PUBLISHED AS: Matthiopoulos, J., J. Fieberg, G. Aarts, F. Barraquand and B.E. Kendall. 2020. Within Reach? Habitat Availability as a Function of Individual Mobility and Spatial Structuring. American Naturalist 195: 1009-1026. DOI: 10.1086/708519

Within reach? Habitat availability as a function of individual mobility and spatial structuring

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 (SDMs) 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.

23 Keywords: conditional availability, Gaussian mixtures, habitat selection, resource selection, step
24 selection functions, species distribution modeling
25

1. 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, McLaughlin et al. 2010, Sinclair et al. 2010, Matthiopoulos et al. 2011, Wenger & 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 & Matthiopoulos 2016). Most of these papers 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 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 paper 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 (section 2), followed by a mathematical model for conditional habitat availability (section 3), 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. In section 4, 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. In section 5, 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, Ciuti & Boyce 2014, Hooten et al. 2017). In section 6 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.

2. *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 & LeLay 2008, Colwell & Rangel 2009, Elith & 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, each dimension representing a biotic or abiotic environmental variable, i.e.

99 a continuous, discrete or qualitative random variable representing a condition (e.g. pH,
100 temperature, sea depth), resource (e.g. soil nutrients, prey, breeding sites) or risk (e.g. predators,
101 pollution). *E*-space can be considered identical to niche-space, as originally conceived by
102 Hutchinson (1957) and MacArthur (1968) although, as extensively argued in the modern
103 literature (Soberón & Nakamura 2009, Peterson et al. 2011, McInerny & Etienne 2013,
104 Matthiopoulos et al. 2015), statistical habitat preference models currently fitted in *E*-space
105 should not be confused with the niche objects as envisaged by these pioneering thinkers.
106 Several papers (Aarts et al 2008, Matthiopoulos et al. 2011, Matthiopoulos et al. 2015 and
107 references therein), conceptualize *habitat* as a point **x** in *E*-space, the combination
108 $\mathbf{x} = \{x_1,\ldots,x_K\}$ of specific values for K environmental variables (e.g. geomorphology and
109 climate variables combining into the characteristic makeup of, say, "polar habitat"). Elsewhere,
110 and in colloquial use, "habitat" has been described in a species-dependent way, as the region in
111 geographical space in which an organism lives (e.g. "polar-bear habitat"). The two definitions
112 are not interchangeable (see Hall et al. 1997). We opt for the former definition because it allows
113 objective comparisons between species and quantitative gradations of suitability. Subject to this
114 definition of habitat, we can introduce the *unconditional availability* ($f_{\mathbf{x}}$) of a particular habitat
115 $\mathbf{x}$ as the relative frequency (i.e. the probability density) with which that habitat occurs across
116 the whole landscape.
117 Data-derived objects recorded in *G*-space are typically complicated and difficult to describe
118 parametrically. For example, describing even a single altitude contour on a map by means of a
119 mathematical formula is a non-trivial task. In contrast, objects in *E*-spaces are generally simpler,
120 as we illustrate in Fig. 1 by visualizing the simple case of a single environmental variable (i.e.,
121 one dimensional *E*-space) measured in a linear region (i.e., one-dimensional *G*-space). The way
122 in which a multimodal variable in *G*-space (Fig. 1b) gives rise to a much simpler (in this case,
123 unimodal) frequency histogram in *E*-space (Fig. 1a) 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, Fraley et al. 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

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

where $f_{l,\mathbf{x}}$ is the l^{th} component (a *K*-dimensional Gaussian probability density function) 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 eq. (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.

3. Conditional habitat availability

The relative simplicity of *E*-spaces (compared to *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*, i.e. the expected availability of habitat **x** to an organism that is currently located in coordinates **s**, characterized by habitat **z**. Importantly, we seek an expression for conditional availability that is not reliant on a neighborhood in *G*-space defined around a particular location **s**, but rather, on the mixture of habitats typically encountered around a particular habitat **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.

Fig. 1 shows this concept in one spatial dimension and for one environmental variable. In this low-dimensional illustration, the general notion of a habitat **x** is simply a particular value x of the single environmental variable X and the general location **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 probability density function in *E*-space (Fig. 1a).

Subsequently, we focus on all the spatial locations $s_1,\ldots,s_n$ (the dots in Fig. 1b) 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 $s_1,\ldots,s_n$ is illustrated in Fig. 1d 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 time scale

168 (say, a year), the kernel could represent Brownian motion over that time scale. Alternatively, if
169 the study organisms are not free-ranging (e.g., because they must provision offspring located at a
170 central place or because they must actively defend a territory), the kernel can be thought of as the
171 result of an Ornstein–Uhlenbeck process (a random walk with a central tendency – Blackwell
172 1997). An isotropic mobility kernel in any number of spatial dimensions can be recast as a
173 function $h(r)$ that describes the probability of an organism reaching a location at distance *r*
174 away from its current position, over the time period of interest. For data collected infrequently
175 enough that locations can be assumed independent, the kernel can be viewed as determining
176 availability at the home range scale, similar to Horne et al.'s (2008) synoptic model of animal
177 space use. Alternatively, our kernel can be used to model perception range. In particular, Fagan
178 et al. (2017) make the case for mathematical formulations of semi-local perception (an
179 intermediate between the extremes of omniscience and purely local information about habitat)
180 and use Gaussian kernels to describe the diminishing ability of an animal to perceive habitats at
181 greater distances.
182 Since the values of environmental variables in *G*-space are spatially autocorrelated,
183 neighboring points in *E*-space (i.e. similar habitats) will tend to be found close to each other in
184 *G*-space as well. Hence, proximity between locations in *G*-space must translate to proximity
185 between their corresponding habitats in *E*-space. This is schematically represented by the single
186 dashed curve in Fig. 1c peaking in the neighborhood of habitat *z* in environmental space.
187 For those animals viewing the world from the vantage points of habitat *z,* this localized
188 sampling in *G*-space (the dots in Fig. 1f) yields a subjective sample of the values of *X* in *E*-space
189 (solid black line in Fig. 1e). The comparison between the two curves shown in Fig. 1e represents
190 the main concern of this paper: although, globally, the landscape contains habitats whose
191 frequency is described by the light grey curve (the unconditional availability f_z), an organism

192 located in a particular habitat z may be surrounded by a considerably different habitat

193 composition as shown by the dark curve (the conditional availability $f_{x|z}$).

194 To write an expression for conditional availability $f_{x|z}$ of habitat x from a position

195 characterized by habitat *z*, we first consider all pairs of points in *G*-space separated by a distance

196 *r.* If the first point is characterized by habitat *z*, then the probability that the second point is of

197 habitat *x* is denoted by $g_{x|z}(r)$. If the organism is at the first point, then the probability that it can

198 reach across a distance *r* is denoted by $h(r)$. Therefore, the product $g_{x|z}(r)h(r)$ represents the

199 probability that habitat *x* is found at distance *r* from habitat *z*, and that it is accessible by the

200 organism located at habitat *z*. To convert this into a probability density function for conditional

201 availability, irrespective of the distance between two points, we can integrate the product across

202 distances *r*:

203 $$f_{x|z} = \frac{1}{C_E} \int_r g_{x|z}(r)h(r)\ dr\,. \tag{2}$$

204 The probability density $g_{x|z}(r)$ encompasses the spatial autocorrelation of habitats, as well as the

205 overall availability of habitat x, and the probability density $h(r)$ represents limitations in

206 accessibility. Since we require $f_{x|z}$ to be a PDF of habitat availability, eq. (2) contains a

207 normalizing constant that integrates over all target habitats *x*:

208 $$C_E = \int_E \int_r g_{x|z}(r)h(r)\ dr\,dx\,. \tag{3}$$

209 In one spatial dimension the accessibility kernel can be defined as one-dimensional Gaussian and

210 its associated distance function will then be half-normal:

211 $$h(r) = \frac{1}{\omega}\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{r^2}{2\omega^2}\right), \tag{4}$$

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) below). 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, 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 probability density function (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, f_z into its L mixture components, and subsequently combine each component from one marginal distribution (following eq. (1), $f_{l,x}$ for $l = 1,...,L$) 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 $m = 1,...,L$). The weighted sum of these pairwise combinations then yields the joint mixture.

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

where $\psi_{l,m}(r)$ are 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

$$\psi_{l,m}(r) \geq 0, \qquad \sum_{l=1}^{L}\sum_{m=1}^{L}\psi_{l,m}(r) = 1, \qquad \sum_{l=1}^{L}\psi_{l,m}(r) = \psi_m, \qquad \sum_{m=1}^{L}\psi_{l,m}(r) = \psi_l \tag{8}$$

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 non-zero. Indeed, it would be possible to allow the variance-covariance structure of each

255 Gaussian component to be unique. These kinds of approaches would lead to efficiencies in the
256 number of Gaussian components needed. At the other extreme, the approach we have used
257 employs large numbers of Gaussian components, all identical and with zero covariances. As the
258 number of components increases, their variances decrease and so does the influence of the
259 assumption of within-component independence. The decision to employ many, identical and
260 simple mixture components was made for analytical tractability. Given that no covariance is
261 assumed within the individual mixture components, the new weights $\psi_{l,m}(r)$ are the only
262 remaining route of generating a covariance in the joint distribution $g_{x,z}(r)$. In other words, we
263 are seeking to construct a covariance structure in $g_{x,z}(r)$ by reweighting radially symmetric
264 Gaussian components. This will introduce some smoothing in the final result (see numerical
265 examples in section 4).

266 To derive the new weights $\psi_{l,m}(r)$, Sawo et al. (2006) propose an algebraic approach, which
267 unfortunately is quite time-consuming for mixtures of multiple components and often fails to
268 satisfy the positivity requirement (the first condition in eq.(8)). We therefore take a more
269 heuristic approach. In Appendix I, we provide an iterative normalization algorithm that
270 constructs a matrix ψ satisfying the conditions in eq. (8) for a given value of *r*. The distance *r*
271 determines the strength of correlation between the two dimensions is. If the distance is small,
272 then the organism will expect to find itself in very similar conditions, which implies that the joint
273 distribution must have high correlation. In contrast, if the organism takes a very large step, then
274 it may find itself in any habitat, with probability proportional to that habitat's global availability.
275 The correlation strength as a function of distance *r* is extracted directly from the environmental
276 data, using an empirical autocorrelation function (ACF – see Appendix I).

277 Placing eqs (7) and (1) into eq. (6) gives

278 $$g_{x|z}(r) = \frac{\sum_{l=1}^{L}\sum_{m=1}^{L}\psi_{l,m}(r) f_{l,x} f_{m,z}}{\sum_{m=1}^{L}\psi_m f_{m,z}}. \quad (9)$$

279 Replacing into eq. (2) and rearranging the integral produces

280 $$f_{x|z} = \frac{\sum_{l=1}^{L}\sum_{m=1}^{L} f_{l,x} f_{m,z} \int_r \psi_{l,m}(r) h(r)\ dr}{C_E \sum_{m=1}^{L}\psi_m f_{m,z}}. \quad (10)$$

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

284 $$\Psi_{l,m} = \int_r \psi_{l,m}(r) h(r)\ dr. \quad (11)$$

285 Note that these quantities satisfy the unit-sum requirement (from the second part of eq. (8)),

286 $$\sum_{l=1}^{L}\sum_{m=1}^{L}\Psi_{l,m} = \int_r \left(\sum_{l=1}^{L}\sum_{m=1}^{L}\psi_{l,m}(r) \right) h(r)\ dr = 1, \quad (12)$$

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

289 $$f_{x|z} = \frac{\sum_{l=1}^{L}\sum_{m=1}^{L} f_{l,x} f_{m,z} \Psi_{l,m}}{\int_E \sum_{l=1}^{L}\sum_{m=1}^{L} f_{l,x} f_{m,z} \Psi_{l,m}\, dx}. \quad (13)$$

290 In Appendix II, we show that the denominator in this expression is the marginal distribution of
291 availability, yielding

292 $$f_{x|z} = \frac{\sum_{l=1}^{L}\sum_{m=1}^{L} f_{l,x} f_{m,z} \Psi_{l,m}}{f_z}. \quad (14)$$

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

294 $$f_{x,z} = \sum_{l=1}^{L}\sum_{m=1}^{L} f_{l,x} f_{m,z} \Psi_{l,m} . \quad (15)$$

295 Plotting the numerator of eq. (14) for different mobility constraints (Fig. 2) illustrates the
296 operation of the calculations of Appendix I. At low mobility (Fig. 2a) the correlation between
297 different types of habitat is strong, but increasing the mobility of the organism (as shown in Figs
298 2b and 2c, by using higher values of ω), moves the joint distribution closer to the independence
299 scenario $f_{x,z} = f_x f_z$.

300 The result in eq. (15) is already applicable to one, two or more spatial dimensions (via an
301 appropriate specification of $h(r)$, see examples in eqs (4) and (5)). In principle, eq. (15) is also
302 applicable to multiple environmental dimensions, but this would also require additional
303 methodological work to generalize the algorithm in Appendix I, so as to include any cross-
304 correlations between environmental variables in addition to their auto-correlations. However,
305 using the algorithm of Appendix I in its current form for many environmental variables is also
306 possible if they can plausibly be assumed to be independent of each other. The extensive
307 literature on collinear environmental variables can be used, either to test for non-independence
308 between environmental dimensions, or (e.g. via principal components analysis) to construct a
309 new set of independent environmental variables (Dormann et al. 2013). Given such a set of
310 orthogonal variables, habitat availability in *K*-dimensional *E*-spaces can be written as

311 $$f_{\mathbf{x}|\mathbf{z}} = f_{x_1|z_1} f_{x_2|z_2} \cdots f_{x_K|z_K} \quad (16)$$

312 ## 4. Illustration using direct sampling of availability from G-space

313 In the preceding sections, we dealt with the problem of restricted accessibility by extending the
314 mathematical definition of habitat availability. A more direct approach to quantifying availability

315 in a particular landscape is to sample around different locations in *G*-space (as we did in Fig. 1).
316 It is therefore useful to visualize the outputs of the sampling and analytical approaches on a
317 simple example for a particular simulated landscape, to help with the interpretation of our
318 method and to motivate a discussion of Monte Carlo error.
319 To generate a joint distribution of habitat availability via sampling, the following steps could
320 be adopted:
321 1. Systematically or randomly select a set $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ of points in *G*-space.
322 2. Randomly sample points (in our case, 50) from the vicinity (in *G*-space) of each $\mathbf{s}_i$
323 according to an accessibility kernel with mobility parameter ω. This will generate a set
324 of satellite points $\mathbf{U}_i = \{\mathbf{u}_{i,1}, \ldots, \mathbf{u}_{i,50}\}$ reflecting the spatial extent of conditional
325 availability (accounting for both mobility constraint and amount of spatial autocorrelation
326 in the environmental variables).
327 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
328 and increment their absolute frequency by one.
329 Using the same simulated landscape throughout this section (see example in Appendix I), we
330 specified two different mobility kernels across the rows of Fig. 3 corresponding to slow-moving
331 animals or short time intervals (Figs 3a, 3b) and fast-moving animals or long time intervals (Figs
332 3c, 3d). The analytical approach in *E*-space derived in section 3 gave the outputs of Figs 3a and
333 3c. We compared these with the corresponding plots (Figs 3b and 3d) obtained via the sampling
334 approach described above. The two approaches give broadly comparable descriptions of the two
335 mobility scenarios, but the model-based approach yielded a smoother description than the
336 sampling algorithm. These differences between the analytical and sampling plots are due to two
337 types of stochasticity. The first relates to Monte Carlo error due to the finite sample sizes taken
338 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 eq. (15) that can allow further applications to make algebraic shortcuts.

5. Applied example: Step selection functions for the analysis of telemetry data

Step selection functions are a method of fitting habitat models to animal telemetry data (Fortin et al. 2005, Thurfjell et al. 2014, Singer et al. 2018). 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 $f_u(\mathbf{s}_j \mid \mathbf{s}_{j-1})$ is described as (see Forester et al. 2009),

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

where $w(\mathbf{x}(\mathbf{s}_j))$ describes habitat preferences and $f_a(\mathbf{s}_j \mid \mathbf{s}_{j-1})$ expresses mobility (the "resource-independent movement kernel" described in Forester et al. 2009). The selection function $w(\mathbf{x})$ is modeled as a log-linear function of predictor variables. Here, as in Matthiopoulos et al. (2015),

361 we employ a curvilinear polynomial form comprising terms up to 2nd order, to allow for the
362 detection of optima in the animal's response to some environmental variables:

363 $$w(\mathbf{x}) = \exp\left(\sum_{k=1}^{K} \sum_{\eta=1}^{2} \gamma_{\eta,k} x_k^{\eta} \right). \quad (18)$$

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

366 $$l(\mathbf{s}_j \mid \mathbf{s}_{j-1}; \gamma) = \log w(\mathbf{x}(\mathbf{s}_j); \gamma) + \log f_a(\mathbf{s}_j \mid \mathbf{s}_{j-1}) - \log \int_{\mathbf{u} \in G} w(\mathbf{x}(\mathbf{u}); \gamma) f_a(\mathbf{u} \mid \mathbf{s}_{j-1}) d\mathbf{u}. \quad (19)$$

367 The log-likelihood of the entire data set of telemetry data is constructed by combining the
368 individual likelihoods of all the observed relocations in the data,

369 $$l = \sum_{j=1}^{J} l(\mathbf{s}_j \mid \mathbf{s}_{j-1}; \gamma). \quad (20)$$

370 Employing this log-likelihood within standard estimation approaches, specifically, conditional
371 logistic regression (Fortin et al. 2005), usually involves two simplifying steps (Forester et al.
372 2009). First, the mobility function $f_a(\mathbf{s}_j \mid \mathbf{s}_{j-1})$ is assumed known, and, second, the non-trivial
373 integral of eq. (19) is approximated by point-sampling methods. The first simplifying step allows
374 the term $\log f_a(\mathbf{s}_j \mid \mathbf{s}_{j-1})$ to be dropped from the log-likelihood, since it contains no parameters
375 that need to be estimated from the data. The second step deals with the integral by organizing the
376 telemetry data into strata, each comprising a single focal telemetry location $\mathbf{s}_j$ and a sample (of
377 size *V*) of control locations $\mathbf{s}_v$. Controls are selected randomly from the geographical vicinity of
378 the telemetry observation $\mathbf{s}_{j-1}$ immediately preceding $\mathbf{s}_j$ so as to represent the habitat options
379 that were available to the animal from that previous position.
380 These two simplifying steps bring the log-likelihood of eq. (19) within the remit of
381 conditional logistic regression, which, for the j^{th} point in a telemetry dataset, is written as

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

where $\mathbf{x}_j = \{x_1,\ldots,x_K\}_j$ 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 & Lumley 2019), and is therefore frequently used for applied analyses (see review in Thurfjell et al. 2014). The estimates of the parameters 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

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

An alternative approach to obtaining a step selection likelihood, without the need to sample control points, is to notice that the sum in eq.(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 probability density function of the availability of habitats around the preceding point (i.e. $f_{\mathbf{z}|\mathbf{s}_{j-1}}$) we could rewrite eq. (21) as

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

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 $f_{\mathbf{z}|\mathbf{s}_{j-1}} \cong f_{\mathbf{z}|\mathbf{x}_{j-1}}$.

402

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

406 $$l_{CLL}(\mathbf{x}_j \mid \mathbf{x}_{j-1}) = \log(w(\mathbf{x}_j)) - \log\left(\int_E w(\mathbf{z}) f_{\mathbf{z}|\mathbf{x}_{j-1}} \, d\mathbf{z} \right). \quad (24)$$

407 Using the results on conditional availability developed in earlier sections, we show in Appendix
408 III that the integral involved in this log-likelihood has a closed form solution. Hence, eq. (20)
409 can be obtained analytically, as

410 $$l_{CLL} = \sum_{j=1}^{J} \log(w(\mathbf{x}_j)) - \sum_{j=1}^{J} \log\left(\frac{1}{f_{\mathbf{x}_{j-1}}} \prod_{k=1}^{K} \sum_{l=1}^{L} \sum_{m=1}^{L} \Psi_{k,l,m} f_{l,x_k} \Theta(\gamma_{1,k}, \gamma_{2,k}, \mu_{m,k}, \sigma_k) \right), \quad (25)$$

411 where $\Theta(\gamma_{1,k}, \gamma_{2,k}, \mu_{m,k}, \sigma_k)$ is an algebraic function of parameters pertaining to habitat
412 preference and availability. This analytical expression can prove useful in studies with imperfect
413 or irregular environmental data sets. For example, a number of modern telemetry tags,
414 particularly in the marine environment, collect in-situ environmental data in addition to location
415 information (Beringer et al. 2004, Biuw et al. 2007, Hooker et al. 2008, Ericsson et al. 2015). For
416 environmental variables that are only measured at the location of the animal, our model could
417 provide a useful description of habitat availability for locations that were potentially accessible
418 but not visited by the animal. If some representative segments of space have been independently
419 surveyed to allow us to characterize the statistical properties of the distribution of these variables
420 (even if high-resolution covariate layers are not available exactly in the vicinity of the telemetry
421 data) then these can supplement the analysis. In addition, for temporally irregular data, our
422 model's mobility kernel can be used to give a varying degree of accessibility, depending on the
423 time interval between locational fixes (a problem also considered in Johnson et al. 2008,

424 Johnson, Hooten & Kuhn 2013). This flexibility can be extended to account for different modes
425 of mobility (e.g. as a result of diurnal activity patterns).

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

428 1. **Representativeness assumption.** The data from which the unconditional habitat
429 distribution is derived must be representative of the landscape on which the method is to
430 be applied. Therefore, we require the marginal distributions to be accurate, even if the
431 environmental layers are not known exactly. In a sufficiently large spatial arena, this
432 assumption can be satisfied without the need for high-resolution data. Any large point-
433 sample will suffice as long as it is collected systematically or randomly from the region
434 of interest or a region with similar properties.
435 2. **Stationary autocorrelation function assumption**. The shape of the autocorrelation
436 function must be the same between the regions used for training the approximation and
437 the geographical region of application. This assumption can be satisfied without the need
438 for spatially expansive data. A single high resolution transect that manages to capture the
439 form of autocorrelation will suffice.

440 As a first practical illustration of the above approach, we conducted a comparison between the
441 sampling and modelling approximations (i.e. eqs (21) and (25), respectively) on a real telemetry
442 data set (Fig. 4), collected from individual harbor seals (*Phoca vitulina*), off the north coast of
443 the Netherlands. We used a simple data set of two environmental covariates corresponding to
444 bathymetry (Fig. 4a) and the percentage of silt in the sediment (Fig. 4b). We selected time
445 intervals between the pairs of successive observations in the data set to be less than 24hrs and
446 sub-sampled from the data set (taking 1 out of every 20 consecutive pairs of locations) to ensure
447 that the successive pairs in the data set were serially independent. The value of the parameter of
448 the mobility kernel $\omega = 2.58$ (in units of grid cell lengths) was derived directly from the data, as

449 the standard deviation of the Rayleigh distribution (calculated as $\sqrt{2\,\mathrm{var}(|\Delta\mathbf{s}|)/(4-\pi)}$, where

450 $|\Delta\mathbf{s}|$ were the observed step lengths in the data). The sampling approximation used 200 control

451 points for each stratum (i.e. combined with each pair of successive locations). The controls for

452 the sampling approach were selected using Rayleigh step lengths with a uniformly random

453 direction on the circle. The modelling approximation used the same Rayleigh distribution and

454 covariate information originating either from a box enclosing the telemetry data (the yellow box

455 in Fig. 4c) or from a strip of the sea that was outside the telemetry set (the blue rectangle in Fig.

456 4c). This comparison allowed us to explore the sensitivity of parameter estimates and spatial

457 predictions to changes in habitat structuring (i.e. violations of the two assumptions of

458 representativeness and stationary autocorrelation). The two regions differed in their area, shape

459 and location. The elongated shape of the blue region precluded averaging over the strong

460 anisotropy in the environment. These differences potentially reduced the representativeness of

461 the blue region.

462 To visualize the differences in habitat composition between the two boxes, we plotted the

463 actual frequency of sea depths and sediment values (the black curves in Figs 4d and 4e

464 respectively) against the modeled availability of those two variables within the yellow box (solid

465 brown line for depth in Fig. 4d and solid blue line for sediment in Fig. 4e) and within the blue

466 rectangle (dotted brown line for depth in Fig. 4d and dotted blue line for sediment in Fig. 4e). In

467 addition, we explored differences in spatial autocorrelation between the yellow and blue

468 rectangles (Fig. 4f). We visualized the results of the analysis in geographic as well as parameter

469 space. The geographic visualization for each of the three analyses looked at the value of the step

470 selection function in each of the map's pixels (Figs 4g,h,i). These values can be interpreted as a

471 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 (Figs 4j,k,l).

The above comparison leads to the following conclusions. When the training data are obtained from the region of interest, the modelling approximation gives similar spatial results to geographic sampling (compare Fig. 4g with Fig. 4h) and the 95% confidence ellipses overlap (Figs 4j and 4k). Using training data outside the region of interest, so that the assumptions of representativeness and stationary autocorrelation are less faithfully preserved (see diagnostics in Figs 4d,e,f), may result in differences between the two approaches, (compare Fig. 4g with 4h and 4i). Yet, the parameter estimates remain within plausible ranges for this particular problem (compare Fig. 4j with 4l). So, 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 modelling 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 only uses 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).

6. 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 $F_x = a_0 + a_1 x$ for some coefficients a_0, 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 non-monotonic contributions to fitness (see Matthiopoulos et al. (2015) for more complex extensions).

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:

$$F_{LO} = \int_E F_x f_x \, dx \,. \tag{26}$$

In Appendix IV we show that this simplifies to

$$F_{LO} = a_0 + a_1 \sum_{l=1}^{L} \psi_l \mu_l \,. \tag{27}$$

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 $\overline{x}$ is the average value of the resource in the environment, under low spatial autocorrelation, we get the intuitive result, corresponding to perfect mixing,

$$F_{LO} = a_0 + a_1 \overline{x} \,. \tag{28}$$

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

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

In Appendix IV we show that this simplifies to

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

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

$$\Psi_{l,k} = \psi_l \psi_k \,. \tag{31}$$

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

$$\overline{F} - \overline{F}_{LO} = \int_z F(z) f_z dz - F_{LO} \,, \tag{32}$$

which rearranges to

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

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

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

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 Appendix V, we show that this expression can be simplified to

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

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 Appendix I, but we can simplify our

556 investigation by comparing the two extremes of very low and very high spatial autocorrelation.
557 The case of very low spatial autocorrelation is represented by eq. (28). The case of very high
558 spatial autocorrelation can be emulated by setting

559 $$\Psi_{l,k} = \begin{cases} \psi_k & \text{if} \quad l = k \\ 0 & \text{otherwise} \end{cases} . \tag{36}$$

560 Within the expression for joint habitat availability (eq. (15)), this works by accumulating a high
561 probability density close to the line of slope 1 (creating joint PDFs similar to those in Figs. 2a or
562 3a), hence enforcing the probability of encountering similar values of z from an animal's current
563 position. Via this simplification, the fitness equation becomes

564 $$\bar{F} = a_0 + a_1 \frac{\sum_{k=1}^{L} \psi_k \mu_k \exp(b_1 \mu_k)}{\sum_{k=1}^{L} \psi_k \exp(b_1 \mu_k)} . \tag{37}$$

565 Subject to the assumption of high spatial autocorrelation, we proceed to explore the behavior of
566 this function by varying the overall resource richness (related to the mean value of the
567 distribution of the available resource) and heterogeneity (the variability of the distribution of the
568 available resource) of the landscape. To do this in a tractable way, we envisage an environmental
569 space that is constructed of L equally spaced and equally weighted Gaussian components (Fig.
570 4a). The mean value ($\bar{z}$) of the mixture determines overall resource richness and the number (L)
571 of individual components, equally split on either side of the mean, represents heterogeneity. We
572 take the spacing between adjacent Gaussian means μ_k, μ_{k+1} to be equal to σ, the standard
573 deviation of each of the Gaussian components. This assumption tends to give approximately
574 uniform distributions of the resource in E-space (thick grey curve in Fig. 5a). Note however that
575 the distribution of the resource in G-space will be heterogeneous. This simplified representation
576 of the environment yields

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

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

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

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

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

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 towards 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 of b_1. We therefore consider three scenarios that give rise to important boundaries in the richness/heterogeneity plane (see 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

$$\lim_{b_1\to\infty}\sum_{k=1}^{L} k\theta_k = L. \tag{41}$$

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

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

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,

$$\sum_{k=1}^{L} k\theta_k = \frac{L+1}{2} \tag{43}$$

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

$$\bar{x} = -\frac{a_0}{a_1} \tag{44}$$

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

$$\lim_{b_1 \to -\infty} \sum_{k=1}^{L} k\theta_k = 1 \tag{45}$$

The mathematical condition for positive fitness is

$$\bar{x} > -\frac{a_0}{a_1} + \sigma\frac{L-1}{2} \tag{46}$$

For graphical convenience, we define resource richness in relation to fitness parameters. Further, 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.

$$\text{Richness} = \bar{x} + \frac{a_0}{a_1} \qquad \text{Heterogeneity} = \sigma\frac{L-1}{2} \tag{47}$$

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 eqs (42), (44) and (46) with the aid of these new definitions gives us the combined results in Fig. 5b that 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)).

7. 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 & 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 Fig. 5b illustrate how the relative

640 scales of mobility and environmental heterogeneity can fundamentally alter the fitness that a
641 landscape can afford a population. Regions 2 and 3 in Fig. 5 have a novel biological
642 interpretation, in which habitat selectivity changes the sign of population growth relative to an
643 assumption of no selectivity. Region 3, in particular, is the direct result of animals in the
644 population being able to aggregate at hotspots of resource distribution, and hence experience
645 higher-than-average fitness, compared to a non-spatial model assuming perfect mixing.
646 Many of these insights would be achievable on a particular landscape by means of intensive
647 sampling of space (as we argued in Section 4); however, our framework offers a flexible
648 abstraction of species-habitat interaction based solely on the statistical properties of the system.
649 This allows us to work in environmental spaces and produce generalizable results, applicable to
650 different landscapes with similar landscape compositions. Similar models can be derived through
651 moment equation modeling incorporating at the same time the spatial autocorrelation in
652 environmental variation, the movement processes through kernels – as done here - as well as
653 deriving the environment-organism covariance from the interaction of dispersal, demography
654 and environmental structure (Murrell & Law 2000, Bolker 2003, North et al. 2011). However,
655 these methods require more complex analytical formulas, place their emphasis on population
656 dynamics and do not operate explicitly in *E*-spaces.
657 A recent review of species range models (Singer et al. 2016) discusses how mechanistic
658 approaches can be used to enhance the predictive ability of correlative models of species'
659 distribution. Hence, the present work can be used to increase the mechanistic content of
660 correlative models, but may also be used for expedient calculation in fully mechanistic
661 approaches. The mechanistic content of our approach can be increased to account for features of
662 movement. For example, the variance of our movement kernel can be assumed to depend on the
663 properties of the local habitat, to account for reductions in mobility due to difficult substrates.
664 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 & Yohe, 2003).

Supplements

Supplementary information is provided for the Appendix derivations as well as an explanation of the algorithms derived in the paper. Archival files for the R-code and data used in the paper can be downloaded from https://zenodo.org/record/3479825#.XZ-MDSV7knc.

Acknowledgments

The impetus for this work came from discussions with Devin Johnson, Otso Ovaskainen and Paul Blackwell. 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.

References

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. & Matthiopoulos, J. (2008) Estimating space use and environmental preference from telemetry data. *Ecography*. 31, 140-160.

Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecol. Evol.* 3: 177-187.

689 Aarts, G., Fieberg, J., Brasseur, S. & Matthiopoulos, J. (2013) Quantifying the effect of habitat
690 availability on species distributions. *Journal of Animal Ecology* 82: 1135-1145.
691 Avgar, T., Potts, J.R., Lewis, M.A. & Boyce, M.S. (2016) Integrated Step Selection Analysis:
692 Bridging the Gap between Resource Selection and Animal Movement. *Methods in Ecol. Evol.*
693 7: 619–630.
694 Barraquand, F. and Murrell, D. J. (2013) Scaling up predator–prey dynamics using spatial
695 moment equations. *Methods in Ecol. Evol*. 4: 276–289.
696 Beringer, J., Millspaugh, J.J., Sartwell, J. and Woeck, R. (2004) Real-time video recording of
697 food selection by captive white-tailed deer. *Wildlife Society Bulletin*, pp.648-654.
698 Beyer, H.L., Haydon, D., Morales, J., Frair, J.L., Hebblewhite, M., Mitchell, M. &
699 Matthiopoulos, J. (2010) Habitat preference: Understanding use versus availability designs.
700 *Philos. Trans. of the Royal Soc. B.* 365: 2245-2254.
701 Biuw, M., Boehme, L., Guinet, C., Hindell, M., Costa, D., Charrassin, J.-B., Roquet, F., Bailleul,
702 F., Meredith, M., Thorpe, S., Tremblay, Y., McDonald, B., Park, Y.-H., Rintoul, S.R., Bindoff,
703 N., Goebel, M., Crocker, D., Lovell, P., Nicholson, J., Monks, F., Fedak, M.A. (2007) Variations
704 in behavior and condition of a Southern Ocean top predator in relation to in situ oceanographic
705 conditions. P.N.A.S. 104: 13705-13710.
706 Bolker, B. M. (2003). Combining endogenous and exogenous spatial variability in analytical
707 population models. *Theoretical Population Biology*, 64(3): 255-270.
708 Blackwell, P. G. (1997). Random diffusion models for animal movement. *Ecol. Mod.* 100, 87–
709 102.
710 Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality: The once and future niche. *PNAS*.
711 106: 19651-19658.

712 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G.,
713 Gruber, B., Lafourcade, B., Leitão, P.J. & Münkemüller, T. (2013). Collinearity: a review of
714 methods to deal with it and a simulation study evaluating their performance. *Ecogr.*, 36: 27-46.
715 Elith, J. & Leathwick, J.R. (2009) Species distribution models: Ecological explanation and
716 prediction across space and time. *An. Rev. Ecol., Evol. and Systematics* 40: 677-697.
717 Ericsson, G., Dettki, H., Neumann, W., Arnemo, J.M. & Singh, N.J. (2015) Offset between GPS
718 collar-recorded temperature in moose and ambient weather station data. *European Journal of*
719 *Wildlife Research*, 61(6), pp.919-922.
720 Fagan, W.F., Gurarie, E., Bewick, S., Howard, A., Cantrell, R.S. & Cosner, C (2017) Perceptual
721 Ranges, Information Gathering, and Foraging Success in Dynamic Landscapes. *Am. Nat.* 189:
722 474–89.
723 Feller, W. (1966) *An Introduction to Probability Theory and Its Applications.* Volume 2. John
724 Wiley & Sons. 696pp.
725 Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005) Wolves
726 influence elk movements: behavior shapes a trophic cascade in yellowstone national park.
727 *Ecology*, 86, 1320–1330.
728 Fraley, C., Raftery A.E. & Wehrens R. (2005). Incremental model-based clustering for large
729 datasets with small clusters. *Journal of Computational and Graphical Statistics* 14:1-18.
730 Fraley, C., Raftery, A.E, Murphy, T.B. & Scrucca, L. (2012). mclust Version 4 for R: Normal
731 Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.
732 Technical Report No. 597, Department of Statistics, University of Washington.
733 Gurarie, E. and Ovaskainen, O. (2011) Characteristic Spatial and Temporal Scales Unify Models
734 of Animal Movement. *The American Naturalist*. 178, 113-123.
735 Hall, L.S., Krausman, P.R. & Morrison, M.L.(1997) The habitat concept and a plea for standard
736 terminology. *Wildl. Soc. Bull*. 25:173-182.

737 Hirzel, A.H. & Le Lay, G. (2008) Habitat suitability modeling and niche theory. *Journal of*
738 *Applied Ecology.* 45: 1372-1381.
739 Hooker, S.K., Heaslip, S.G., Matthiopoulos, J., Cox, O. & Boyd, I. (2008) Data sampling options
740 for animal-borne video cameras: considerations based on deployments with Antarctic fur seals.
741 *Marine Technology Society Journal.* 42: 65-75.
742 Hooten, M.B., Johnson, D.S., McClintock, B.T. & Morales, J.M. (2017) *Animal Movement:*
743 *Statistical models for telemetry data*. CRC Press. Boca Raton. 306pp.
744 Horne, J.S., Garton, E.O. & Rachlow J.L. (2008) A Synoptic Model of Animal Space Use:
745 Simultaneous Estimation of Home Range, Habitat Selection, and Inter/Intra-Specific
746 Relationships. *Ecological Modelling* 214: 338–48.
747 Hughes, B. (1995) *Random walks and random environments.* Vol. I, Clarendon Press, 631p.
748 Hutchinson, G.E. (1957) Concluding remarks. *Cold spring Harb. Symp. Quant. Biol.* 22:415-
749 427.
750 Joe, H. (2015) *Dependence modeling with copulas.* Chapman and Hall/CRC. Boca Raton Fl.
751 USA. 480p.
752 Johnson, D.H. (1980) The comparison of usage and availability measurements for evaluating
753 resource preference. *Ecology* 61:65-71.
754 Johnson, D.S., Hooten, M.B. & Kuhn, C.E. (2013) Estimating animal resource selection from
755 telemetry data using point process models. Journal of Animal Ecology, 82, 1155–1164.
756 Johnson, D.S., Thomas, D.L., Ver Hoef, J.M. and Christ, A. (2008) A general framework for the
757 analysis of animal resource selection from telemetry data. *Biometrics,* 64: 968-976.
758 Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*. 73: 1943-1967.
759 MacArthur, R.H. (1968) *The theory of the niche. In Population Biology and Evolution,*
760 Lewontin, R.C. (Ed.) Syracuse University Press, New York.

761 Manly, B., McDonald, L., Thomas, D., McDonald, T. & Erickson, W.P. (2004) *Resource*
762 *selection by Animals - Statistical design and analysis for field studies.* Chapman & Hall,
763 London, 177 pp.
764 Matthiopoulos, J. (2003) The use of space by animals as a function of accessibility and
765 preference. Ecological Modelling. 159:239–268.
766 Matthiopoulos, J., Hebblewhite, M., Aarts, G. & Fieberg, J. (2011) Generalized functional
767 responses for species distributions. *Ecology.* 92: 583-589.
768 Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M. and Haydon, D. T. (2015)
769 Establishing the link between habitat selection and animal population dynamics. *Ecol. Monogr.*
770 85: 413–436.
771 Matthiopoulos, J., Field, C. & MacLeod, R. (2019) Predicting population change from models
772 based on habitat availability and utilization. *Proceedings of the Royal Society B*, 20182911.
773 McInerny, G.J. & Etienne, R.S. (2013) 'Niche' or 'distribution' modeling? A response to
774 Warren. *T.R.E.E.* 28: 191-192.
775 McLoughlin, P.D., Morris, D.W., Fortin, D., Vander Wal, E. & Constantini, A.L. (2010)
776 Considering ecological dynamics in resource selection functions. *J. Animal Ecol*. 79: 4-12.
777 Murrell, D. J., & Law, R. (2000). Beetles in fragmented woodlands: a formal framework for
778 dynamics of movement in ecological landscapes. *J. Animal Ecol.* 69(3), 471-483.
779 North, A., Cornell, S., & Ovaskainen, O. (2011). Evolutionary responses of dispersal distance to
780 landscape structure and habitat loss. *Evolution*, 65(6): 1739-1751.
781 Northrup, J.M., Hooten, M.B., Anderson, C.R. & Wittemyer, G. (2013) Practical guidance on
782 characterizing availability in resource selection functions under a use–availability design.
783 *Ecology*, 94: 1456-1463.
784 Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts
785 across natural systems. *Nature*, 421: 37-42.

786 Paton, R. S. and Matthiopoulos, J. (2016) Defining the scale of habitat availability for models of
787 habitat selection. *Ecology*, 97: 1113–1122.
788 Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M.
789 & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton University
790 Press. Princeton. 314p.
791 R Core Team (2016). R: A language and environment for statistical computing. R
792 Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
793 Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006)
794 Are niche-based species distribution models transferable in space? Journal of Biogeography.
795 33: 1689-1703.
796 Sawo, F., Brunn, D. and Hanebeck, U.D. (2006) Parameterized Joint Densities with Gaussian
797 and Gaussian Mixture Marginals. *Proceedings of the 9th International Conference on*
798 *Information Fusion* (Fusion 2006), Florence, Italy.
799 Schneider, D.S. (2001) The rise of the concept of scale in ecology: The concept of scale is
800 evolving from verbal expression to quantitative expression. *BioScience*. 51: 545-553.
801 Signer, J., Fieberg, J. and Avgar, T., 2019. Animal Movement Tools (amt): R-Package for
802 Managing Tracking Data and Conducting Habitat Selection Analyses. Ecology and Evolution,
803 9: 880-890.
804 Sinclair, S. J., White, M.D. & Newell, G.R. (2010) How useful are species distribution models
805 for managing biodiversity under future climates? *Ecology and Society* 15: 8.
806 Singer, A., Johst, K., Banitz, T., Fowler, M., Groeneveld, J., Gutiérrez, A., Hartig, F., Krug, R.,
807 Liess, M., Matlack, G., Meyer, K., Pe'er, G., Radchuk, V., Voinopol-Sassu, A. & Travis, J.
808 (2016). Community dynamics under environmental change: How can next generation
809 mechanistic models improve projections of species distributions? *Ecol. Mod.* 326: 63-74.

Sklar, L. (2015) Philosophy of Statistical Mechanics, in The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.)

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: Concepts, methods, and assumptions. P.N.A.S. 106: 19644-19650.

Therneau, T.M. & Lumley, T. (2019) Survival analysis. https://github.com/therneau/survival

Thurfjell, H., Ciuti, S. & Boyce, M.S. (2014) Application of step-selection functions in ecology and conservation. *Movement Ecology*, DOI: 10.1186/2051-3933-2-4.

Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecol. Evol*. 3: 260-267.

Wiens, J. A. (1989) Spatial scaling in ecology. *Functional Ecology*. 3: 385–397.

Zurell, D., Jeltch, F., Dormann, C. & Schröder, B. (2009) Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*. 32: 733-744.

Figure legends

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, 7 values in *G*-space (the dots 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 plate 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 dots in f). The resulting plot of frequencies for these localized measurements is shown as the dark curve in (e).

Figure 2: The joint distribution of habitat availability ($f_{\mathbf{x},\mathbf{z}}$), for a given marginal habitat availability ($f_{\mathbf{z}}$, $f_{\mathbf{x}}$) 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.

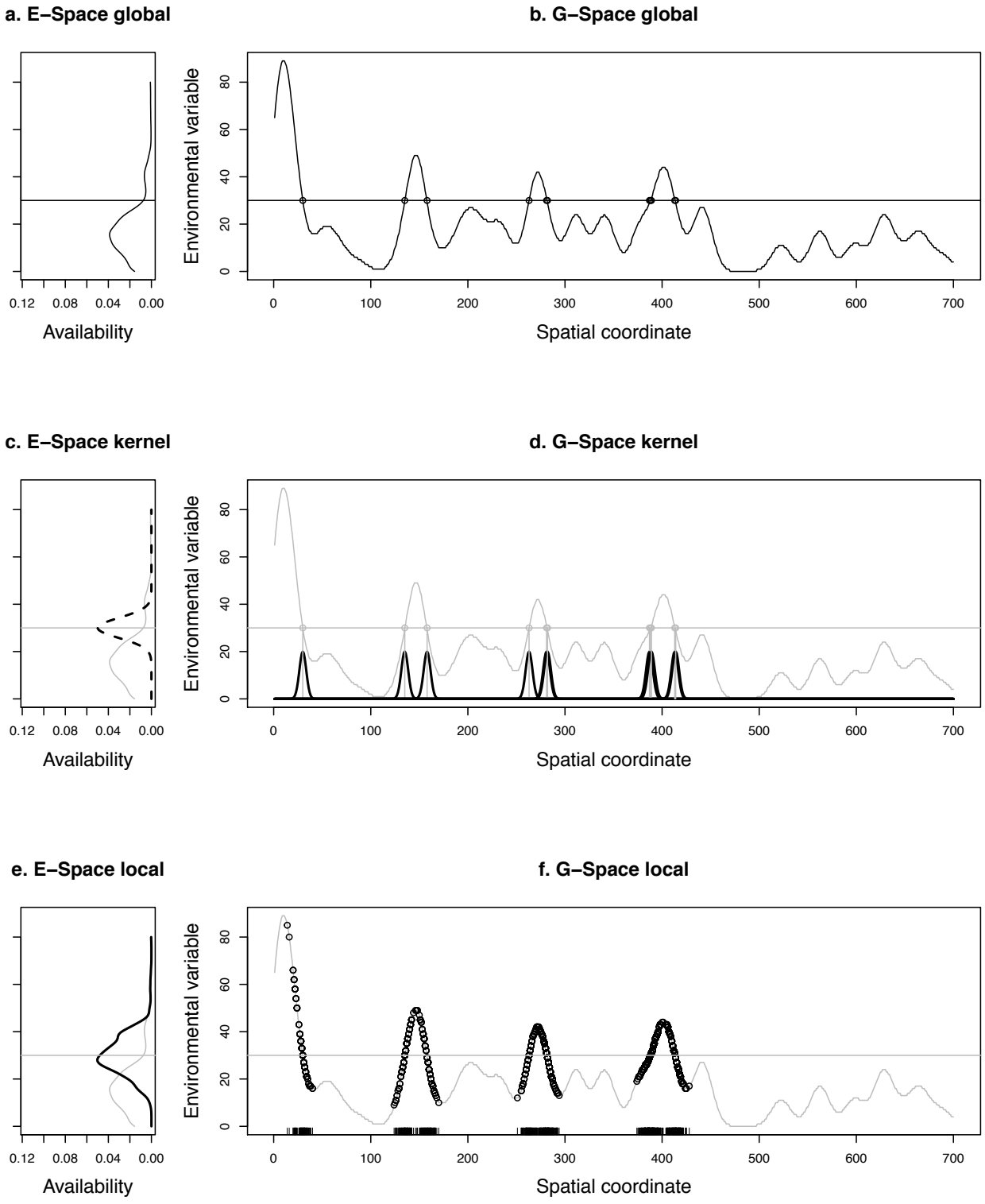
Figure 3: Comparison between the analytical and numerical forms of the joint distribution (left and right columns respectively) under scenarios of low and high mobility (top and bottom rows, respectively). The axes represent the variables z (focal habitat) and x (target habitat) for a one-dimensional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

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 substrate to movement, observed via satellite telemetry data (c). Step selection by *E*-space approximation used two different rectangular areas for learning about the environment (shown as yellow and blue in plate 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 f) brown colour is used for depth and blue for sediment. Solid lines correspond to the yellow box in part c and dotted lines correspond to the extrapolation in the blue rectangle in part 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 cross-hairs 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.

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

871 successive components). This arrangement allows us to reduce the description of *E*-space to the
872 two traits of resource richness (the position of the mixture along the resource axis) and
873 heterogeneity (the dispersion of the mixture, driven here by the number of participating Gaussian
874 components. b. Summary of findings in the graphical plane of resource richness and
875 heterogeneity in the case of a highly autocorrelated resource distribution in G-space. Four
876 regions arise indicating population viability depending on the habitat selectivity displayed by the
877 individuals making up the population.
878
879
880

881 # Figures



882

883 Figure 1

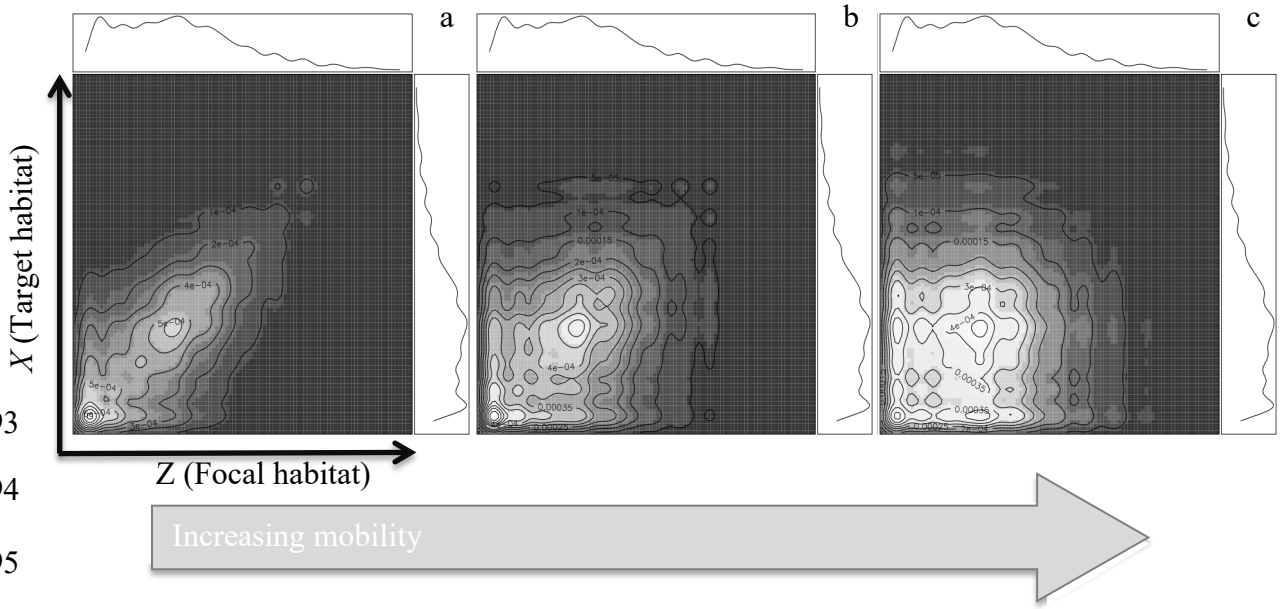
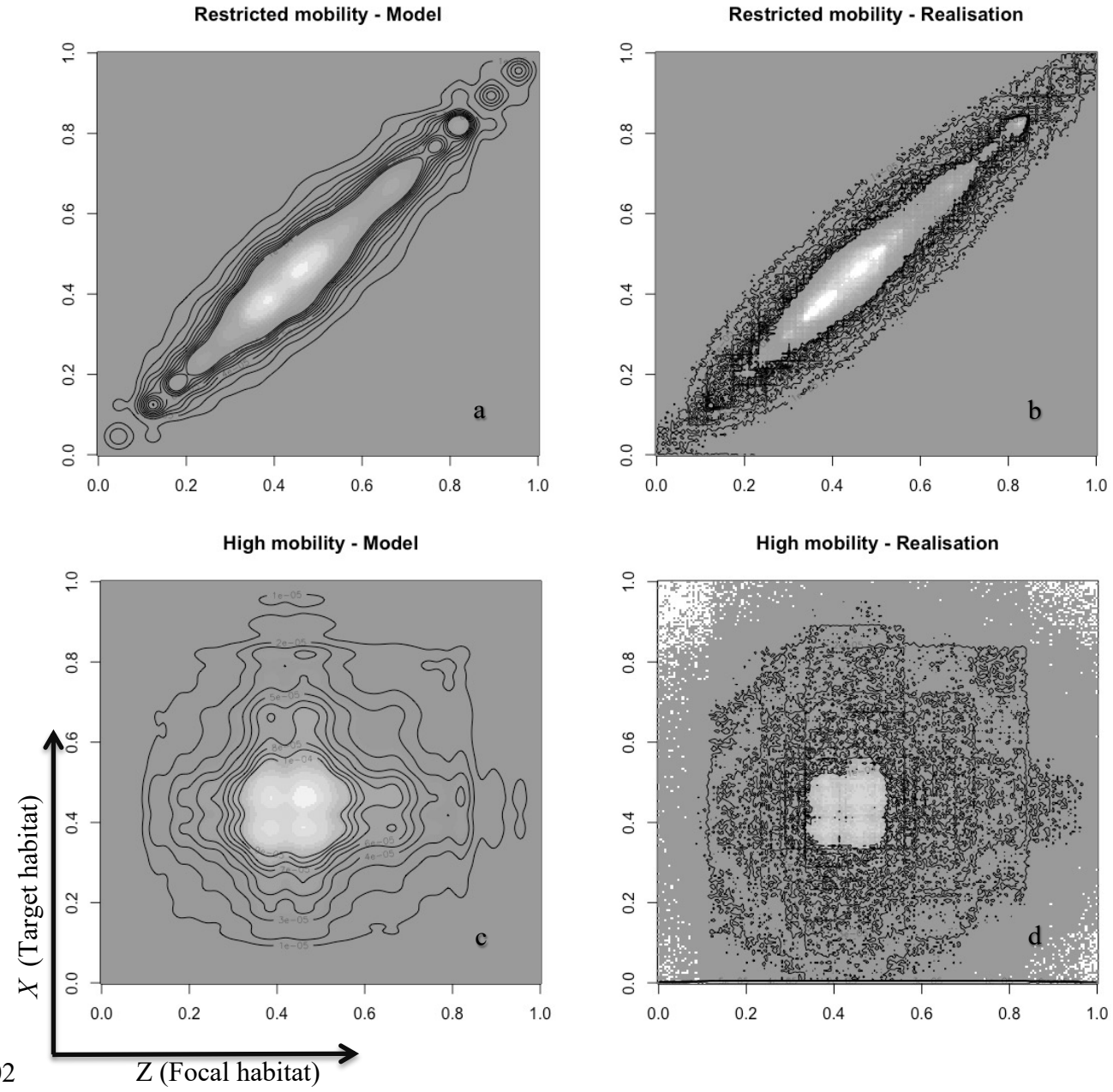


Figure 2



902

903

904

905 Figure 3

906

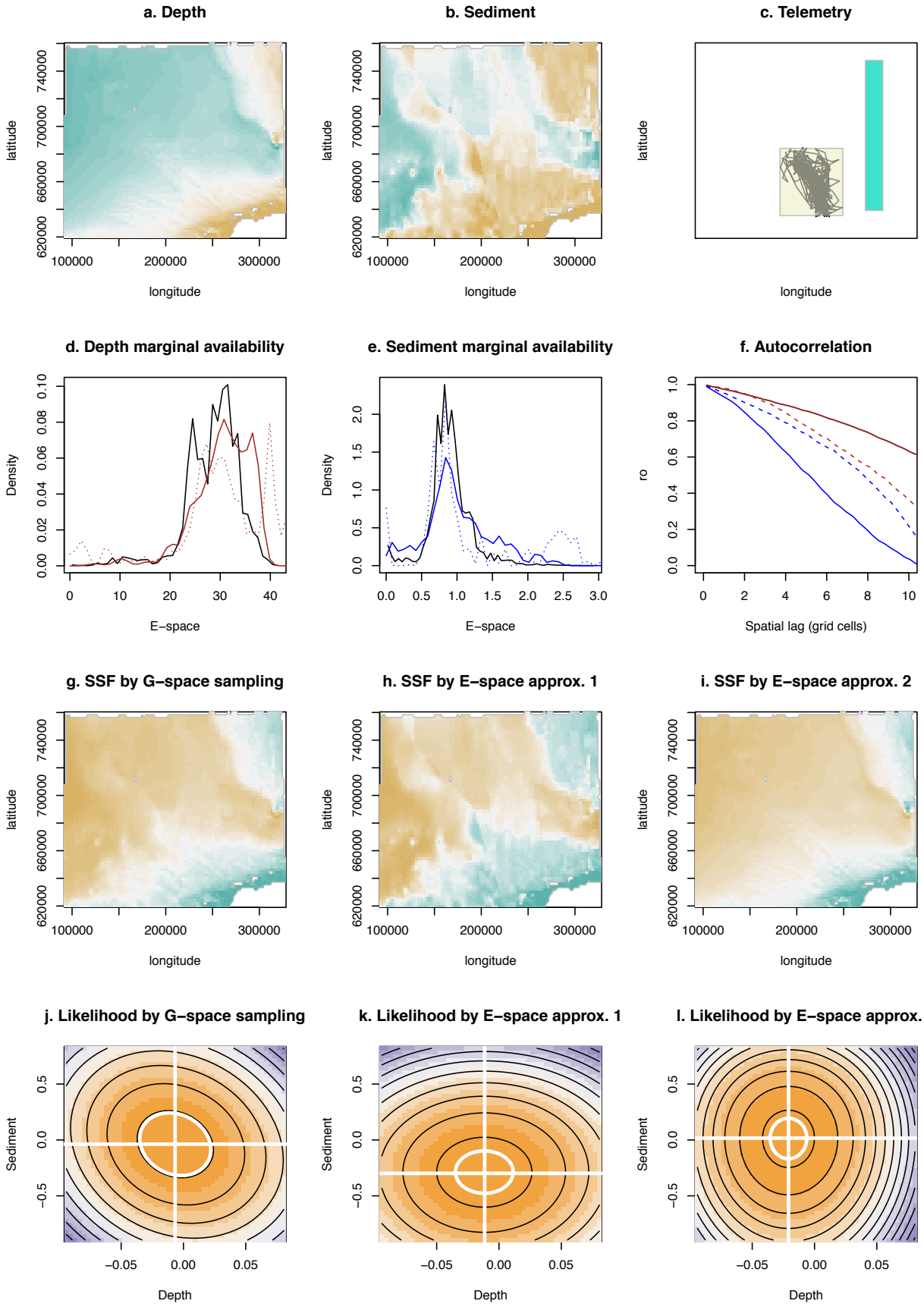

a. Depth
b. Sediment
c. Telemetry
latitude
longitude
d. Depth marginal availability
e. Sediment marginal availability
f. Autocorrelation
Density
E−space
ro
Spatial lag (grid cells)
g. SSF by G−space sampling
h. SSF by E−space approx. 1
i. SSF by E−space approx. 2
j. Likelihood by G−space sampling
k. Likelihood by E−space approx. 1
l. Likelihood by E−space approx. 2
Sediment
Depth

Figure 4

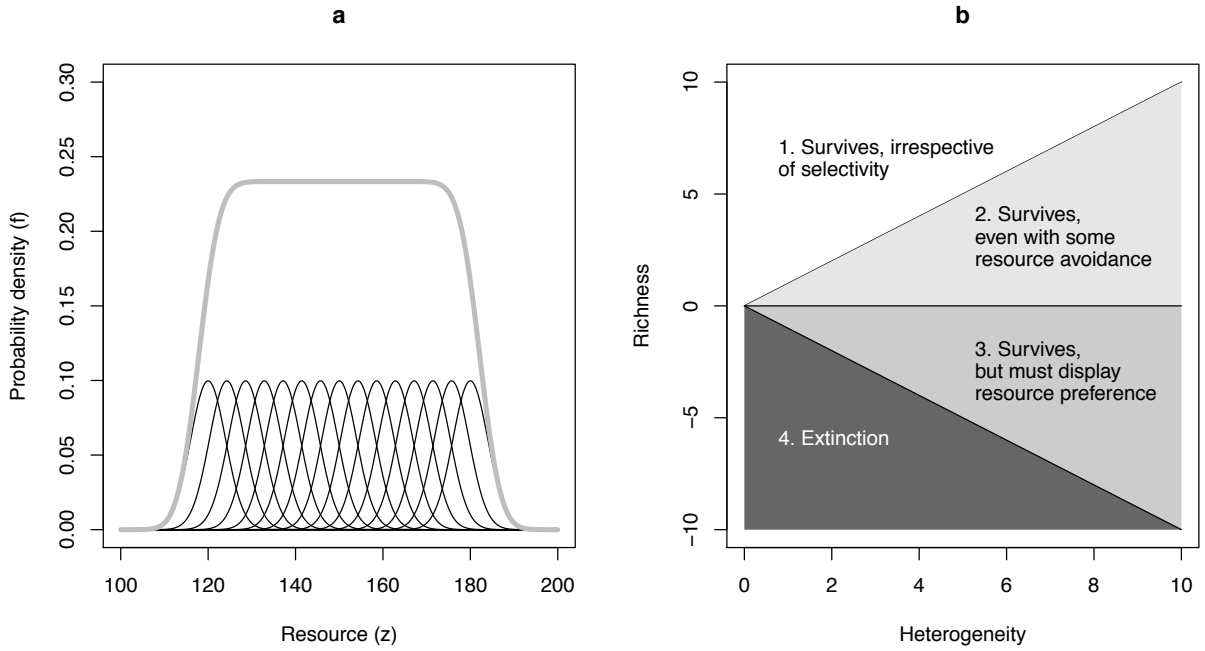

a
b
Probability density (f)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
100
120
140
160
180
200
Resource (z)
Richness
10
5
0
−5
−10
0
2
4
6
8
10
Heterogeneity
1. Survives, irrespective of selectivity
2. Survives, even with some resource avoidance
3. Survives, but must display resource preference
4. Extinction

Figure 5