

Within Reach? Habitat Availability as a Function of Individual Mobility and Spatial Structuring

Jason Matthiopoulos,^{1,*} John Fieberg,² Geert Aarts,³ Frédéric Barraquand,⁴ and Bruce E. Kendall⁵

1. Institute of Biodiversity, Animal Health, and Comparative Medicine, Graham Kerr Building, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom; 2. Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, Saint Paul, Minnesota 55108; 3. Wageningen Marine Research, Wageningen University and Research, Ankerpark 27, 1781 AG Den Helder, The Netherlands; and Department of Coastal Systems, NIOZ Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, The Netherlands; 4. Institute of Mathematics of Bordeaux, CNRS, University of Bordeaux, 351 Cours de la Libération, 33400 Talence, France; 5. Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106

Submitted August 8, 2019; Accepted November 8, 2019; Electronically published April 21, 2020

Online enhancements: supplemental PDF.

ABSTRACT: Organisms need access to particular habitats for their survival and reproduction. However, even if all necessary habitats are available within the broader environment, they may not all be easily reachable from the position of a single individual. Many species distribution models consider populations in environmental (or niche) space, hence overlooking this fundamental aspect of geographical accessibility. Here, we develop a formal way of thinking about habitat availability in environmental spaces by describing how limitations in accessibility can cause animals to experience a more limited or simply different mixture of habitats than those more broadly available. We develop an analytical framework for characterizing constrained habitat availability based on the statistical properties of movement and environmental autocorrelation. Using simulation experiments, we show that our general statistical representation of constrained availability is a good approximation of habitat availability for particular realizations of landscape-organism interactions. We present two applications of our approach, one to the statistical analysis of habitat preference (using step-selection functions to analyze harbor seal telemetry data) and a second that derives theoretical insights about population viability from knowledge of the underlying environment. Analytical expressions for habitat availability, such as those we develop here, can yield gains in analytical speed, biological realism, and conceptual generality by allowing us to formulate models that are habitat sensitive without needing to be spatially explicit.

Keywords: conditional availability, Gaussian mixtures, habitat selection, resource selection, step-selection functions, species distribution modeling.

* Corresponding author; email: jason.matthiopoulos@glasgow.ac.uk.

ORCID: Matthiopoulos, <https://orcid.org/0000-0003-3639-8172>; Fieberg, <https://orcid.org/0000-0002-3180-7021>; Aarts, <https://orcid.org/0000-0003-3196-4235>; Barraquand, <https://orcid.org/0000-0002-4759-0269>; Kendall, <https://orcid.org/0000-0003-1782-8106>.

Am. Nat. 2020. Vol. 195, pp. 000–000. © 2020 by The University of Chicago. 0003-0147/2020/19506-59405\$15.00. All rights reserved.
DOI: 10.1086/708519

Introduction

Habitats within an environment can be thought of as a combination of different values of environmental variables (e.g., abiotic conditions or biotic resources). Individual organisms may require multiple habitats to meet their biological needs, but these habitats may not all be equally accessible. Across species and life stages, individuals vary in their mobility, from complete sessility (e.g., individual plants), through central-place foraging (e.g., colonial breeders), to expansive nomadism (e.g., free-ranging grazers). Additionally, spatial structuring of the landscape may create separation between different types of vital habitats. Therefore, spatial heterogeneity and an organism's mobility determine the availability of habitats experienced from any given position in geographical space. Approaches used to quantify and understand space use (e.g., resource-selection functions; Manly et al. 2004) and spatial population dynamics (e.g., Matthiopoulos et al. 2015, 2019) are often formulated in environmental (or niche) spaces. Because such approaches are not explicitly geographic, they are prefaced by an “equal accessibility” assumption, hence ignoring this issue. However, it is becoming increasingly clear that the precise calculation of habitat availability can dramatically affect the inferences and predictions drawn from such models. For example, when analyzing animal usage data, we can be led to infer preference, avoidance, or indifference for the same habitats depending on our definition of habitat availability (Beyer et al. 2010). This makes inferences from species distribution models sensitive to habitat availability (Randin et al. 2006; Zurell et al. 2009; McLoughlin et al. 2010; Sinclair et al. 2010; Matthiopoulos et al. 2011; Wenger and Olden 2012; Aarts et al. 2013; Northrup et al. 2013).

To account for accessibility, some approaches use expert opinion (e.g., ad hoc buffers in step-selection functions; Thurfjell et al. 2014), simultaneous estimation (e.g., Horne

et al. 2008; Avgar et al. 2016), or empirical heuristics (e.g., post hoc model selection criteria in Paton and Matthiopoulos 2016). Most of these articles describe habitat availability in terms of summaries of samples taken from some spatial domain of relevance. For example, the average temperature prevailing in the neighborhood of a foraging animal can be calculated from a sample of temperatures measured (or remotely sensed) at points selected randomly or systematically from within a circular buffer centered at the position of the forager. Such summaries allow us to incorporate availability into analyses of space use in a particular landscape but at the expense of analytical tractability and generality across new landscapes with similar properties but limited data. Performing mathematical rather than sampling-based or numerical analyses with geographical layers is particularly difficult because parametric descriptions of heterogeneous landscapes are challenging to construct. Furthermore, not all of the details and geographical features of landscapes are necessarily relevant for summaries of habitat availability.

An alternative approach is to statistically describe the salient attributes of species mobility and landscape structure and use such statistical summaries to define habitat availability in a compact mathematical form that is an adequate approximation of the neighboring environment from the position of any given individual. This approach is both numerically efficient in describing population processes within a given landscape and generalizable across spatially similar landscapes.

The objective of this article is to formally develop the concept of habitat availability, starting from first principles, and to derive expectations about the environment based on its global statistical properties rather than any particular local configuration of habitats, an approach similar to statistical mechanics in the physics literature (e.g., Sklar 2015). Principally, the ability of an individual to move between two or more habitats will depend on how far apart they are (a distance determined by the spatial autocorrelations of the environmental variables making up these habitats) and how easily the individual can move between them (as determined by various mobility constraints; Matthiopoulos 2003). Formalizing these effects of spatial autocorrelation and mobility requires clear conceptual definitions of habitat and of unconstrained (unconditional) habitat availability. We therefore begin with a brief review of useful notation and terminology and follow it with a mathematical model for conditional habitat availability, which quantitatively maps accessibility in geographical space (the spatial locations that an organism can access from any given position) to accessibility in environmental space (the habitats that an organism can be expected to access from the habitat corresponding to its given spatial position). Given the rather abstract nature of this framework, we provide three types of

intuition-building illustrations. We compare measures of habitat availability calculated using our new framework and using a direct sampling approach applied to a buffer zone around a particular location. We use our framework to derive an analytical form of the likelihood for step-selection functions (an approach commonly used to quantify habitat selection from fine-scale telemetry data; Thurfjell et al. 2014; Hooten et al. 2017). We use our framework to investigate population fitness for territorial species. This application introduces additional mathematical tools that can allow the formulation of general results connecting habitat accessibility, habitat use, and population viability. We conclude by placing this work in its broader context.

G-spaces, E-spaces, Habitats, and Unconditional Habitat Availability

Models that deal with species-environment interactions frequently differentiate between geographical space (*G*-space) and environmental space (*E*-space), a distinction historically known as Hutchinson's duality (Hirzel and Le Lay 2008; Colwell and Rangel 2009; Elith and Leathwick 2009). *G*-space comprises the three dimensions of latitude, longitude, and altitude/depth, often projected onto a Cartesian system of coordinates. In contrast, *E*-space can be high dimensional, with each dimension representing a biotic or abiotic environmental variable, that is, a continuous, discrete, or qualitative random variable representing a condition (e.g., pH, temperature, sea depth), resource (e.g., soil nutrients, prey, breeding sites), or risk (e.g., predators, pollution). *E*-space can be considered identical to niche space, as originally conceived by Hutchinson (1957) and MacArthur (1968), although, as extensively argued in the modern literature (Soberón and Nakamura 2009; Peterson et al. 2011; McNerny and Etienne 2013; Matthiopoulos et al. 2015), statistical habitat preference models currently fitted in *E*-space should not be confused with the niche objects as envisaged by these pioneering thinkers.

Several articles (Aarts et al. 2008; Matthiopoulos et al. 2011; Matthiopoulos et al. 2015 and references therein), conceptualize habitat as a point \mathbf{x} in *E*-space, the combination $\mathbf{x} = \{x_1, \dots, x_K\}$ of specific values for K environmental variables (e.g., geomorphology and climate variables combining into the characteristic makeup of, say, "polar habitat"). Elsewhere and in colloquial use, "habitat" has been described in a species-dependent way as the region in geographical space in which an organism lives (e.g., "polar bear habitat"). The two definitions are not interchangeable (see Hall et al. 1997). We opt for the former definition because it allows objective comparisons between species and quantitative gradations of suitability. Subject to this definition of habitat, we can introduce the unconditional

availability (f_x) of a particular habitat \mathbf{x} as the relative frequency (i.e., the probability density) with which that habitat occurs across the whole landscape.

Data-derived objects recorded in G -space are typically complicated and difficult to describe parametrically. For example, describing even a single altitude contour on a map by means of a mathematical formula is a nontrivial task. In contrast, objects in E -spaces are generally simpler, as we illustrate in figure 1 by visualizing the simple case of a single environmental variable (i.e., one-dimensional E -space) measured in a linear region (i.e., one-dimensional G -space). The way in which a multimodal variable in G -space (fig. 1*b*) gives rise to a much simpler (in this case, unimodal) frequency histogram in E -space (fig. 1*a*) is typical of all landscapes because multiple occurrences in G -space of the same habitat are condensed into a single habitat frequency in E -space. Therefore, most species distribution models (SDMs) are developed and fitted in E -space rather than G -space (Hooten et al. 2017). That is not to say that a description of habitat availability in E -space based on a simple unimodal distribution is always sufficient. Since habitat availability can be a complicated object (Matthiopoulos et al. 2015), a parametric description of the unconditional availability of habitats may be suitably obtained as a mixture of multiple (e.g., Gaussian) components in K -dimensional space. For example, Matthiopoulos et al. (2015) used the well-established numerical library `mclust` (Fraley et al. 2005, 2012) in the R environment (R Core Team 2016) to approximate unconditional habitat availability in K environmental dimensions as a Gaussian mixture of L components:

$$f_x = \sum_{l=1}^L \psi_l f_{l,x} = \frac{1}{(2\pi)^{K/2} \prod_{k=1}^K \sigma_k} \sum_{l=1}^L \psi_l \exp\left(-\frac{1}{2} \sum_{k=1}^K \left(\frac{x_k - \mu_{l,k}}{\sigma_k}\right)^2\right), \quad (1)$$

where $f_{l,x}$ is the l th component (a K -dimensional Gaussian probability density function [PDF]) of the mixture, ψ_l is the weight associated with the l th component (such that $\sum_{l=1}^L \psi_l = 1$), $\mu_{l,k}$ is the mean (i.e., the location in E -space) of the l th mixture component along the k th environmental dimension, and σ_k is the characteristic standard deviation along the k th environmental dimension. Such Gaussian mixtures are universal approximators. Economy in the number L of mixture components could be achieved by extending equation (1) to allow a different standard deviation for each component. However, as in Matthiopoulos et al. (2015), we prioritize mathematical uniformity of the mixture components over parsimony. We therefore allow for a large number of components but constrain them to have the same standard deviation σ_k .

Conditional Habitat Availability

The relative simplicity of E -spaces (compared with G -spaces) comes at a price, because by condensing the environment into the relative frequencies of different habitats, we lose information on the geographical nearness between habitats. Correcting this problem requires an appropriate augmentation of E -space to account for spatial proximity, leading to the notion of conditional availability, that is, the expected availability of habitat \mathbf{x} to an organism that is currently located in coordinates \mathbf{s} , characterized by habitat \mathbf{z} . Importantly, we seek an expression for conditional availability that is not reliant on a neighborhood in G -space defined around a particular location \mathbf{s} but rather is reliant on the mixture of habitats typically encountered around a particular habitat \mathbf{z} . Such an expression would enable us to describe the key patterns in spatially local availability, without the need for models to become spatially explicit.

Figure 1 shows this concept in one spatial dimension and for one environmental variable. In this low-dimensional illustration, the general notion of a habitat \mathbf{x} is simply a particular value x of the single environmental variable X and the general location \mathbf{s} in G -space is the position s on a single spatial axis S . The unconditional availability (i.e., the frequency in E -space) of a particular value x of the environmental variable X is f_x . Collecting such frequencies for all values of the environmental variable forms a PDF in E -space (fig. 1*a*).

Subsequently, we focus on all of the spatial locations s_1, \dots, s_n (circles in fig. 1*b*) that are characterized by a particular habitat $z = 30$. An organism with constrained mobility that finds itself in one of these locations will only be able to experience neighboring locations in space. Such localized access to G -space in the neighborhoods of the points s_1, \dots, s_n is illustrated in figure 1*d* using Gaussian kernels, which describe the accessibility of a point at distance r from the current location s_i . These kernels represent the constraints on organism mobility. For example, if we were considering habitat selection by a free-ranging animal over a particular timescale (say, a year), the kernel could represent Brownian motion over that timescale. Alternatively, if the study organisms are not free-ranging (e.g., because they must provision offspring located at a central place or because they must actively defend a territory), the kernel can be thought of as the result of an Ornstein-Uhlenbeck process (a random walk with a central tendency; Blackwell 1997). An isotropic mobility kernel in any number of spatial dimensions can be recast as a function $h(r)$ that describes the probability of an organism reaching a location at distance r away from its current position over the time period of interest. For data collected infrequently enough that locations can be assumed independent, the kernel can be viewed as determining availability at the

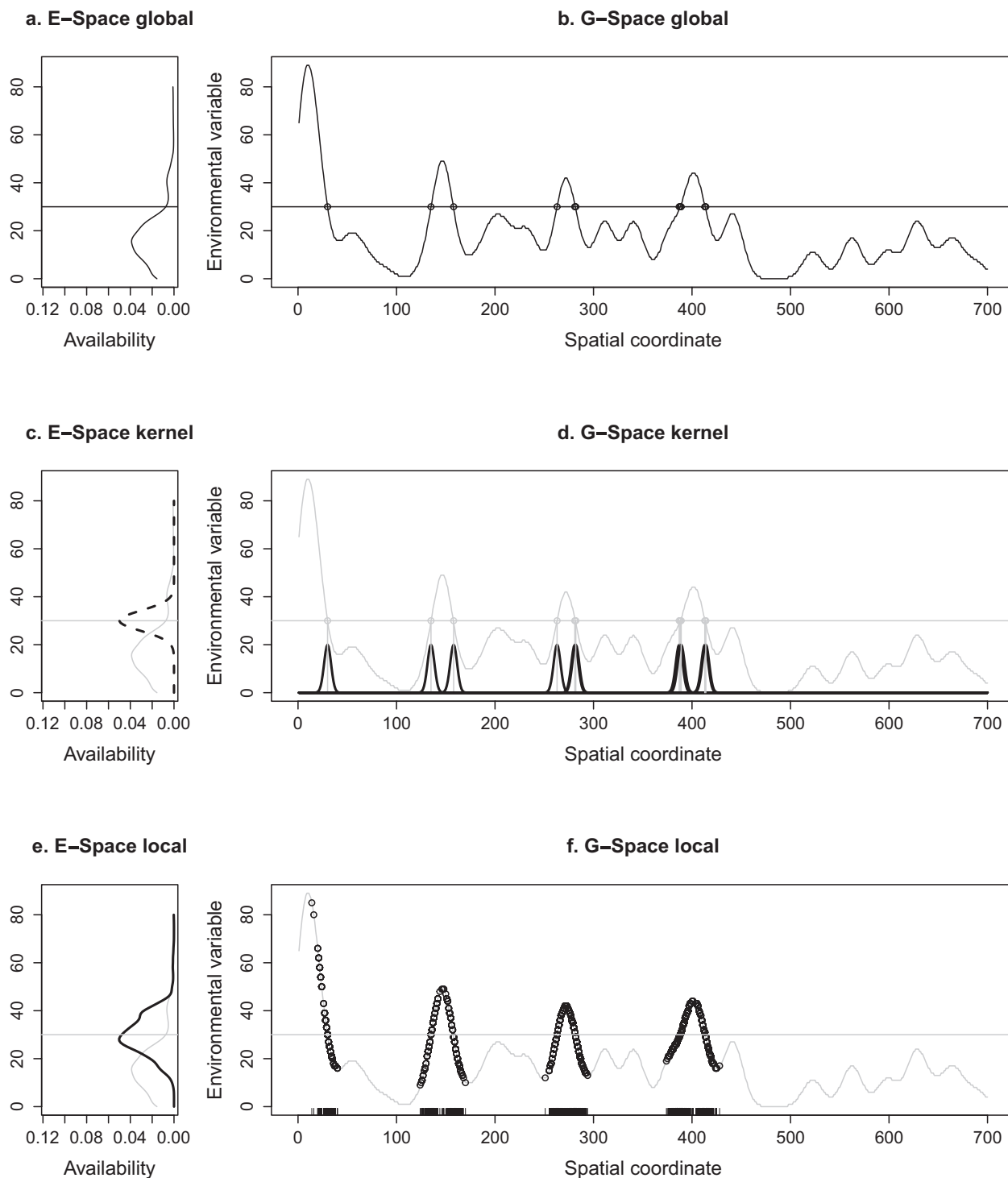


Figure 1: The panels in the left column represent *E*-space (comprising a single environmental variable) corresponding to the *G*-space (comprising a single spatial dimension) in the right column. The values plotted in *G*-space are the local values of the environmental variable X , and *E*-space summarizes the frequency with which each value of the environmental variable occurs. *G*-spaces are usually more complicated objects to describe because the same habitat can occur several times. In this example, seven values in *G*-space (the circles in *b*) are condensed to one value in the plot of *E*-space (*a*). The accessibility of space around a particular habitat can be represented by symmetric kernels (the dark curves in *d*). The existence of spatial autocorrelation in the proximity of each of these spatial locations guarantees that similar environmental values will be found within these kernels of accessibility. We represent this by the dashed curve in *c*—an imagined kernel in *E*-space that represents the correspondence between spatial and environmental proximity. A realization of the sampling process from the kernels (using Gaussian forms) provides a scattering of observations in *G*-space (shown as circles in *f*). The resulting plot of frequencies for these localized measurements is shown as the dark curve in *e*.

home range scale, similar to Horne et al.'s (2008) synoptic model of animal space use. Alternatively, our kernel can be used to model perception range. In particular, Fagan et al. (2017) make the case for mathematical formulations of semilocal perception (an intermediate between the extremes of omniscience and purely local information about habitat) and use Gaussian kernels to describe the diminishing ability of an animal to perceive habitats at greater distances.

Since the values of environmental variables in G -space are spatially autocorrelated, neighboring points in E -space (i.e., similar habitats) will tend to be found close to each other in G -space as well. Hence, proximity between locations in G -space must translate to proximity between their corresponding habitats in E -space. This is schematically represented by the single dashed curve in figure 1c peaking in the neighborhood of habitat z in environmental space.

For those animals viewing the world from the vantage points of habitat z , this localized sampling in G -space (circles in fig. 1f) yields a subjective sample of the values of X in E -space (solid black line in fig. 1e). The comparison between the two curves shown in figure 1e represents the main concern of this article: although globally the landscape contains habitats whose frequency is described by the light gray curve (the unconditional availability f_x), an organism located in a particular habitat z may be surrounded by a considerably different habitat composition, as shown by the dark curve (the conditional availability $f_{x|z}$).

To write an expression for conditional availability $f_{x|z}$ of habitat x from a position characterized by habitat z , we first consider all pairs of points in G -space separated by a distance r . If the first point is characterized by habitat z , then the probability that the second point is of habitat x is denoted by $g_{x|z}(r)$. If the organism is at the first point, then the probability that it can reach across a distance r is denoted by $h(r)$. Therefore, the product $g_{x|z}(r)h(r)$ represents the probability that habitat x is found at distance r from habitat z and that it is accessible by the organism located at habitat z . To convert this into a PDF for conditional availability irrespective of the distance between two points, we can integrate the product across distances r :

$$f_{x|z} = \frac{1}{C_E} \int_r g_{x|z}(r)h(r) dr. \quad (2)$$

The probability density $g_{x|z}(r)$ encompasses the spatial autocorrelation of habitats as well as the overall availability of habitat x , and the probability density $h(r)$ represents limitations in accessibility. Since we require $f_{x|z}$ to be a PDF of habitat availability, equation (2) contains a normalizing constant that integrates over all target habitats x :

$$C_E = \int_E \int_r g_{x|z}(r)h(r) dr dx. \quad (3)$$

In one spatial dimension the accessibility kernel can be defined as one-dimensional Gaussian, and its associated distance function will then be half-normal:

$$h(r) = \frac{1}{\omega} \left(\frac{2}{\pi} \right)^{1/2} \exp \left(-\frac{r^2}{2\omega^2} \right), \quad (4)$$

where the parameter ω determines the rate at which accessibility decays with distance from the current position. This can be extended to two spatial dimensions by assuming a two-dimensional Gaussian density for the position of the organism. This diffusion-type model implies a Rayleigh distribution (Hughes 1995) for the distance function:

$$h(r) = \frac{r}{\omega^2} \exp \left(-\frac{r^2}{2\omega^2} \right). \quad (5)$$

Non-Gaussian formulations of the mobility constraint are possible, as long as they are well behaved under integration (for an explanation of this constraint, see eq. [11]). For example, if we wished to capture the behavior of animals that interspersed localized movement by occasional long-distance forays, we may choose to implement the kernel as a fat-tailed distribution. In that case, a probabilistic model such as the t -distribution would be preferable to one whose expectations are pathological, like the Cauchy distribution (Feller 1966).

The conditional habitat availability at distance r can be derived from the relationship linking conditional and joint probabilities:

$$g_{x|z}(r) = \frac{g_{x,z}(r)}{f_z}, \quad (6)$$

where $g_{x,z}(r)$ is the joint probability density of habitats x and z , quantifying the probability that they can be encountered at distance r from each other. In this expression, f_z is the marginal distribution $f_z(r) = \int_x g_{x,z}(r) dx$. Since the destination habitat is integrated out of this expression, the marginal is independent of the distance between x and z ; hence, $f_z(r) = f_z$. Therefore, irrespective of the particular form of the joint probability distribution of habitats under the requisite distance r , the unconditional availability of habitats is preserved.

The joint distribution $g_{x,z}(r)$ must be constructed from the two marginals f_x and f_z (i.e., the unconditional habitat availabilities of habitat x and z , respectively) by introducing a dependence structure. Dependence structures between any two marginal distributions can be constructed by the method of copulas (Joe 2014), but this is a computationally prohibitive approach because it relies on two inversions of the PDF (PDF to quantile function and back again). Furthermore, in our application the problem is particularly challenging because the marginal distributions are

high-dimensional mixtures (eq. [1]) describing the availability of multiple dimensions in E -space.

An alternative proposed by Sawo et al. (2006) for constructing joint PDFs from mixture marginals is to first decompose each marginal f_x and f_z into its L mixture components and subsequently combine each component from one marginal distribution (following eq. [1], $f_{l,x}$ for $l = 1, \dots, L$) with every mixture component from the other marginal (which is identical to the first but specified for another habitat z ; so again, following eq. [1], $f_{m,z}$ for $m = 1, \dots, L$). The weighted sum of these pairwise combinations then yields the joint mixture

$$g_{x,z}(r) = \sum_{l=1}^L \sum_{m=1}^L \psi_{l,m}(r) f_{l,x} f_{m,z}, \quad (7)$$

where $\psi_{l,m}(r)$ is a new set of weights for the pairwise combinations between $f_{l,x}$ and $f_{m,z}$ that, for any given pairwise distance r , must satisfy the following conditions:

$$\begin{aligned} \psi_{l,m}(r) &\geq 0, & \sum_{l=1}^L \sum_{m=1}^L \psi_{l,m}(r) &= 1, \\ \sum_{l=1}^L \psi_{l,m}(r) &= \psi_m, & \sum_{m=1}^L \psi_{l,m}(r) &= \psi_l. \end{aligned} \quad (8)$$

Constructing the joint availability function by means of this weighted superposition of products ($f_{l,x} f_{m,z}$) implies independence within the pairwise combinations. For a given number of mixture components, the quality of the approximation of G -space correlations within the joint distribution $g_{x,z}(r)$ could be improved by allowing the covariances of each mixture component to be nonzero. Indeed, it would be possible to allow the variance-covariance structure of each Gaussian component to be unique. These kinds of approaches would lead to efficiencies in the number of Gaussian components needed. At the other extreme, the approach we have used employs large numbers of Gaussian components, all identical and with zero covariances. As the number of components increases, their variances decrease and so does the influence of the assumption of within-component independence. The decision to employ many identical and simple mixture components was made for analytical tractability. Given that no covariance is assumed within the individual mixture components, the new weights $\psi_{l,m}(r)$ are the only remaining route of generating a covariance in the joint distribution $g_{x,z}(r)$. In other words, we are seeking to construct a covariance structure in $g_{x,z}(r)$ by reweighting radially symmetric Gaussian components. This will introduce some smoothing in the final result (see the numerical examples in “Illustration Using Direct Sampling of Availability from G -space”).

To derive the new weights $\psi_{l,m}(r)$, Sawo et al. (2006) propose an algebraic approach, which unfortunately is quite

time-consuming for mixtures of multiple components and often fails to satisfy the positivity requirement (the first condition in eq. [8]). We therefore take a more heuristic approach. In supplement 1 (supplements 1–5 are available online in the supplemental PDF), we provide an iterative normalization algorithm that constructs a matrix ψ satisfying the conditions in equation (8) for a given value of r . The distance r determines the strength of correlation between the two dimensions. If the distance is small, then the organism will expect to find itself in very similar conditions, which implies that the joint distribution must have high correlation. In contrast, if the organism takes a very large step, then it may find itself in any habitat, with probability proportional to that habitat’s global availability. The correlation strength as a function of distance r is extracted directly from the environmental data, using an empirical autocorrelation function (see supplement 1).

Placing equations (7) and (1) into equation (6) gives

$$g_{x|z}(r) = \frac{\sum_{l=1}^L \sum_{m=1}^L \psi_{l,m}(r) f_{l,x} f_{m,z}}{\sum_{m=1}^L \psi_m f_{m,z}}. \quad (9)$$

Replacing into equation (2) and rearranging the integral produces

$$f_{x|z} = \frac{\sum_{l=1}^L \sum_{m=1}^L f_{l,x} f_{m,z} \int_r \psi_{l,m}(r) h(r) dr}{C_E \sum_{m=1}^L \psi_m f_{m,z}}. \quad (10)$$

In supplement 1, we discuss how the integral in the above expression can be evaluated numerically for a single environmental variable. Henceforth, we replace these integrals by the shorthand notation $\Psi_{l,m}$, defined as

$$\Psi_{l,m} = \int_r \psi_{l,m}(r) h(r) dr. \quad (11)$$

Note that these quantities satisfy the unit-sum requirement (from the second part of eq. [8]):

$$\sum_{l=1}^L \sum_{m=1}^L \Psi_{l,m} = \int_r \left(\sum_{l=1}^L \sum_{m=1}^L \psi_{l,m}(r) \right) h(r) dr = 1, \quad (12)$$

so they can be thought of as a set of new weights to replace the original quantities $\psi_{l,m}(r)$. This simplifies the overall expression in equation (10), even after expanding the normalization constant:

$$f_{x|z} = \frac{\sum_{l=1}^L \sum_{m=1}^L f_{l,x} f_{m,z} \Psi_{l,m}}{\int_E \sum_{l=1}^L \sum_{m=1}^L f_{l,x} f_{m,z} \Psi_{l,m} dx}. \quad (13)$$

In supplement 2, we show that the denominator in this expression is the marginal distribution of availability, yielding

$$f_{x|z} = \frac{\sum_{l=1}^L \sum_{m=1}^L f_{l,x} f_{m,z} \Psi_{l,m}}{f_z}. \quad (14)$$

This prompts the identification of the numerator as the joint probability density $f_{x,z}$:

$$f_{x,z} = \sum_{l=1}^L \sum_{m=1}^L f_{l,x} f_{m,z} \Psi_{l,m}. \quad (15)$$

Plotting the numerator of equation (14) for different mobility constraints (fig. 2) illustrates the operation of the calculations of supplement 1. At low mobility (fig. 2a) the correlation between different types of habitat is strong, but increasing the mobility of the organism (as shown in fig. 2b and 2c by using higher values of ω) moves the joint distribution closer to the independence scenario $f_{x,z} = f_x f_z$.

The result in equation (15) is already applicable to one, two, or more spatial dimensions (via an appropriate specification of $h(r)$; see examples in eqq. [4] and [5]). In principle, equation (15) is also applicable to multiple environmental dimensions, but this would also require additional methodological work to generalize the algorithm in supplement 1 so as to include any cross correlations between environmental variables in addition to their autocorrelations. However, using the algorithm of supplement 1 in its current form for many environmental variables is also possible if they can plausibly be assumed to be independent of each other. The extensive literature on collinear environ-

mental variables can be used, either to test for nonindependence between environmental dimensions or (e.g., via principal component analysis) to construct a new set of independent environmental variables (Dormann et al. 2013). Given such a set of orthogonal variables, habitat availability in K -dimensional E -spaces can be written as

$$f_{x|z} = f_{x_1|z_1} f_{x_2|z_2} \dots f_{x_K|z_K}. \quad (16)$$

Illustration Using Direct Sampling of Availability from G -space

In the preceding sections, we dealt with the problem of restricted accessibility by extending the mathematical definition of habitat availability. A more direct approach to quantifying availability in a particular landscape is to sample around different locations in G -space (as we did in fig. 1). It is therefore useful to visualize the outputs of the sampling and analytical approaches on a simple example for a particular simulated landscape, to help with the interpretation of our method and to motivate a discussion of Monte Carlo error.

To generate a joint distribution of habitat availability via sampling, the following steps could be adopted:

Step 1. Systematically or randomly select a set $S = \{s_1, \dots, s_n\}$ of points in G -space.

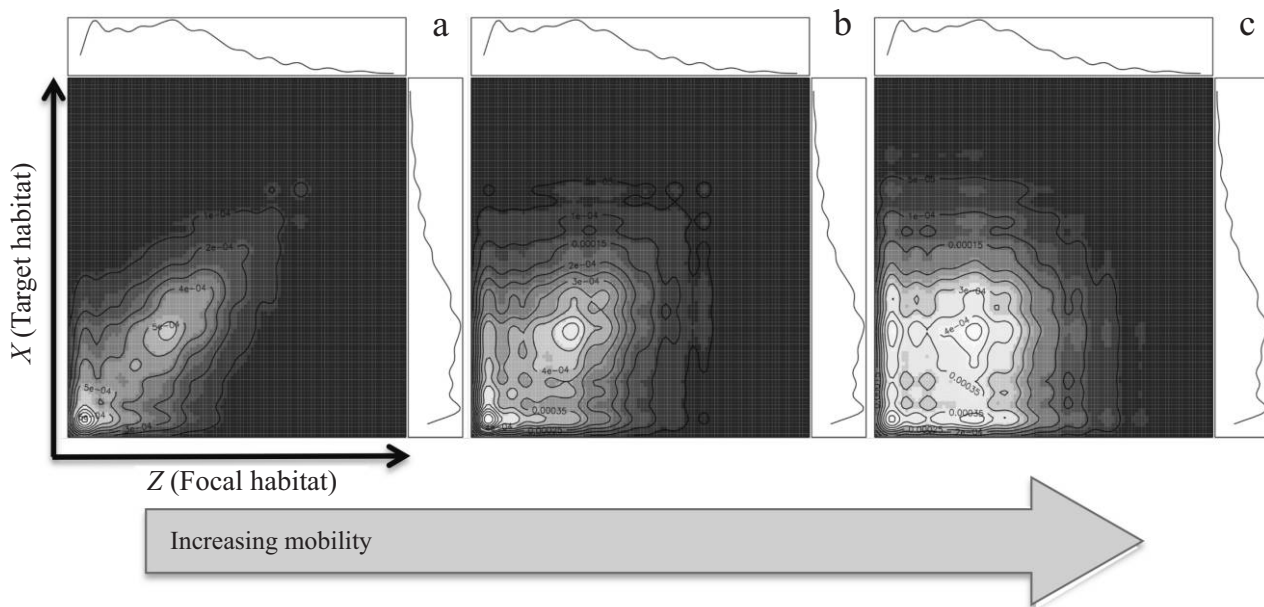


Figure 2: Joint distribution of habitat availability ($f_{x,z}$) for a given marginal habitat availability (f_z, f_x) under three examples of the mobility constraint ($\omega = 3, 5, 10$). Lighter colors represent higher probability density. The figure explores the case of a single environmental variable X ; therefore, the axes have identical units and scales, representing the support of that environmental variable in E -space. The marginal distributions shown on the sides of the main plots are identical, representing the fact that the overall habitat availability across the landscape is not affected by mobility.

Step 2. Randomly sample points (in our case, 50) from the vicinity (in G -space) of each \mathbf{s} , according to an accessibility kernel with mobility parameter ω . This will generate a set of satellite points $\mathbf{U}_i = \{\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,50}\}$ reflecting the spatial extent of conditional availability (accounting for both mobility constraint and amount of spatial autocorrelation in the environmental variables).

Step 3. For every combination of points $(\mathbf{s}_i, \mathbf{u}_{i,j})$, extract their location $(\mathbf{x}_i, \mathbf{z}_{i,j})$ in joint E -space and increment their absolute frequency by one.

Using the same simulated landscape throughout this section (see the example in supplement 1), we specified two different mobility kernels across the rows of figure 3 corresponding to slow-moving animals or short time intervals (fig. 3a, 3b) and fast-moving animals or long time intervals (fig. 3c, 3d). The analytical approach in E -space derived earlier gave the outputs of figure 3a and 3c. We compared these with the corresponding plots (fig. 3b, 3d) obtained via the sampling approach described above. The two approaches give broadly comparable descriptions of the two mobility scenarios, but the model-based approach yielded a smoother description than the sampling algorithm. These differences between the analytical and sampling plots are due to two types of stochasticity. The first relates to Monte Carlo error due to the finite sample sizes taken from each buffer zone. Small sample sizes will tend to introduce stochasticity in the representation. The second relates to the dependence of the sampling approach on the particular realization of the landscape. Many of its features are essentially a result of chance because they are likely to change if a different landscape with the same statistical properties is sampled. By relying on summaries of spatial autocorrelation, the analytical approach is likely to be more generally applicable to landscapes whose habitat geographies are shaped by similar mechanisms.

In general, sampling is more direct but has three disadvantages: (1) it is computationally expensive (because a large number of focal and satellite points is needed to overcome Monte Carlo error; this increases rapidly with the dimension of E -space); (2) it is specific to the particular realization of the environment presented in the study landscape, inhibiting both understanding about how spatial patterns affect availability and extrapolation to similar landscapes; and (3) it does not yield a compact mathematical expression such as equation (15) that can allow further applications to make algebraic shortcuts.

Applied Example: Step-Selection Functions for the Analysis of Telemetry Data

A step-selection function is a method of fitting a habitat model to animal telemetry data (Fortin et al. 2005; Thurfjell et al. 2014; Signer et al. 2019). The general step-selection

model operates in G -space and describes the likelihood that an animal performs a particular relocation from position \mathbf{s}_{j-1} to position \mathbf{s}_j with environmental attributes $\mathbf{x}(\mathbf{s}_j)$. The likelihood $f_u(\mathbf{s}_j|\mathbf{s}_{j-1})$ is described as (see Forester et al. 2009)

$$f_u(\mathbf{s}_j|\mathbf{s}_{j-1}) = \frac{w(\mathbf{x}(\mathbf{s}_j))f_a(\mathbf{s}_j|\mathbf{s}_{j-1})}{\int_{\mathbf{u} \in G} w(\mathbf{x}(\mathbf{u}))f_a(\mathbf{u}|\mathbf{s}_{j-1}) d\mathbf{u}}, \quad (17)$$

where $w(\mathbf{x}(\mathbf{s}_j))$ describes habitat preferences and $f_a(\mathbf{s}_j|\mathbf{s}_{j-1})$ expresses mobility (the “resource-independent movement kernel” described in Forester et al. 2009). The selection function $w(\mathbf{x})$ is modeled as a loglinear function of predictor variables. Here, as in Matthiopoulos et al. (2015), we employ a curvilinear polynomial form comprising terms up to the second order, to allow for the detection of optima in the animal’s response to some environmental variables:

$$w(\mathbf{x}) = \exp\left(\sum_{k=1}^K \sum_{\eta=1}^2 \gamma_{\eta,k} x_k^\eta\right). \quad (18)$$

The objective of statistical inference focuses on the selection coefficients γ . The log-likelihood function corresponding to equation (17) is

$$l(\mathbf{s}_j|\mathbf{s}_{j-1}; \gamma) = \log w(\mathbf{x}(\mathbf{s}_j); \gamma) + \log f_a(\mathbf{s}_j|\mathbf{s}_{j-1}) - \log \int_{\mathbf{u} \in G} w(\mathbf{x}(\mathbf{u}); \gamma) f_a(\mathbf{u}|\mathbf{s}_{j-1}) d\mathbf{u}. \quad (19)$$

The log likelihood of the entire data set of telemetry data is constructed by combining the individual likelihoods of all of the observed relocations in the data:

$$l = \sum_{j=1}^J l(\mathbf{s}_j|\mathbf{s}_{j-1}; \gamma). \quad (20)$$

Employing this log likelihood within standard estimation approaches—specifically, conditional logistic regression (Fortin et al. 2005)—usually involves two simplifying steps (Forester et al. 2009). First, the mobility function $f_a(\mathbf{s}_j|\mathbf{s}_{j-1})$ is assumed known, and second, the nontrivial integral of equation (19) is approximated by point-sampling methods. The first simplifying step allows the term $\log f_a(\mathbf{s}_j|\mathbf{s}_{j-1})$ to be dropped from the log likelihood, since it contains no parameters that need to be estimated from the data. The second step deals with the integral by organizing the telemetry data into strata, each comprising a single focal telemetry location \mathbf{s}_j and a sample (of size V) of control locations \mathbf{s}_v . Controls are selected randomly from the geographical vicinity of the telemetry observation \mathbf{s}_{j-1} immediately preceding \mathbf{s}_j , so as to represent the habitat options that were available to the animal from that previous position.

These two simplifying steps bring the log likelihood of equation (19) within the remit of conditional logistic

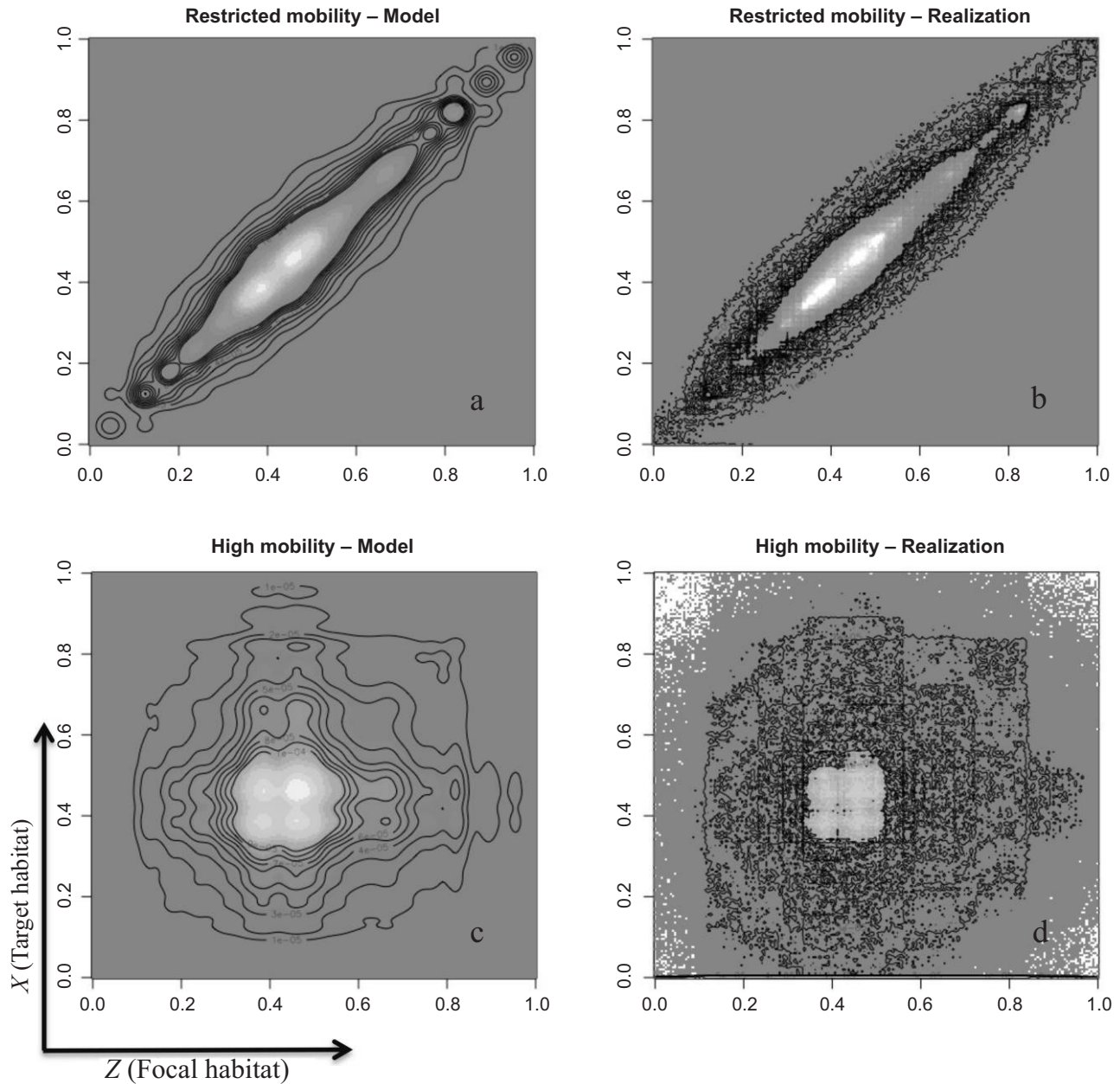


Figure 3: Comparison between the analytical and numerical forms of the joint distribution (left and right columns, respectively) under scenarios of low and high mobility (top and bottom rows, respectively). The axes represent the variables z (focal habitat) and x (target habitat) for a one-dimensional environmental space. Shading of contours (lighter shades for higher probability) is on the same scale for plots on the same row, for comparison.

regression, which, for the j th point in a telemetry data set, is written as

$$l_{\text{CLL}}(\mathbf{s}_j | \mathbf{s}_{j-1}; \gamma) = \log(w(\mathbf{x}_j; \gamma)) - \log\left(w(\mathbf{x}_j; \gamma) + \sum_{v=1}^V w(\mathbf{z}_v; \gamma)\right), \quad (21)$$

where $\mathbf{x}_j = \{x_1, \dots, x_K\}_j$ is the habitat at the j th telemetry location and \mathbf{z}_v is the habitat at the v th control location. The likelihood is conditional on the location \mathbf{s}_{j-1} in the sense that the control points are selected from within a neighborhood of that location. The above form of the likelihood is implemented in R, in the form of the `clogit()` model in the `survival` library (Therneau and Lumley 2019),

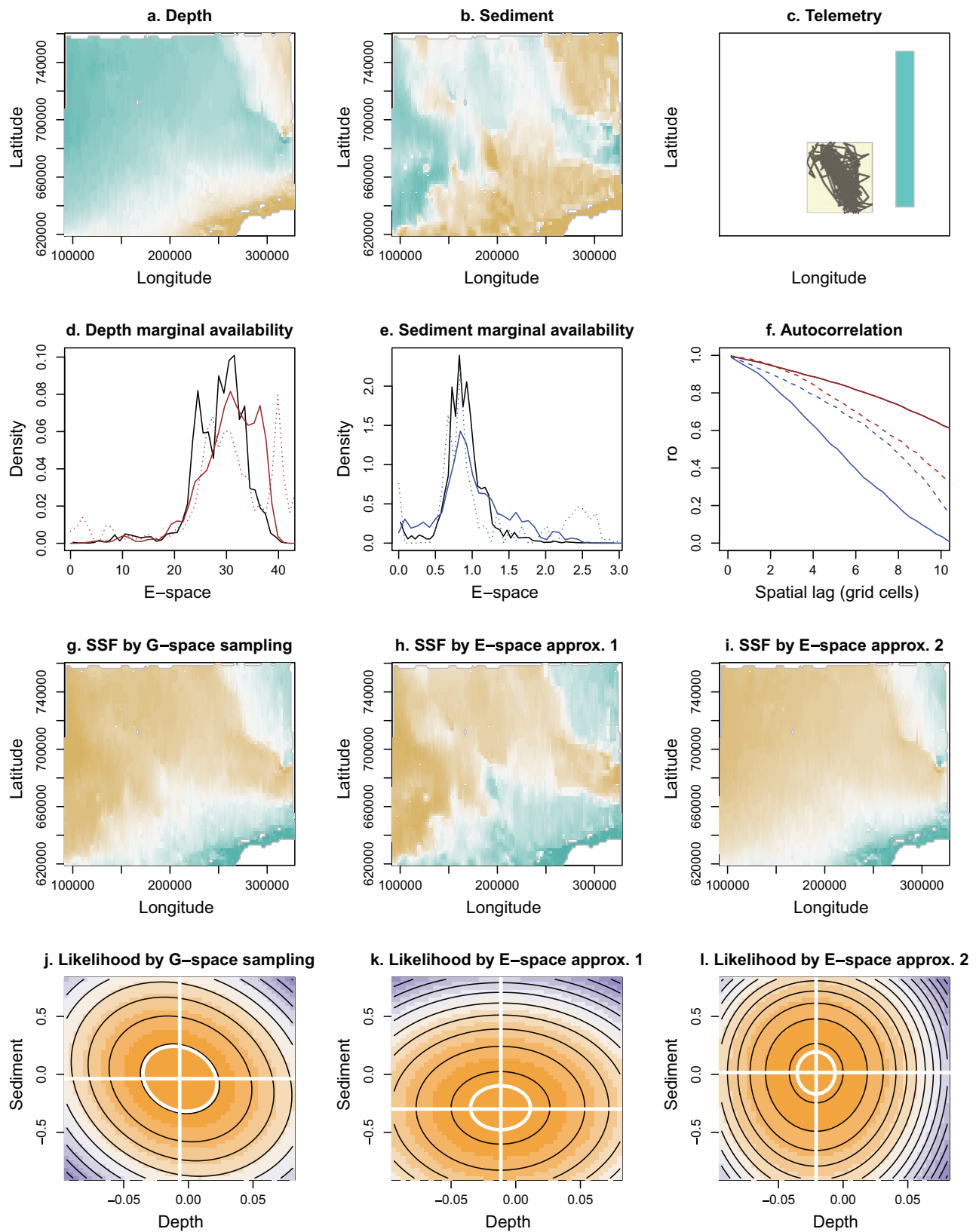


Figure 4: Step-selection analysis of harbor seal telemetry data by *G*-space sampling and *E*-space approximation. Two environmental variables (*a* and *b*) were used to characterize the seascape in which harbor seal GPS tracking data were collected (*c*). Step selection by *E*-space approximation used two different rectangular areas for learning about the environment (shown as yellow and blue in plot *c* and giving rise to *E*-space approximation 1 and 2, respectively). Each variable was summarized in terms of its marginal availability (*d* and *e*) and spatial autocorrelation (*f*). In plots *d*, *e*, and

and is therefore frequently used for applied analyses (for a review, see Thurfjell et al. 2014). The estimates of the parameter γ stabilize as the number V of controls selected becomes large, subject to data storage and computational speed capacity. Indeed, if V tends to infinity (e.g., $V > 100$), the likelihood can be replaced by the simpler form

$$l_{\text{CLL}}(\mathbf{s}_j | \mathbf{s}_{j-1}) = \log(w(\mathbf{x}_j)) - \log\left(\sum_{v=1}^V w(\mathbf{z}_v)\right). \quad (22)$$

An alternative approach to obtaining a step-selection likelihood without the need to sample control points is to notice that the sum in equation (22) is proportional to the expected value of the step-selection function in the vicinity of the point \mathbf{s}_{j-1} . Therefore, given an exact PDF of the availability of habitats around the preceding point (i.e., $f_{\mathbf{z}|\mathbf{s}_{j-1}}$) we could rewrite equation (21) as

$$l_{\text{CLL}}(\mathbf{x}_j | \mathbf{s}_{j-1}) = \log(w(\mathbf{x}_j)) - \log\left(\int_E w(\mathbf{z}) f_{\mathbf{z}|\mathbf{s}_{j-1}} d\mathbf{z}\right). \quad (23)$$

However, in general an exact form of $f_{\mathbf{z}|\mathbf{s}_{j-1}}$ will not be available for any given point \mathbf{s}_{j-1} . We can, instead, approximate this function by using the habitat characteristics \mathbf{x}_{j-1} at the point \mathbf{s}_{j-1} , so that $f_{\mathbf{z}|\mathbf{s}_{j-1}} \cong f_{\mathbf{z}|\mathbf{x}_{j-1}}$.

This approximation requires knowledge of the unconditional availability of habitats and the spatial autocorrelation in each environmental variable. If these assumptions hold (see below), then the log likelihood in equation (23) can be rewritten as

$$l_{\text{CLL}}(\mathbf{x}_j | \mathbf{x}_{j-1}) = \log(w(\mathbf{x}_j)) - \log\left(\int_E w(\mathbf{z}) f_{\mathbf{z}|\mathbf{x}_{j-1}} d\mathbf{z}\right). \quad (24)$$

Using the results on conditional availability developed in earlier sections, we show in supplement 3 that the integral involved in this log likelihood has a closed-form solution. Hence, equation (20) can be obtained analytically as

$$l_{\text{CLL}} = \sum_{j=1}^J \log(w(\mathbf{x}_j)) - \sum_{j=1}^J \log\left(\frac{1}{f_{\mathbf{x}_{j-1}}} \prod_{k=1}^K \sum_{l=1}^L \sum_{m=1}^L \Psi_{k,l,m} f_{l,x_k} \Theta(\gamma_{1,k}, \gamma_{2,k}, \mu_{m,k}, \sigma_k)\right), \quad (25)$$

where $\Theta(\gamma_{1,k}, \gamma_{2,k}, \mu_{m,k}, \sigma_k)$ is an algebraic function of parameters pertaining to habitat preference and availability. This analytical expression can prove useful in studies with

imperfect or irregular environmental data sets. For example, a number of modern telemetry tags, particularly in the marine environment, collect in situ environmental data in addition to location information (Berlinger et al. 2004; Biuw et al. 2007; Hooker et al. 2008; Ericsson et al. 2015). For environmental variables that are measured only at the location of the animal, our model could provide a useful description of habitat availability for locations that were potentially accessible but not visited by the animal. If some representative segments of space have been independently surveyed to allow us to characterize the statistical properties of the distribution of these variables (even if high-resolution covariate layers are not available exactly in the vicinity of the telemetry data), then these can supplement the analysis. In addition, for temporally irregular data, our model's mobility kernel can be used to give a varying degree of accessibility, depending on the time interval between locational fixes (a problem also considered in Johnson et al. 2008, 2013). This flexibility can be extended to account for different modes of mobility (e.g., as a result of diurnal activity patterns).

Two key assumptions are required to ensure the modeling approximation in E -space provides an adequate approximation to conditional habitat availability:

Representativeness assumption. The data from which the unconditional habitat distribution is derived must be representative of the landscape on which the method is to be applied. Therefore, we require the marginal distributions to be accurate, even if the environmental layers are not known exactly. In a sufficiently large spatial arena, this assumption can be satisfied without the need for high-resolution data. Any large point sample will suffice, as long as it is collected systematically or randomly from the region of interest or a region with similar properties.

Stationary autocorrelation function assumption. The shape of the autocorrelation function must be the same between the regions used for training the approximation and the geographical region of application. This assumption can be satisfied without the need for spatially expansive data. A single high-resolution transect that manages to capture the form of autocorrelation will suffice.

As a first practical illustration of this approach, we conducted a comparison between the sampling and modeling approximations (i.e., eqq. [21] and [25], respectively) on a real telemetry data set (fig. 4) collected from individual

f , brown color is used for depth, and blue is used for sediment. The solid lines correspond to the yellow box in plot c , and dotted lines correspond to the extrapolation in the blue rectangle in plot c . The solid black curves in d and e represent the actual frequency of depth and sediment values in the yellow rectangle. The third row of plots shows the maps of relative preference derived from each step-selection analysis, specifically, the G -sampling approach (g); E -space approximation 1, using the yellow rectangle (h); and E -space approximation 2, using the blue rectangle (i). The final row of plots shows the likelihood profiles in 2D parameter space derived from each of the three analyses: G -space sampling (j), E -space approximation 1 (k), and E -space approximation 2 (l). The white crosshairs indicate maximum likelihood parameter estimates, accompanied by asymptotic 95% confidence ellipses (also drawn in white). The coloration from purple to brown reflects increasing likelihood for different parameter combinations. SSF = step-selection function.

harbor seals (*Phoca vitulina*), off the northern coast of the Netherlands. We used a simple data set of two environmental covariates corresponding to bathymetry (fig. 4a) and the percentage of silt in the sediment (fig. 4b). We selected time intervals between the pairs of successive observations in the data set to be less than 24 h and sub-sampled from the data set (taking 1 out of every 20 consecutive pairs of locations) to ensure that the successive pairs in the data set were serially independent. The value of the parameter of the mobility kernel $\omega = 2.58$ (in units of grid cell lengths) was derived directly from the data as the standard deviation of the Rayleigh distribution (calculated as $(2 \text{ var}(|\Delta \mathbf{s}|)/(4 - \pi))^{1/2}$, where $|\Delta \mathbf{s}|$ are the observed step lengths in the data). The sampling approximation used 200 control points for each stratum (i.e., combined with each pair of successive locations). The controls for the sampling approach were selected using Rayleigh step lengths with a uniformly random direction on the circle. The modeling approximation used the same Rayleigh distribution and covariate information originating either from a box enclosing the telemetry data (the yellow box in fig. 4c) or from a strip of the sea that was outside the telemetry set (the blue rectangle in fig. 4c). This comparison allowed us to explore the sensitivity of parameter estimates and spatial predictions to changes in habitat structuring (i.e., violations of the two assumptions of representativeness and stationary autocorrelation). The two regions differed in their area, shape, and location. The elongated shape of the blue region precluded averaging over the strong anisotropy in the environment. These differences potentially reduced the representativeness of the blue region.

To visualize the differences in habitat composition between the two boxes, we plotted the actual frequency of sea depths and sediment values (the black curves in fig. 4d and 4e, respectively) against the modeled availability of those two variables within the yellow box (solid brown line for depth in fig. 4d and solid blue line for sediment in fig. 4e) and within the blue rectangle (dotted brown line for depth in fig. 4d and dotted blue line for sediment in fig. 4e). In addition, we explored differences in spatial autocorrelation between the yellow and blue rectangles (fig. 4f). We visualized the results of the analysis in geographic as well as parameter space. The geographic visualization for each of the three analyses looked at the value of the step-selection function in each of the map's pixels (fig. 4g–4i). These values can be interpreted as a relative measure of preference in comparison to nearby cells. The parameter space visualization examined the estimates and 95% confidence ellipses generated by each of the three methods for the coefficients of the two environmental variables (fig. 4j–4l).

The above comparison leads to the following conclusions. When the training data are obtained from the region of interest, the modeling approximation gives spatial re-

sults similar to that of geographic sampling (compare fig. 4g with fig. 4h), and the 95% confidence ellipses overlap (fig. 4j, 4k). Using training data outside the region of interest, so that the assumptions of representativeness and stationary autocorrelation are less faithfully preserved (see the diagnostics in fig. 4d–4f), may result in differences between the two approaches (compare fig. 4g with fig. 4h, 4i). Yet the parameter estimates remain within plausible ranges for this particular problem (compare fig. 4j with 4l). Thus, while the method gives plausible parameter estimates outside the range of the data, we can conclude that there are increasing differences as the training data deviate from the region of interest. Therefore, although the proposed approach of modeling spatial accessibility in *E*-space is not a substitute for direct sampling of controls in *G*-space, it is a method that can provide informative results when environmental data are sparse or of limited geographic coverage.

We note that the above application uses only the most rudimentary form of step-selection estimation. As part of future work, it would be interesting to explore how the above likelihood could be extended to perform simultaneous estimation of movement characteristics and habitat preferences (e.g., Forester et al. 2009; Avgar et al. 2016). Additionally, the approach taken here assumes independence of the conditional availability of the different environmental variables. It is reassuring that the approximation above works reasonably well despite this simplification, given that depth and sediment were moderately cross correlated ($r^2 = 0.64$).

Theoretical Example: The Effects of Spatial Autocorrelation on the Fitness of Territorial Animals

To illustrate how our approach can be used to derive theoretical results, we consider the effects of spatial autocorrelation on the average fitness of populations of animals holding territories of identical size. To derive some useful baseline results, we begin by assuming that space is saturated by territories (i.e., no apparent habitat preference) but relax this assumption later. We consider a habitat described by a single covariate (e.g., a single resource) where z refers to the value of the resource at the territory's centroid and x refers to values of the resource found elsewhere within the territory. The fitness contribution of a habitat (i.e., a particular value of the resource x) is denoted by F_x , such that $F_x = a_0 + a_1x$ for some coefficients a_0 and a_1 . We require fitness to be negative when the resource x is low (i.e., $a_0 < 0$) and to have a positive relationship with increasing resource values (i.e., $a_1 > 0$). This example can be extended (with more elaborate algebra but no loss of analytical tractability) by introducing several covariates,

possibly having nonmonotonic contributions to fitness (for more complex extensions, see Matthiopoulos et al. 2015).

Fitness in the Absence of Habitat Preference

When a population lives in a landscape of very low spatial autocorrelation (LO), all habitats (i.e., all values of the resource) should, on average, be present within each territory in proportion to their broader availability (f_x). In other words, the composition of each territory—and therefore also the fitness afforded by each territory—will be representative of the broader landscape:

$$F_{\text{LO}} = \int_E F_x f_x dx. \quad (26)$$

In supplement 4, we show that this simplifies to

$$F_{\text{LO}} = a_0 + a_1 \sum_{l=1}^L \psi_l \mu_l. \quad (27)$$

Thus, the fitness of the organism is derived from a weighted sum of the means of the Gaussian mixture describing habitat availability; this sum is equal to the mean of x . In other words, if \bar{x} is the average value of the resource in the environment, under low spatial autocorrelation we get the intuitive result, corresponding to perfect mixing:

$$F_{\text{LO}} = a_0 + a_1 \bar{x}. \quad (28)$$

More generally, for animals living in more realistic landscapes with some spatial autocorrelation, the expected fitness for a territory centered at habitat z will be

$$F(z) = \int_E F_x f_{x|z} dx. \quad (29)$$

In supplement 4, we show that this simplifies to

$$F(z) = a_0 + \frac{a_1}{f_z} \sum_{l=1}^L \sum_{k=1}^L f_{k,z} \Psi_{l,k} \mu_l. \quad (30)$$

Incidentally, a comparison between equations (30) and (27) implies that in the case of perfect mixing, the joint weights of the habitat availability formula take the form

$$\Psi_{l,k} = \psi_l \psi_k. \quad (31)$$

We explore the difference between the average fitness, across the landscape, in the absence and presence of spatial autocorrelation:

$$\bar{F} - \bar{F}_{\text{LO}} = \int_z F(z) f_z dz - F_{\text{LO}}, \quad (32)$$

which rearranges to

$$\bar{F} - \bar{F}_{\text{LO}} = a_1 \left(\sum_{l=1}^L \sum_{k=1}^L \Psi_{l,k} \mu_l - \sum_{l=1}^L \psi_l \mu_l \right). \quad (33)$$

We note that $\sum_{k=1}^L \Psi_{l,k} = \psi_l$, which gives $\bar{F} - \bar{F}_{\text{LO}} = 0$. This makes intuitive sense and has been anticipated by

previous work (Barraquand and Murrell 2013; fig. 1). In an autocorrelated landscape tessellated by territories, some individuals will benefit from aggregations of high resource, while others will lose out by having their territories at resource troughs.

Fitness in the Presence of Habitat Preference

We now relax the assumption of uniform placement of territories by introducing a model of heterogeneity that is affected by an underlying habitat preference function $w(z) = \exp(b_0 + b_1 z)$ to the single resource z . We assume for this exploration that habitat preference operates on the selection of the territory centroid but that the organism uses parts of the territory uniformly. The average fitness afforded by the environment to a population of such animals would therefore be

$$\bar{F} = A^{-1} \int_E w(z) f_z F(z) dz, \quad (34)$$

where $F(z)$ is the fitness associated with a territory centred at habitat z (as defined in eq. [30]) and $A = \int_E w(z) f_z dz$ is a normalizing constant for the preference function. In supplement 5, we show that this expression can be simplified to

$$\bar{F} = a_0 + a_1 \frac{\sum_{l=1}^L \sum_{k=1}^L \Psi_{l,k} \mu_l \exp(b_1 \mu_k)}{\sum_{k=1}^L \psi_k \exp(b_1 \mu_k)}. \quad (35)$$

This expression describes the average population fitness as a function of unitary fitness parameters (a_0, a_1), marginal resource availability (expressed by the parameters ψ_k, μ_k), spatial autocorrelation (contained in the joint weights $\Psi_{l,k}$), and the selectivity (b_1) in choosing the centroid of a territory. For any particular landscape, the joint weights will generally need to be derived using methods such as the ones presented in supplement 1, but we can simplify our investigation by comparing the two extremes of very low and very high spatial autocorrelation. The case of very low spatial autocorrelation is represented by equation (28). The case of very high spatial autocorrelation can be emulated by setting

$$\Psi_{l,k} = \begin{cases} \psi_k & \text{if } l = k, \\ 0 & \text{otherwise.} \end{cases} \quad (36)$$

Within the expression for joint habitat availability (eq. [15]), this works by accumulating a high probability density close to the line of slope 1 (creating joint PDFs similar to those in figs. 2a or 3a), hence enforcing the probability of encountering similar values of z from an animal's current position. Via this simplification, the fitness equation becomes

$$\bar{F} = a_0 + a_1 \frac{\sum_{k=1}^L \psi_k \mu_k \exp(b_1 \mu_k)}{\sum_{k=1}^L \psi_k \exp(b_1 \mu_k)}. \quad (37)$$

Subject to the assumption of high spatial autocorrelation, we proceed to explore the behavior of this function by varying the overall resource richness (related to the mean value of the distribution of the available resource) and heterogeneity (the variability of the distribution of the available resource) of the landscape. To do this in a tractable way, we envisage an environmental space that is constructed of L equally spaced and equally weighted Gaussian components (fig. 4a). The mean value (\bar{z}) of the mixture determines overall resource richness, and the number (L) of individual components, equally split on either side of the mean, represents heterogeneity. We take the spacing between adjacent Gaussian means μ_k and μ_{k+1} to be equal to σ , the standard deviation of each of the Gaussian components. This assumption tends to give approximately uniform distributions of the resource in E -space (thick gray curve in fig. 5a). Note, however, that the distribution of the resource in G -space will be heterogeneous. This simplified representation of the environment yields

$$\psi_k = \frac{1}{L} \quad \text{and} \quad \mu_k = \bar{x} + \sigma \left(k - \frac{L+1}{2} \right). \quad (38)$$

For fitness (eq. [37]), these simplifications imply that

$$\bar{F} = a_0 + a_1 \left(\bar{x} - \sigma \frac{L+1}{2} + \sigma \sum_{k=1}^L k \theta_k \right), \quad (39)$$

where θ_k are weights driven by the habitat selectivity parameter (b_1):

$$\theta_k = \frac{\exp(b_1 k \sigma)}{\sum_{k=1}^L \exp(b_1 k \sigma)}. \quad (40)$$

The parameter b_1 represents the ability of an organism to express preference for placing its territory centroid at high-resource locations. In a very small population it is expected that b_1 will be very large, because when unobstructed by conspecifics, an organism will be able to place its territory at the peak of resource concentration. If the landscape is completely saturated, so that space is covered by territories, apparent selectivity will move toward zero. If the centroid of a territory serves a life-history function that is mutually exclusive to resource acquisition (e.g., a ground nest that needs to be placed within high but inedible grass), the apparent selectivity for the resource may give negative values

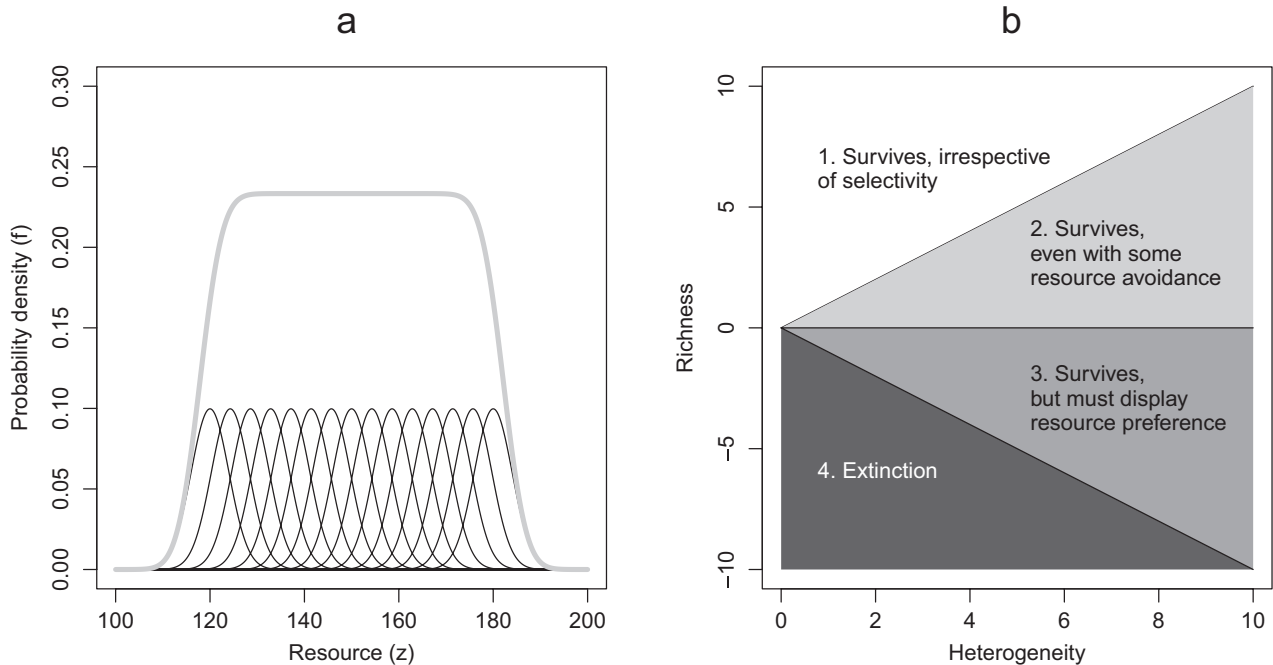


Figure 5: *a*, Example of a uniform marginal distribution in environmental space in one resource variable constructed from the superposition of equally weighted Gaussian components (each having a standard deviation of σ , which is also used as the placement distance between successive components). This arrangement allows us to reduce the description of E -space to the two traits of resource richness (the position of the mixture along the resource axis) and heterogeneity (the dispersion of the mixture), driven here by the number of participating Gaussian components. *b*, Summary of findings in the graphical plane of resource richness and heterogeneity in the case of a highly autocorrelated resource distribution in G -space. Four regions arise indicating population viability depending on the habitat selectivity displayed by the individuals making up the population.

of b_1 . We therefore consider three scenarios that give rise to important boundaries in the richness/heterogeneity plane (see the collected results in fig. 5b).

Scenario 1: Fitness is negative, even when the population displays high selectivity (i.e., very high values of b_1). This corresponds to environments where even small populations, with the ability to concentrate around the best available habitat, become extinct. The scenario of very high selectivity is written

$$\lim_{b_1 \rightarrow \infty} \sum_{k=1}^L k\theta_k = L. \quad (41)$$

Using equation (39), the mathematical condition for negative fitness is

$$\bar{x} < -\frac{a_0}{a_1} - \sigma \frac{L-1}{2}. \quad (42)$$

Scenario 2: Fitness is zero in a saturated population that has completely filled up space with territories, giving the impression of no selectivity ($b_1 = 0$). In this scenario,

$$\sum_{k=1}^L k\theta_k = \frac{L+1}{2}. \quad (43)$$

Using equation (39), the mathematical condition for zero fitness is

$$\bar{x} = -\frac{a_0}{a_1}. \quad (44)$$

Scenario 3: Fitness is positive even when the organism avoids high concentrations of the resource (i.e., for very large negative values of b_1). This scenario implies

$$\lim_{b_1 \rightarrow -\infty} \sum_{k=1}^L k\theta_k = 1. \quad (45)$$

The mathematical condition for positive fitness is

$$\bar{x} > -\frac{a_0}{a_1} + \sigma \frac{L-1}{2}. \quad (46)$$

For graphical convenience, we define resource richness in relation to fitness parameters. Furthermore, we define environmental heterogeneity in terms of the number and dispersion of Gaussian components used to describe the range of resource values in environmental space:

$$\text{richness} = \bar{x} + \frac{a_0}{a_1}, \quad \text{heterogeneity} = \sigma \frac{L-1}{2}. \quad (47)$$

These definitions are biologically intuitive. In particular, this index of resource richness takes the value zero when experiencing the average availability of the resource barely allows an organism to survive. The index of heterogeneity becomes zero when the minimum number of Gaussian components ($L = 1$) is used to describe the environment.

Recasting the conditions in equations (42), (44), and (46) with the aid of these new definitions gives us the combined results in figure 5b, which enable us to summarize population viability in terms of resource richness and heterogeneity in the case of a highly spatially autocorrelated resource distribution. The figure illustrates that spatial heterogeneity expands the ability of a population to persist and quantifies the thresholds of extinction and persistence in scale-independent coordinates (thanks to the scalings of richness and heterogeneity emerging from this analysis in the form of eq. [47]).

Discussion

Assumptions about habitat accessibility can drastically affect the predictions of population models in space and time. Models that assume either perfectly mixed or completely sessile populations are liable to err for different reasons. We therefore need a theoretical and quantitative framework for describing habitat accessibility. The two basic determinants of habitat accessibility from any given geographical position are the speed with which organisms move and the spatial scales over which the environment varies. Starting from this fact, we have derived a compact expression for conditional habitat availability (eq. [14]) in environmental (or niche) space. This was achieved by describing the availability of all habitats from the vantage point of any given habitat, using functions of distance (reflecting both mobility and spatial autocorrelation). The benefits of this framework are both conceptual and applied. From a conceptual viewpoint, this work can be seen as a contribution to the historical and ongoing discussions about scale in ecology (Wiens 1989; Levin 1992; Schneider 2001; Gurarie and Ovaskainen 2011). Our work offers a quantitative formalization of the interplay between the scale of spatial autocorrelation and the scale of organism mobility over particular time frames. A correctly scaled view of accessibility can quantify relationships that would not have been evident via qualitative arguments alone. The collected findings in figure 5b illustrate how the relative scales of mobility and environmental heterogeneity can fundamentally alter the fitness that a landscape can afford a population. Regions 2 and 3 in figure 5 have a novel biological interpretation in which habitat selectivity changes the sign of population growth relative to an assumption of no selectivity. Region 3, in particular, is the direct result of animals in the population being able to aggregate at hot spots of resource distribution and, hence, experience higher-than-average fitness compared with a nonspatial model assuming perfect mixing.

Many of these insights would be achievable on a particular landscape by means of intensive sampling of space; however, our framework offers a flexible abstraction of

species-habitat interaction based solely on the statistical properties of the system. This allows us to work in environmental spaces and produce generalizable results, applicable to different landscapes with similar landscape compositions. Similar models can be derived through moment equation modeling incorporating at the same time the spatial autocorrelation in environmental variation and the movement processes through kernels (as done here) as well as deriving the environment-organism covariance from the interaction of dispersal, demography, and environmental structure (Murrell and Law 2000; Bolker 2003; North et al. 2011). However, these methods require more complex analytical formulas, place their emphasis on population dynamics, and do not operate explicitly in *E*-spaces.

A recent review of species range models (Singer et al. 2016) discusses how mechanistic approaches can be used to enhance the predictive ability of correlative models of species' distribution. Hence, the present work can be used to increase the mechanistic content of correlative models but may also be used for expedient calculation in fully mechanistic approaches. The mechanistic content of our approach can be increased to account for features of movement. For example, the variance of our movement kernel can be assumed to depend on the properties of the local habitat, to account for reductions in mobility due to difficult substrates. Accessibility may also be thought of in larger spatiotemporal scales from the viewpoint of dispersal processes. SDMs based on snapshots of species abundance assume that, over many generations, dispersal events that are hard but not impossible will have been made at some point, allowing the species to occupy all the locations that have suitable habitat. Our framework can accommodate both timescales of dispersal by varying how far out in the tails of the availability kernel we sample. Indeed, that can become an index of how fast a species can fill up the landscape, which becomes relevant as we try to figure out whether species will be able to shift their geographic ranges fast enough to keep up with climate change (Parmesan and Yohe 2003).

Acknowledgments

The impetus for this work came from discussions with Devin Johnson, Otso Ovaskainen, and Paul Blackwell. Sophie Brasseur provided us with data for the seal case study. The associate editor, Ben Bolker, made pivotal suggestions that influenced the structure of the manuscript. We appreciate the help and encouragement of two anonymous reviewers.

Data and Code Availability

Archival files for the R code and data used in the article can be downloaded from <https://zenodo.org/record/3479825#.XZ-MDSV7knc>.

Literature Cited

- Aarts, G., J. Fieberg, S. Brasseur, and J. Matthiopoulos. 2013. Quantifying the effect of habitat availability on species distributions. *Journal of Animal Ecology* 82:1135–1145.
- Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos. 2008. Estimating space use and environmental preference from telemetry data. *Ecography* 31:140–160.
- Avgar, T., J. R. Potts, M. A. Lewis, and M. S. Boyce. 2016. Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution* 7:619–630.
- Barraquand, F., and D. J. Murrell. 2013. Scaling up predator-prey dynamics using spatial moment equations. *Methods in Ecology and Evolution* 4:276–289.
- Beringer, J., J. J. Millsbaugh, J. Sartwell, and R. Woeck. 2004. Real-time video recording of food selection by captive white-tailed deer. *Wildlife Society Bulletin* 32:648–654.
- Beyer, H. L., 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.
- Biuw, M., L. Boehme, C. Guinet, M. Hindell, D. Costa, J.-B. Charrassin, F. Roquet, et al. 2007. Variations in behavior and condition of a Southern Ocean top predator in relation to in situ oceanographic conditions. *Proceedings of the National Academy of Sciences of the USA* 104:13705–13710.
- Bolker, B. M. 2003. Combining endogenous and exogenous spatial variability in analytical population models. *Theoretical Population Biology* 64:255–270.
- Blackwell, P. G. 1997. Random diffusion models for animal movement. *Ecological Modelling* 100:87–102.
- Colwell, R. K., and T. F. Rangel. 2009. Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the USA* 106:19651–19658.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46.
- 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.
- Ericsson, G., H. Dettki, W. Neumann, J. M. Arnemo, and N. J. Singh. 2015. Offset between GPS collar-recorded temperature in moose and ambient weather station data. *European Journal of Wildlife Research* 61:919–922.
- Fagan, W. F., E. Gurarie, S. Bewick, A. Howard, R. S. Cantrell, and C. Cosner. 2017. Perceptual ranges, information gathering, and foraging success in dynamic landscapes. *American Naturalist* 189:474–489.
- Feller, W. 1966. *An introduction to probability theory and its applications*. Vol. 2. Wiley, New York.
- Forester, J. D., H. K. Im, and P. J. Rathouz. 2009. Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology* 90:3554–3565.
- Fortin, D., H. L. Beyer, M. S. Boyce, D. W. Smith, T. Duchesne, and J. S. Mao. 2005. Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. *Ecology* 86:1320–1330.

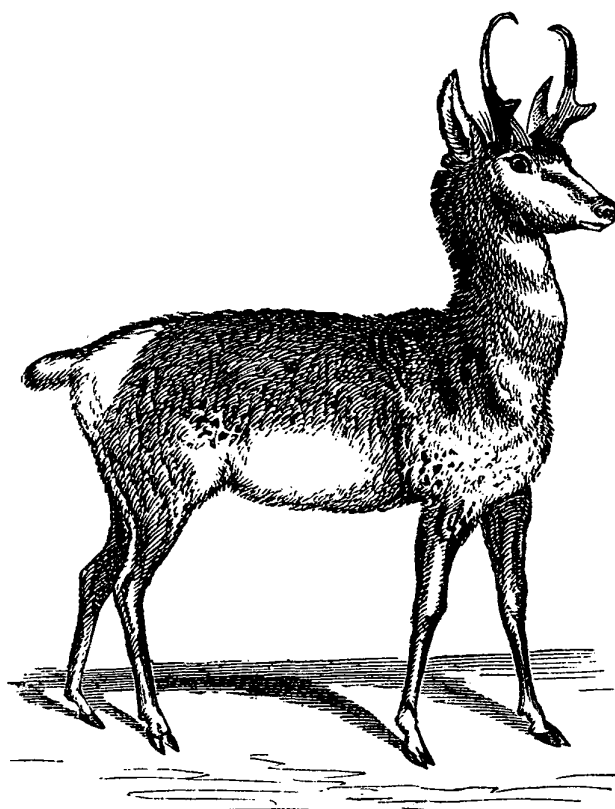
- Fraley, C., A. E. Raftery, T. B. Murphy, and L. Scrucca. 2012. *mclust* version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report no. 597, Department of Statistics, University of Washington.
- Fraley, C., A. E. Raftery, and R. Wehrens. 2005. Incremental model-based clustering for large datasets with small clusters. *Journal of Computational and Graphical Statistics* 14:1–18.
- Gurarie, E., and O. Ovaskainen. 2011. Characteristic spatial and temporal scales unify models of animal movement. *American Naturalist* 178:113–123.
- 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.
- Hirzel, A. H., and G. Le Lay. 2008. Habitat suitability modeling and niche theory. *Journal of Applied Ecology* 45: 1372–1381.
- Hooker, S. K., S. G. Heaslip, J. Matthiopoulos, O. Cox, and I. Boyd. 2008. Data sampling options for animal-borne video cameras: considerations based on deployments with Antarctic fur seals. *Marine Technology Society Journal* 42:65–75.
- Hooten, M. B., D. S. Johnson, B. T. McClintock, and J. M. Morales. 2017. *Animal movement: statistical models for telemetry data*. CRC, Boca Raton, FL.
- Horne, J. S., E. O. Garton, and J. L. Rachlow. 2008. A synoptic model of animal space use: simultaneous estimation of home range, habitat selection, and inter/intra-specific relationships. *Ecological Modelling* 214:338–348.
- Hughes, B. 1995. *Random walks and random environments*. Vol. 1. Clarendon, Oxford.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology* 22:415–427.
- Joe, H. 2014. *Dependence modeling with copulas*. Chapman & Hall/CRC, Boca Raton, FL.
- Johnson, D. S., M. B. Hooten, and C. E. Kuhn. 2013. Estimating animal resource selection from telemetry data using point process models. *Journal of Animal Ecology* 82:1155–1164.
- Johnson, D. S., D. L. Thomas, J. M. Ver Hoef, and A. Christ. 2008. A general framework for the analysis of animal resource selection from telemetry data. *Biometrics* 64:968–976.
- Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943–1967.
- MacArthur, R. H. 1968. The theory of the niche. Pages 159–179 *in* R. C. Lewontin, ed. *Population biology and evolution*. Syracuse University Press, Syracuse, NY.
- Manly, B., L. McDonald, D. Thomas, T. McDonald, and W. P. Erickson. 2004. *Resource selection by animals—statistical design and analysis for field studies*. Chapman & Hall, London.
- Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. *Ecological Modelling* 159:239–268.
- Matthiopoulos, J., J. Fieberg, G. Aarts, H. L. Beyer, J. M. Morales, and D. T. Haydon. 2015. Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs* 85:413–436.
- Matthiopoulos, J., C. Field, and R. MacLeod. 2019. Predicting population change from models based on habitat availability and utilization. *Proceedings of the Royal Society B* 286:20182911.
- Matthiopoulos, J., M. Hebblewhite, G. Aarts, and J. Fieberg. 2011. Generalized functional responses for species distributions. *Ecology* 92:583–589.
- McInerney, G. J., and R. S. Etienne. 2013. “Niche” or “distribution” modeling? a response to Warren. *Trends in Ecology and Evolution* 28:191–192.
- McLoughlin, P. D., D. W. Morris, D. Fortin, E. Vander Wal, and A. L. Constantini. 2010. Considering ecological dynamics in resource selection functions. *Journal of Animal Ecology* 79:4–12.
- Murrell, D. J., and R. Law. 2000. Beetles in fragmented woodlands: a formal framework for dynamics of movement in ecological landscapes. *Journal of Animal Ecology* 69:471–483.
- North, A., S. Cornell, and O. Ovaskainen. 2011. Evolutionary responses of dispersal distance to landscape structure and habitat loss. *Evolution* 65:1739–1751.
- Northrup, J. M., M. B. Hooten, C. R. Anderson, and G. Wittemyer. 2013. Practical guidance on characterizing availability in resource selection functions under a use-availability design. *Ecology* 94:1456–1463.
- Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37–42.
- Paton, R. S., and J. Matthiopoulos. 2016. Defining the scale of habitat availability for models of habitat selection. *Ecology* 97:1113–1122.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ.
- Randin, C. F., T. Dirnbock, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33:1689–1703.
- R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Sawo, F., D. Brunn, and U. D. Hanebeck. 2006. Parameterized joint densities with Gaussian and Gaussian mixture marginals. *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy.
- Schneider, D. S. 2001. The rise of the concept of scale in ecology: the concept of scale is evolving from verbal expression to quantitative expression. *BioScience* 51:545–553.
- Signer, J., J. Fieberg, and T. Avgar. 2019. *Animal movement tools (amt): R-package for managing tracking data and conducting habitat selection analyses*. *Ecology and Evolution* 9:880–890.
- Sinclair, S. J., M. D. White, and G. R. Newell. 2010. How useful are species distribution models for managing biodiversity under future climates? *Ecology and Society* 15:8.
- Singer, A., K. Johst, T. Banitz, M. Fowler, J. Groeneveld, A. Gutiérrez, F. Hartig, et al. 2016. Community dynamics under environmental change: how can next generation mechanistic models improve projections of species distributions? *Ecological Modelling* 326:63–74.
- Sklar, L. 2015. Philosophy of statistical mechanics. *In* Edward N. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2015/entries/statphys-statmech/>.
- Soberón, J., and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the USA* 106:19644–19650.
- Therneau, T. M., and T. Lumley. 2019. *Survival analysis*. <https://github.com/therneau/survival>
- Thurfjell, H., S. Ciuti, and M. S. Boyce. 2014. Application of step-selection functions in ecology and conservation. *Movement Ecology* 2:4. <https://doi.org/10.1186/2051-3933-2-4>.

- Wenger, S. J., and J. D. Olden. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* 3:260–267.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3:385–397.
- Zurell, D., F. Jeltch, C. Dormann, and B. Schröder. 2009. Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography* 32:733–744.

References Cited Only in the Online Enhancements

- Wood, S. N. 2006. *Generalized additive models: an introduction* with R. Chapman & Hall/CRC, Boca Raton, FL.

Associate Editor: Benjamin M. Bolker
Editor: Alice A. Winn



“The horn, when shed, seems to be a mass of agglutinated hairs enclosed by a substance resembling whalebone in appearance; some of the hairs, however, never amalgamating with the horn, but retaining their natural condition, and, passing entirely through the horn, will be found protruding on the inside and outside of the horn.” From “The Prong-Horn Antelope” by W. J. Hays (*The American Naturalist*, 1868, 2:131–133).