrix of the GF. Thus, the model presented here could be extended using a non-separable spatiotemporal correlation matrix instead of the purely spatial Matérn correlation matrix. However, this is a numerical challenge. Another modification of this idea could be to use circular time instead of linear time. Circular time is convenient when making inference about patterns in a repeating time window. In our case, the time window could be a 24-h diel cycle, such as that used by Shirota and Gelfand (2017). Although their model was used for crime data, a similar approach could be used for telemetry data. For instance, Benoit-Bird et al. (2009) analysed the effect of nocturnal light on the diel migration of micronekton in the water column. In our case, the model of Shirota and Gelfand (2017) could account for the missing sunlight covariate, supposing that the sunlight had an impact on movement decisions, but the users do not dispose of this variable.

In summary, our study demonstrates how to use the GF-iSSA to account for unobserved spatial effects in habitat selection analyses. This approach has three advantages. The first and largest advantage is an appropriately estimated uncertainty, which is key to correct biological inference. Second, our method has a high predictive quality compared to methods not accounting for spatial autocorrelation. Via the GF, the GF-iSSA formally estimates unobserved spatial random effects, and can use these effects in addition to observed fixed effects to make predictions. Users additionally interested in predictions (rather than inference regarding the effects of particular habitat types on movement and space use) would thus benefit from the GF-iSSA. Although not representing the whole linear prediction, since the movement kernel is not included, the selection function of the GF-iSSA and other SSAs is used for predicting long-term movement and space use of animals (Potts & Schlägel, 2020; Signer et al., 2017). Consequently, we encourage users aiming to make such long-term predictions to use the GF-iSSA, since the predictions that account for unobserved spatial variation are likely to have better predictive power than analyses that cannot account for this variation. Third, based on our simulation study, the mean estimates of the movement kernel are slightly less biased including a GF in the model. Nonetheless, as expected, the selection coefficients estimates do not deviate meaningfully from those fitted using a normal iSSA. However, we recommend the users to focus not only on the mean estimates but also on their uncertainty, which in our simulation study was better estimated by GF-iSSA than iSSA.

AUTHOR CONTRIBUTIONS

Rafael Arce Guillen, Finn Lindgren, Stefanie Muff and Ulrike E. Schlägel contributed conceptually and formulated the model. Rafael Arce Guillen, Finn Lindgren and Stefanie Muff provided the practical implementation of the model. Thomas W. Glass and Greg A. Breed contributed with data and ecological concepts of the manuscript. All authors assisted in writing and editing the manuscript.

ACKNOWLEDGEMENTS

This project was funded by the *German Research Foundation* (DFG) Grant SCHL 2259/1-1 to UES. In addition, this article was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 491466077. We thank the BioMove Research Training Group (DFG-GRK 2118) for scientific exchange and the *Wildlife Conservation Society* for providing us wolverine data. We thank two anonymous reviewers for their valuable suggestions, which helped to improve the manuscript. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14208.

DATA AVAILABILITY STATEMENT

For obtaining the spatial covariates layers, see Glass et al. (2021). The wolverine movement data are archived at Movebank under https://doi.org/10.5441/001/1.290 (Glass & Robards, 2023). The code for the simulation study, the model fitting, the wolverine data preparation as well as the wolverine tracks inference are archived on https://doi.org/10.5281/zenodo.8260549 (Arce Guillen et al., 2023).

ORCID

Rafael Arce Guillen https://orcid.org/0000-0003-1246-7962
Finn Lindgren https://orcid.org/0000-0002-5833-2011
Stefanie Muff https://orcid.org/0000-0002-8425-0757
Thomas W. Glass https://orcid.org/0000-0001-7494-4918
Greg A. Breed https://orcid.org/0000-0002-7958-1877
Ulrike E. Schlägel https://orcid.org/0000-0001-6640-9042

REFERENCES

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*(1), 177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Arce Guillen, R., Lindgren, F., Muff, S., Glass, T. W., Breed, G. A., & Schlägel, U. E. (2023). gaussianboy/code-gfissa: 0.1. *Zenodo*, https://doi.org/10.5281/zenodo.8260549

Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*(5), 619–630.

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*(6), 760–766.

Benoit-Bird, K. J., Au, W. W. L., & Wisdoma, D. W. (2009). Nocturnal light and lunar cycle effects on diel migration of micronekton. *Limnology and Oceanography*, *54*(5), 1789–1800.

Börger, L., Dalziel, B. D., & Fryxell, J. M. (2008). Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology Letters*, *11*(6), 637–650.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, *157*(2–3), 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G., & Stewart, B. S. (2002). Employing stochastic differential equations to model wildlife motion. *Bulletin of the Brazilian Mathematical Society*, *33*, 385–408.

Clarke, K. A. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science*, *22*(4), 341–352.

Diggle, P. J., Moraga, P., Rowlingson, B., & Taylor, B. M. (2013). Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, *28*(4), 542–563.

Dupont, E., Wood, S. N., & Augustin, N. (2020). Spatial+: A novel approach to spatial confounding. *arXiv preprint* arXiv:2009.09420.

Engel, M., Mette, T., & Falk, W. (2022). Spatial species distribution models: Using Bayes inference with INLA and SPDE to improve the tree species choice for important European tree species. *Forest Ecology and Management*, *507*, 119983.

Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'how to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, *90*(5), 1027–1043.

Forester, J. D., Im, H. K., & Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: Sampling and data analysis. *Ecology*, *90*(12), 3554–3565.

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in yellowstone national park. *Ecology*, *86*(5), 1320–1330. https://doi.org/10.1890/04-0953

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, *114*(525), 445–452.

Gallagher, A. J., Creel, S., Wilson, R. P., & Cooke, S. J. (2017). Energy landscapes and the landscape of fear. *Trends in Ecology & Evolution*, *32*(2), 88–96.

Gaynor, K. M., Brown, J. S., Middleton, A. D., Power, M. E., & Brashares, J. S. (2019). Landscapes of fear: Spatial patterns of risk perception and response. *Trends in Ecology & Evolution*, *34*(4), 355–368.

Gehr, B., Hofer, E. J., Muff, S., Ryser, A., Vimercati, E., Vogt, K., & Keller, L. F. (2017). A landscape of coexistence for a large predator in a human dominated landscape. *Oikos*, *126*(10), 1389–1399.

Gelfand, A. E., & Schliep, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, *18*, 86–104. https://doi.org/10.1016/j.spasta.2016.03.006

Glass, T., & Robards, M. (2023). Data from: Spatiotemporally variable snow properties drive habitat use of an arctic mesopredator. https://doi.org/10.5441/001/1.290

Glass, T. W., Breed, G. A., Liston, G. E., Reinking, A. K., Robards, M. D., & Kielland, K. (2021). Spatiotemporally variable snow properties drive habitat use of an arctic mesopredator. *Oecologia*, *195*(4), 887–899.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. CRC Press.

Hanks, E. M., Johnson, D. S., & Hooten, M. B. (2017). Reflected stochastic differential equation models for constrained animal movement. *Journal of Agricultural, Biological, and Environmental Statistics*, *22*, 353–372.

Hooten, M. B., Johnson, D. S., McClintock, B. T., & Morales, J. M. (2017). *Animal movement: Statistical models for telemetry data*. CRC Press.

Jullum, M. (2020). Investigating mesh-based approximation methods for the normalization constant in the log Gaussian Cox process likelihood. *Stat*, *9*(1), e285. https://doi.org/10.1002/sta4.285

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240), aaa2478. https://doi.org/10.1126/science.aaa2478

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Lele, S. R., Merrill, E. H., Keim, J., & Boyce, M. S. (2013). Selection, use, choice and occupancy: Clarifying concepts in resource selection studies. *Journal of Animal Ecology*, *82*(6), 1183–1191.

Lezama-Ochoa, N., Pennino, M. G., Hall, M. A., Lopez, J., & Murua, H. (2020). Using a Bayesian modelling approach (INLA-SPDE) to predict the occurrence of the Spinetail Devil Ray (*Mobular mobular*). *Scientific Reports*, *10*(1), 1–11.

Lindgren, F., Bolin, D., & Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spatial Statistics*, *50*, 100599. https://doi.org/10.1016/j.spasta.2022.100599

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498.

Manly, B. F., McDonald, L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2007). *Resource selection by animals: Statistical design and analysis for field studies*. Springer Science & Business Media.

Marshall, B. M., Crane, M., Silva, I., Strine, C. T., Jones, M. D., Hodges, C. W., Suwanwaree, P., Artchawakom, T., Waengsothorn, S., & Goode, M. (2020). No room to roam: King cobras reduce movement in agriculture. *Movement Ecology*, *8*(1), 1–14.

Matthews, C. J. D., Breed, G. A., LeBlanc, B., & Ferguson, S. H. (2020). Killer whale presence drives bowhead whale selection for sea ice in arctic seascapes of fear. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6590–6598. https://doi.org/10.1073/pnas.1911761117

Muff, S., Signer, J., & Fieberg, J. (2019). *R code and output supporting 'Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation'*.

Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, *89*(1), 80–92.

Munden, R., Börger, L., Wilson, R. P., Redcliffe, J., Brown, R., Garel, M., & Potts, J. R. (2021). Why did the animal turn? Time-varying step selection analysis for inference between observed turning-points in high frequency data. *Methods in Ecology and Evolution*, *12*(5), 921–932.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19052–19059.

Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., & Bijleveld, A. I. (2022). Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*, *375*(6582), eabg1780.

Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., Altmann, J., Antunes, P. C., Belant, J. L., & Beyer, D. (2019). A comprehensive analysis of autocorrelation and bias in home range estimation. *Ecological Monographs*, *89*(2), e01344.

Northrup, J. M., Vander Wal, E., Bonar, M., Fieberg, J., Laforge, M. P., Leclerc, M., Prokopenko, C. M., & Gerber, B. D. (2022). Conceptual and methodological advances in habitat-selection modeling: Guidelines for ecology and evolution. *Ecological Applications*, *32*(1), e02470. https://doi.org/10.1002/eap.2470

Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., & Lewis, M. A. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in Ecology and Evolution*, *5*(3), 253–262.

Potts, J. R., & Schlägel, U. E. (2020). Parametrizing diffusion-taxis equations from animal movement trajectories using step selection analysis. *Methods in Ecology and Evolution*, *11*(9), 1092–1105.

Preisler, H. K., Ager, A. A., Johnson, B. K., & Kie, J. G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, *15*(7), 643–657.

Prokopenko, C. M., Boyce, M. S., & Avgar, T. (2017). Characterizing wildlife behavioural responses to roads using integrated step selection analysis. *Journal of Applied Ecology*, *54*(2), 470–479.

Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., & Warton, D. I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution*, *6*(4), 366–379.

Ribeiro, P. J., Jr., Diggle, P. J., Jr, R., Maintainer Paulo, J., & Imports, M. (2020). *Package 'geor'*.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392.

Schlägel, U. E., & Lewis, M. A. (2016). A framework for analyzing the robustness of movement models to variable step discretization. *Journal of Mathematical Biology*, *73*(4), 815–845. https://doi.org/10.1007/s00285-016-0969-5

Shirota, S., & Gelfand, A. E. (2017). Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, *11*, 481–503. https://doi.org/10.1214/16-AOAS960

Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere*, *8*(4), e01771.
Signer, J., Fieberg, J., & Avgar, T. (2019). Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution*, *9*(2), 880–890.
Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, *103*(1), 49–70.
Spencer, W. D. (2012). Home ranges and the value of spatial information. *Journal of Mammalogy*, *93*(4), 929–947.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.
Thaden, H., & Kneib, T. (2018). Structural equation models for dealing with spatial confounding. *The American Statistician*, *72*(3), 239–252. https://doi.org/10.1080/00031305.2017.1305290
Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, *2*(1), 1–12.
van Ee, J. J., Ivan, J. S., & Hooten, M. B. (2022). Community confounding in joint species distribution models. *Scientific Reports*, *12*(1), 1–14.
Ward, E. J., Jannot, J. E., Lee, Y.-W., Ono, K., Shelton, A. O., & Thorson, J. T. (2015). Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications*, *25*(8), 2198–2209.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Code example and supplementary information.

Data S2: RData file containing 25 animal tracks from one scenario as well as spatial covariates.

How to cite this article: Arce Guillen, R., Lindgren, F., Muff, S., Glass, T. W., Breed, G. A., & Schlägel, U. E. (2023). Accounting for unobserved spatial variation in step selection analyses of animal movement via spatial random effects. *Methods in Ecology and Evolution*, *14*, 2639–2653. https://doi.org/10.1111/2041-210X.14208