Mathematical Biology manuscript No.
(will be inserted by the editor)

Robustness of movement models: can models bridge the gap between temporal scales of data sets and behavioural processes?

Ulrike E. Schlägel · **Mark A. Lewis**

Received: date / Accepted: date

Abstract Discrete-time random walks and their extensions are common tools for analyzing animal movement data. In these analyses, resolution of temporal discretization is a critical feature. Ideally, a model both mirrors the relevant temporal scale of the biological process of interest and matches the data sampling rate. Challenges arise when resolution of data is too coarse due to technological constraints, or when we wish to extrapolate results or compare results obtained from data with different resolutions. Drawing loosely on the concept of robustness in statistics, we propose a rigorous mathematical framework for studying movement models' robustness against changes in temporal resolution. In this framework, we define varying levels of robustness as formal model properties, focusing on random walk models with spatially-explicit component. With the new framework, we can investigate whether models can validly be applied to data across varying temporal resolutions and how we can account for these different resolutions in statistical inference results. We apply the new framework to movement-based resource selection models, demonstrating both analytical and numerical calculations, as well as a Monte Carlo simulation approach. While exact robustness is rare, the concept of approximate robustness provides a promising new direction for analyzing movement models.

Keywords animal movement · sampling rate · resource selection · GPS data · parameter estimation · Markov model

Mathematics Subject Classification (2000) 92B05 · 60J20 · 62M05 · 62-07

U.E. Schlägel
Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada
E-mail: ulrike.schlaegel@gmail.com
Present address: Institute of Biochemistry and Biology, Plant Ecology and Conservation Biology, University of Potsdam, Am Mühlenberg 3, 14476 Potsdam, Germany

M.A. Lewis
Department of Mathematical and Statistical Sciences & Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

1 Introduction

Major advances in tracking technology during the last decades have made large datasets of animal movement available to ecologists, and analyses of data have become widespread in ecology. These analyses have shed light on mechanisms that underly fundamental processes such as migration (Robinson et al 2009; Costa et al 2012), navigation (Tsoar et al 2011; Benhamou et al 2011), or home range behaviour and territoriality (Borger et al 2008; Potts and Lewis 2014; Giuggioli and Kenkre 2014). They have helped to identify conservation goals by revealing habitat preferences and critical environmental features for populations (Sawyer et al 2009; Colchero et al 2010; Ito et al 2013; Masden et al 2012), as well as the role of important mutualistic interactions between mobile animals and immobile plants (Côrtes and Uriarte 2013; Mueller et al 2014). These are only few of the many facets of movement ecology.

Mathematical and statistical models provide a framework for studying movement (Schick et al 2008; Smouse et al 2010; Langrock et al 2013). When linking models to data, we can estimate model parameters and identify best-fitting models, thus inferring unknown quantities or mechanisms in movement behaviour. Although movement itself is a continuous process, many individual-based movement models treat time as a discrete variable, viewing movement as a series of locations in space, or equivalently as a series of steps (Turchin 1998; McClintock et al 2014). This may largely be ascribed to data being available in this format. Discrete-time models may thus be an intuitive first choice to describe a sampled movement path. However, there may be more reasons to use discrete-time models. The continuous movement path of an animal may consist of various behaviours at different scales (Johnson et al 2002; Benhamou 2013). Using a discrete-time model at the scale of interest allows us to focus on the behavioural mechanisms at that scale, while, for example, combining other unknown processes as stochastic effects. Also, the choice of time formulation in a movement model can have side effects that impact inference results. For example, McClintock et al (2014) demonstrated that using a continuous-time Ornstein-Uhlenbeck process in a hierarchical model for identifying behavioural states led to difficulties discriminating between states, due to an inherent correlation between the variables step length and bearing in the Ornstein-Uhlenbeck process.

When linking discrete-time models to data, the temporal resolution of the discretization is a critical feature that must be chosen with care. Different time scales may come into play and need to be consolidated. On the one hand, a time scale is given by the biological process of interest. For example, we may be interested in inferring behavioural mechanisms of a movement process and thus need to consider the time scale at which these mechanisms are relevant. The discretization of a model should represent this scale appropriately. On the other hand, a different time scale may be given by the data collection rate. In practice, the sampling rate of data is subject to technological constraints. One of the major limitations of electronic tagging devices such as Argos or GPS tags is battery life, imposing a tradeoff between measurement rate and total deployment time (Ryan et al 2004; Breed et al 2011). Also, to avoid a large noise to signal ratio, the time interval should be chosen so that measurement error relative to distance travelled during a time interval is small (Ryan et al 2004). For slow moving animals and depending on the accuracy of the tagging device, a minimum time interval of an hour may be necessary (Jerde and Visscher 2005). Therefore, the resolution of the data may not always match the time scale of the behavioural process of interest. In this case, it becomes a challenge for a model to overcome the conflict.

A related problem is that sampling rate can affect data analysis results (Codling and Hill 2005; Rowcliffe et al 2012; Postlethwaite and Dennis 2013). A common measure calculated from raw movement data is the total distance travelled, which can provide useful information about an animal's energetic expenditures. It is well documented that this quantity is highly influenced by the sampling rate of the data (Ryan et al 2004; Mills et al 2006; Tanferna et al 2012; Rowcliffe et al 2012). A range of studies demonstrated that other fundamental movement characteristics vary with data sampling frequency as well, for example path sinuosity and apparent speed (Codling and Hill 2005), movement rate and turning angle (Postlethwaite and Dennis 2013), and estimates of territory size (Mills et al 2006). One of the main problems underlying these effects is information loss when subsampling a movement path. This also impairs our capacity to correctly estimate behavioural states through hierarchical modelling approaches that have become widespread in movement analyses (Breed et al 2011; Rowcliffe et al 2012). These findings demonstrate that great care is needed when extrapolating movement analysis results beyond the temporal scale of a study. Comparisons of results may not be appropriate if the temporal resolution of the data varies too much, but it is unclear what constitutes 'too much'.

Despite the evidence of the extent of the problem, little is known about how to solve it. Previous approaches have been mainly empirical, using very fine scale data or synthetic data from simulations, which are subsampled at various resolutions. Movement characteristics calculated at these varying sampling rates are then compared to the values based on the full data, which represent the 'true' values. Some studies have fitted functions to the relationships of movement characteristics and sampling rate (Pépin et al 2004; Codling and Hill 2005; Mills et al 2006). These empirically obtained functions may be used to correct movement characteristics for sampling rate. While correction factors derived from movement data remain situation-specific and cannot easily be applied across species (Ryan et al 2004; Rowcliffe et al 2012), we can obtain more general results by analyzing the effects of sampling rate at the level of the model (Codling and Hill 2005; Rosser et al 2013). Often, important characteristics of movement can be well captured by models, and therefore analyzing the properties of models can provide more general insights. However, only few such studies exist. An approach to circumvent the problem of scale-dependent statistical inference has been taken by Fleming et al (2014), who use the semivariance function of a stochastic movement process to identify multiple movement modes acting at different temporal scales. The method takes into account all possible time lags between observations. However, there are limitations as to the movement processes that can be included in this analysis (Fleming et al 2014).

Here, we present a rigorous framework for studying how movement models react to changes in sampling rate, and we use this framework to analyze a class of models based on random walks. With our analysis, we seek to understand whether, and how, models can help to compensate mismatching temporal scales between different data sets or between data and behavioural process of interest. The framework is based on the movement model robustness presented in Schlägel and Lewis (2016), where we analyzed classic random walks. Here we extend this to spatially-explicit random walks, as for example used in resource-selection studies (Forester et al 2009; Potts et al 2014). We investigate whether there are models that can validly be applied to data with different temporal resolutions and how we can account for the differences in resolutions in our interpretation of statistical inference results. In particular, we are interested in how model parameters, and their estimates, change as we decrease the temporal resolution. While estimates may change due to a shift in behavioural scale, which we always need to be aware of, we are interested in the changes that arise from the method, that is the model. Our framework is related to the statistical concept of robustness, which aims at safeguarding statistical procedures against violations of model assumptions (Hampel 1986; Huber and Ronchetti 2009). Often, such violations refer to deviations from assumed probability distributions (e.g. Normal errors), which may result in outliers, misspecified relationships between response and explanatory variables in regression analyses, or violations of the common independence assumption. In this paper, we define robustness of movement models against changes in temporal discretization. In our framework, we treat robustness as a formal property of a model, namely the movement model. If a model has this property, it can be applied to data with varying resolutions. Additionally, while model parameters do not stay the same, they change systematically and can be translated between resolutions.

As a cautionary note, we emphasize that the purpose of our paper is to highlight the sensitivity of movement data analyses based on discrete-time models to temporal resolution and to explore potential remedies. There will always be limitations as to the mismatch in resolution between process and data that a model can handle. As data becomes coarser behavioural detail is lost, and a model that is suitable at a fine scale, e.g. the scale of area-restricted search, is most likely unsuitable at a larger scale, e.g. the scale of patch selection (Benhamou 2013). Our analysis is directed towards a better and more precise understanding of the impact of temporal discretization on movement analyses, in particular when it is still reasonable to assume that the data's resolution is still within the scale of interest (e.g. 15-minute data versus 4-hour data for a large mammal). In our study of simple random walks, we found that movement model robustness is a very strong condition (Schlägel and Lewis 2016). Therefore, we here extend our framework to include approximate robustness, which slightly relaxes the assumptions of exact robustness.

Our paper is outlined as follows. In section 2, we define what we mean by a movement model to be robust against changes in temporal resolution. We provide three different definitions, varying in their strength of conditions. In section 3, we present different approaches how the definitions can be used to analyze robustness of movement models. Depending on models' complexity and preexisting information, we can use formal analytical methods, numerical calculations, as well as Monte Carlo and simulation approaches. We use these approaches to examine robustness of spatially-explicit random walks and resource-selection models, and we summarize our findings in section 4. In section 5, we discuss the relevance of our robustness framework for statistical inference and also draw specific conclusions for spatially-explicit resource-selection models.

2 Robustness of Markovian movement models

We consider movement models that are discrete-time Markov processes of the form $(\boldsymbol{X}_t, t \in T)$, where $\boldsymbol{X}_t \in \mathbb{R}^2$ is an individual's location and $T = \{0, \tau, 2\tau, \dots\}$ is a set of regularly spaced times. This means that we assume that the time interval $\tau > 0$ between two successive location measurements is fixed. Such data often arise from terrestrial animals fitted with GPS devices (Frair et al 2010). The time interval τ of the model is usually specified by the resolution of the data. We denote the one-step transition density for the probability of moving from location $\boldsymbol{y}$ to $\boldsymbol{x}$ between times $t-\tau$ and t by $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a vector of model parameters. This notation highlights that the transition density can be time-heterogeneous.

We consider sub-models that consist of every nth location of the original model for $n \in \mathbb{N}$. The transition density of the nth sub-model for the probability of moving from location y to x between times $t-n\tau$ and t is denoted by $p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$; compare Fig. 1. A priori, the function $p_{t-n\tau,t}$ can have an entirely different form than $p_{t-\tau,t}$ and may correspond to a different probability distribution. However, via the Chapman-Kolmogorov equation, the n-step transition density can be written as a marginal density,

$$p_{t-n\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}) = \int_{\mathbb{R}^2\times\cdots\times\mathbb{R}^2} p_{\text{joint}}(\boldsymbol{x}_t,\boldsymbol{x}_{t-\tau},\dots,\boldsymbol{x}_{t-(n-1)\tau}|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta})\, d\boldsymbol{x}_{t-\tau}\dots \mathrm{d}\boldsymbol{x}_{t-(n-1)\tau}, \quad (1)$$

where we marginalize over all intermediate locations visited between times $t-n\tau$ and t. For simplicity, we use the general subscript 'joint' to denote any joint density of multiple locations. From the notation of the locations it is clear which joint density is meant. The Markov property of the model allows us to stepwise split up the joint density as follows

$$p_{\text{joint}}(\boldsymbol{x}_t,\boldsymbol{x}_{t-\tau},\dots,\boldsymbol{x}_{t-(n-1)\tau}|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-\tau},\boldsymbol{\theta})\, p_{\text{joint}}(\boldsymbol{x}_{t-\tau},\dots,\boldsymbol{x}_{t-(n-1)\tau}|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}). \quad (2)$$

We can continue this until we obtain

$$p_{t-n\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}) = \int_{\mathbb{R}^2\times\cdots\times\mathbb{R}^2} \prod_{k=1}^{n-1} p_{t-k\tau,t-(k-1)\tau}(\boldsymbol{x}_{t-(k-1)\tau}|\boldsymbol{x}_{t-k\tau},\boldsymbol{\theta})\, d\boldsymbol{x}_{t-\tau}\dots \mathrm{d}\boldsymbol{x}_{t-(n-1)\tau}. \quad (3)$$

Therefore, we can use the one-step densities to calculate the n-step density; compare Fig. 1. For statistical inference, and thus for our robustness concept, the model parameter vector $\boldsymbol{\theta}$ plays a crucial role. Although the n-step density may belong to a different distribution than the one-step density, equation (3) justifies that we use the same parameter $\boldsymbol{\theta}$ in the notation of the n-step density as in the one-step density.

We define robustness in terms of the one-step and n-step densities of a model.

Definition 1 (Robustness of degree n**)** Let $n \in \mathbb{N}$ be finite. A movement model of the above type is *robust of degree n* if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta})) \text{ for all } t \in T \text{ and } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^2. \tag{4}$$

This definition requires that the n-step densities are of the same functional form as the one-step transitions, where parameters of the model are appropriately transformed via the function g_n. This means that if a model is robust, the nth sub-model is in fact the same as the original model but with systematically adjusted parameters. The parameter transformation g_n allows us to extrapolate the original parameter $\boldsymbol{\theta}$ to the coarser temporal discretization of the nth sub-model. Additionally, we can use the nth sub-model to infer the parameter $\boldsymbol{\theta}$ of the original model, because we can invert $g_n(\boldsymbol{\theta})$. Note, however, that this rests on the assumption that the original model defines the process of interest. If, instead we start at the coarser resolution, we would also need surjectivity of the function g_n to conclude the existence of the finer model.

Robustness of degree n has important implications. Given a behavioural process of interest, described by a robust model with parameter $\boldsymbol{\theta}$, we can apply the model not only to data with matching temporal resolution τ but also to coarser data with resolution $n\tau$ (e.g. double time interval for $n = 2$). The parameter estimate $\boldsymbol{\psi}$ that we obtain from the coarser data is in fact an estimate of $g_n(\boldsymbol{\theta})$. From this, we can infer the value of $\boldsymbol{\theta}$ via $\boldsymbol{\theta} = g_n^{-1}(\boldsymbol{\psi})$. Additionally, robustness allows us to compare studies pertaining to the same behavioural process but using data sets with different resolutions. If $\boldsymbol{\theta}$ is the estimate based on the finer data, it can be extrapolated to the coarser scale via the parameter transformation $g_n(\boldsymbol{\theta})$, for all degrees n for which the model is robust.

Robustness as in Definition 1 is a strong condition that we do not expect to hold but in few special cases of the density $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$. However, equation (4) may hold up to a function $v(\boldsymbol{x},\boldsymbol{y})$, where v is a bounded function that could also depend on n or τ. For practical applications, such *approximate* or *asymptotic robustness* may be sufficient. Therefore, we provide two additional definitions.

Definition 2 (Asymptotic robustness of degree n**)** Let $n \in \mathbb{N}$ be finite. A movement model of the above type is said to be *asymptotically robust of degree n* if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ and a function $v : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^+$ with the property $v(\boldsymbol{x},\boldsymbol{y};\tau) - 1 = \mathcal{O}(\tau)$ on $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^+$, such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))\, v(\boldsymbol{x},\boldsymbol{y};\tau) \text{ for all } t \in T \text{ and } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^2. \tag{5}$$

Here, $\mathcal{O}$ denotes the Landau symbol for the order of a function. If a model is asymptotically robust, the n-step densities are not exactly the same as the one-step densities, as was required in Definition 1. However, the discrepancy between the densities is bounded by a function that is proportional to τ. More precisely, for an asymptotically robust model we have

$$1 - C\tau \leq \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))} \leq 1 + C\tau \tag{6}$$

for all $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\theta}$, for some constant $C > 0$. Therefore, if the time interval τ of the model is sufficiently small, the n-step density will closely resemble the one-step density with appropriately adjusted parameters. Asymptotic robustness of degree n implies that robustness of degree n is achieved as $\tau \to 0$, that is when the time interval τ approaches zero.

In applications, the time interval τ may not be chosen sufficiently small for Definition 2 to be useful. Therefore, we give a variation of Definition 2, in which the function v does not depend on τ.

Definition 3 (Approximate robustness of magnitude δ and degree n) Let $n \in \mathbb{N}$ be finite. A movement model of the above type is said to be *approximately robust of magnitude δ and degree n* if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ and a function $v : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^+$ with the property $0 < 1 - \delta \leq v(\boldsymbol{x},\boldsymbol{y}) \leq 1 + \delta$ for all $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^2$, for a $\delta > 0$, such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))\, v(\boldsymbol{x},\boldsymbol{y}) \text{ for all } t \in T \text{ and } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^2. \tag{7}$$

Analogously to equation (6), condition (7) can be written as

$$1 - \delta \leq \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))} \leq 1 + \delta. \tag{8}$$

In fact, we may consider two different types of magnitudes. Setting

$$v(\boldsymbol{x},\boldsymbol{y}) := \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))}, \tag{9}$$

this function depends a priori on the parameters, that is we have $v(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})$, and the magnitude is $\delta_{\boldsymbol{\theta}}$. If $\max_{\boldsymbol{\theta}} \delta_{\boldsymbol{\theta}}$ exists, then this is the overall magnitude for the model with all possible parameter values. The magnitude determines how close n-step densities are to the parameter-adjusted one-step densities. If δ is small, then the correction function v is close to one everywhere, and thus the n-step density has similar values as the one-step density over its entire domain.

Asymptotic and approximate robustness have similar implications for inference as robustness, but only approximately. The quality of the approximation depends on τ or the magnitude δ. Suppose we wish to estimate parameters of a behavioural process that we formulate in a model. Suppose we consider the time interval τ as suitable for the process. If the model is robust of degree n, we can use data not only at the matching scale but also at a coarser scale. For example, if the model is robust of degree 2, we can use data obtained at time interval 2τ. Because the model is also valid for the coarser scale, we can translate parameter estimates between the scales via the function g_n. If a model is asymptotically or approximately robust, the model is not exactly but still approximately valid for the coarser scale. To see this, consider the likelihood function

$$L_1(\boldsymbol{\theta}|\{\boldsymbol{x}_0,\boldsymbol{x}_\tau,\boldsymbol{x}_{2\tau},\ldots,\}) = \prod_{t\in\{\tau,2\tau,\ldots\}} p_{t-\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-\tau},\boldsymbol{\theta}). \tag{10}$$

If a model is robust of degree n, the likelihood for data at time interval $n\tau$ is

$$\begin{aligned} L_n(\boldsymbol{\theta}|\{\boldsymbol{x}_0,\boldsymbol{x}_{n\tau},\boldsymbol{x}_{(n+1)\tau},\ldots,\}) &= \prod_{t\in\{n\tau,(n+1)\tau,\ldots\}} p_{t-n\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}) \\ &= L_1(g_n(\boldsymbol{\theta})|\{\boldsymbol{x}_0,\boldsymbol{x}_{n\tau},\boldsymbol{x}_{(n+1)\tau},\ldots,\}). \end{aligned} \tag{11}$$

If a model is asymptotically robust, we have instead

$$L_1(g_n(\boldsymbol{\theta}))\cdot(1-C\tau+\mathscr{O}(\tau^2)) \le L_n(\boldsymbol{\theta}) \le L_1(g_n(\boldsymbol{\theta}))\cdot(1+C\tau+\mathscr{O}(\tau^2)), \tag{12}$$

omitting the notation of the data, which is the same as in equation (11). Analogously, for approximate robustness we have

$$L_1(g_n(\boldsymbol{\theta}))\cdot(1-\delta+\mathscr{O}(\delta^2)) \le L_n(\boldsymbol{\theta}) \le L_1(g_n(\boldsymbol{\theta}))\cdot(1+C\delta+\mathscr{O}(\delta^2)). \tag{13}$$

Therefore, if a model is asymptotically or approximately robust of degree n, we may loosely write $L_n(\boldsymbol{\theta}) \approx L_1(g_n(\boldsymbol{\theta}))$, that is the likelihood functions based on data at time interval τ and on data at interval $n\tau$ are approximately the same. Thus, if data at time interval τ is not available, we can analyze data at time interval $n\tau$ instead, using the likelihood L_1 of the original model. Parameter estimates obtained in this way can be translated to the scale τ by using the inverse parameter transformation g_n^{-1}. Such results from statistical inference based on L_1 may be close to results based on the correct L_n, which may be difficult to compute. How close results are depends on the quality of the approximations in Definitions 2 and 3 via τ or δ. For example, if a model is approximately robust with a very small magnitude δ, the likelihood L_1 will describe data at time interval $n\tau$ almost as well as L_n.

3 Analyzing spatially-explicit random walks

We used the robustness definitions to analyze spatially-explicit random walk models. These models merge general movement tendencies of an individual with decisions based on specific characteristics of locations, such as environmental features and available resources. We investigated how the models react when applied to data with increasingly coarser temporal resolution.

Our robustness definitions have two key features. First, the one-step transition densities of the model and the n-step densities of the sub-models need to have the same form. Second, model parameters, which are parameters of the densities, need to be transformed by a known function g_n. We can assume different approaches to investigate robustness properties of a model, depending on whether we have a candidate for the parameter transformation g_n or not. If prior knowledge allows us to investigate robustness for a given or hypothesized parameter transformation, we can calculate and compare the n-step density $p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ and the parameter-adjusted one-step density $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))$. By showing equality of the two densities, we can verify robustness. For complex models, analytical calculations may be difficult, or even impossible. In these cases, we may resort to numerical calculations, especially when approximate robustness is sufficient.

In many situations, we may not know g_n a priori, nor have any anticipation. Or, we may have tested robustness for a hypothesized parameter transformation but got poor results. In these cases, we need to establish some information on possible forms of the parameter transformation. Additionally, for complex models, numerical calculation of the high-dimensional integral required for the n-step density (compare equation (3)) may become inaccurate. A solution is then to draw on the ideas of Monte Carlo sampling. Monte Carlo methods and simulations are useful when probability densities are difficult to compute in closed-form but can conveniently be sampled from (e.g., Robert and Casella 2000). In the following, we demonstrate both approaches for analyzing movement models' robustness.

3.1 Analytical and numerical approach

Spatially-explicit random walks can be created by merging two elements in the transition density of the model. One component is the general movement kernel $k_{\boldsymbol{\theta}_1}(x;y)$, which can be the transition density of any standard random walk, describing the probability that an individual takes a step from y to x if there were no environmental information available. A second part of the model, given by the weighting function $w_{\boldsymbol{\theta}_2}(x)$, rates each possible step based on the location x. The transition densities of the full model takes the form

$$p_{t-\tau,t}(x|y,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{k_{\boldsymbol{\theta}_1}(x;y)\,w_{\boldsymbol{\theta}_2}(x)}{\int_{\mathbb{R}} k_{\boldsymbol{\theta}_1}(z;y)w_{\boldsymbol{\theta}_2}(z)\,\mathrm{d}z}. \tag{14}$$

The integral in the denominator serves as a normalization constant.

For simplicity, we restricted our analysis to the one-dimensional case, that is we assumed that $X_t \in \mathbb{R}$. We further focused on Gaussian kernels $k_{\boldsymbol{\theta}_1}(x;y) = k_\sigma(x;y)$, where $k_\sigma(x;y)$ is a Gaussian density with mean y and standard deviation σ. The weighting function $w_{\boldsymbol{\theta}_2}(x)$ was assumed to be positive everywhere to ensure that equation (14) defines a density. In the following we simply use $\boldsymbol{\theta}$ for the parameter vector of the weighting function, or, when it is clear which parameters refer to the weighting function, we drop the subscript for the parameter in the notation of the weighting function entirely.

Note that the transition density (14) does not depend on time explicitly. Still, as the individual moves through space over time, the centre location y of the kernel shifts. Although the kernel is a function of the distance $\|x-y\|$ only, the weighting function adds a spatially explicit component. Therefore, unless the individual remains at the same location, the transition kernel effectively changes at every time step. In the following, we omit the time-related subscript in the notation of the density and simply write p_1 for the transition density (14) and p_n for the n-step density. The time interval of the original process is always assumed to be τ. The distinction between one-step and n-step density is still important, because the n-step density is in fact an integral over multiple one-step densities; compare equation (3).

We investigated whether we could find weighting functions $w_{\boldsymbol{\theta}}(x)$ such that the model with transition density (14) is robust, asymptotically robust or approximately robust. We started by verifying Definition 1 for simple cases of the weighting function for a fixed parameter transformation g_n. As highlighted above, the parameter transformation is a key element, translating parameters between different temporal resolutions. For the parameter of the Gaussian movement kernel k_σ, we obtained a candidate for the transformation based on the linearity of the Gaussian distribution. If we only consider the kernel k_σ, we have a simple random walk with normally distributed steps between locations. The n-step density (3) is then the density of a sum of n normally distributed random variables, which is again normal with standard deviation $\sqrt{n}\sigma$. Therefore, we assumed that the transformation of the kernel's standard deviation was given by $g_n(\sigma) = \sqrt{n}\sigma$. For the parameters of the weighting function we assumed that they remain unaffected, that is $g_n(\boldsymbol{\theta}) = \boldsymbol{\theta}$. This is an ideal property for a weighting function, as it guarantees validity of inference results across different sampling rates without further translation.

In a next step, we used the same parameter transformation $g_n(\sigma, \boldsymbol{\theta}) = (\sqrt{n}\sigma, \boldsymbol{\theta})$ to establish conditions on the weighting function such that the model is asymptotically robust. For this, we assumed that the parameter of the kernel, the standard deviation, was influenced by the time interval τ, that is $\sigma = \sigma(\tau)$. This reflects that an individual may travel larger distances during longer time intervals. Because of the linearity of the Gaussian distribution, we assumed the relationship $\sigma(\tau) = \sqrt{\tau}\omega$, for some $\omega > 0$. For certain conditions on the weighting function, we verified Definition 2 analytically for the robustness degree $n = 2$ by calculating the function $v(x, y; \tau)$ and placing bounds on it.

As alternative to an analytical approach, we can calculate the ratio of two-step and one-step density numerically to see whether we can find a function $v(x, y; \tau)$ according to Definition 2 for the degree $n = 2$. Define $\delta(\tau) := \max_{x,y} |v(x, y; \tau) - 1|$. Note that since step densities depend on τ through $\sigma(\tau)$, we may equivalently consider $\delta(\sigma)$. If this is independent of the other parameters $\boldsymbol{\theta}$, we can obtain the bound on v as $\delta := \max_\sigma \delta(\sigma)$, if this maximum exists. More generally, we can consider $v(x, y, \sigma, \boldsymbol{\theta})$ and calculate $\delta_{\boldsymbol{\theta}}(\sigma) := \max_{x,y} |v(x, y; \sigma, \boldsymbol{\theta}) - 1|$. This $\delta_{\boldsymbol{\theta}}(\sigma)$ is the magnitude of approximate robustness (degree 2) for a model with a fixed weighting function, including parameter values. An overall magnitude for the family of models consisting of the model for all parameter values can be obtained as $\delta := \max_{\sigma,\boldsymbol{\theta}} \delta_{\boldsymbol{\theta}}(\sigma)$. We demonstrate these two numerical approaches with an example weighting function.

3.2 Simulation approach

3.2.1 Resource selection models

Resource selection analyses link animal location data and environmental variables to understand animals' space-use patterns in relation to their habitat. These studies provide insight into species' preferences or avoidance of habitat characteristics, which is important information for wildlife management and conservation purposes (Hebblewhite and Merrill 2008; Latham et al 2011; Squires et al 2013). Central methodological elements are resource selection functions (RSF) and resource selection probability functions (RSPF), describing the probability of selection of certain units (e.g. pixels of land) by an organism based on environmental covariates (Manly et al 2002; Boyce et al 2002; Lele and Keim 2006). RSF and RSPF have been used on their own in a mere statistical framework (Boyce et al 2002; Courbin et al 2013), incorporated into spatially-explicit models (Rhodes et al 2005; Aarts et al 2011), and become part of mechanistic movement models (Moorcroft and Barnett 2008; Potts et al 2014)). We refer to Lele et al (2013) for details about the distinction of RSF and RSPF and use RSF as a general term for both concepts, unless otherwise stated.

We include resource selection in the spatially-explicit random walk with transition density (14) by letting the weighting function take the form of an RSF, $w_{\boldsymbol{\theta}}(x) = w_{\boldsymbol{\theta}}(\boldsymbol{r}(x))$, where $\boldsymbol{r}(x) = (r_1(x), \ldots, r_n(x))$ is a vector of resource covariates at location x. Each r_j is a function over space, representing resource covariates such as elevation, biomass measures, land cover type, and much more. The transition density becomes

$$p_1(x|y, \sigma, \boldsymbol{\theta}) = \frac{k_\sigma(x;y)\, w_{\boldsymbol{\theta}}(\boldsymbol{r}(x))}{\int_{\mathbb{R}} k_\sigma(z;y) w_{\boldsymbol{\theta}}(\boldsymbol{r}(z))\, \mathrm{d}z}. \tag{15}$$

In practice, geographical information is spatially discrete, and therefore the normalizing integral in equation (15) becomes a sum over pixels, or cells, of land. Note that we still restrict our attention to one-dimensional models.

The RSF can take various forms, and here we consider the two most commonly used ones (Manly et al 2002; Lele and Keim 2006), the exponential RSF,

$$w_{\exp}(\boldsymbol{r}(x)) = \exp(\boldsymbol{\beta} \cdot \boldsymbol{r}(x)) \tag{16}$$

and the logistic function,

$$w_{\log}(\boldsymbol{r}(x)) = \frac{\exp(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{r}(x))}{1 + \exp(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{r}(x))}. \tag{17}$$

The vector $\boldsymbol{\beta}$ comprises all selection parameters with respect to resource covariates $\boldsymbol{r}$. A higher selection parameter means stronger selection with respect to the corresponding resource. In the logistic form, α is an intercept parameter, which can shift the inflection point of the logistic function away from zero. The inflection point is the point where the logistic function attains a value of 0.5, that is where the probability of selecting a resource is 50%. If the exponential form (16) is used, an intercept similarly to the one used in equation (17) is not identifiable, because it cancels in the definition of the transition density (15). Therefore we have omitted it in equation (16). The function $w_{\log}$ has range $(0,1)$ and can therefore be used to describe probabilities. This means that this form can be used as RSPF, which for a given location y specifies the probability that an animal selects this location, given the covariate values of the location. In contrast, the exponential RSF can only specify values proportional to this probability, with unknown proportionality constant (Lele et al 2013).

3.2.2 Sampling models and sub-models

We examined the two models with weighting functions $w_{\exp}$ and $w_{\log}$ for their robustness. Because the weighting functions depend on space through environmental information $\boldsymbol{r}$ they are highly non-linear, and therefore the transition densities are difficult to examine analytically. Sampling probability distributions is a convenient work around and has the additional advantage that we can control parameters and isolate processes of interest. We thus simulated sample trajectories from the model with transition densities (15). The joint density of a movement trajectory $(x_1,\ldots,x_N)\in\mathbb{R}^N$ of length $N\in\mathbb{N}$ is given by

$$p_{\text{joint}}(x_1,\ldots,x_N,\boldsymbol{\theta}) = p_1(x_1,\boldsymbol{\theta})\prod_{t=2}^{N} p_1(x_t|x_{t-1},\boldsymbol{\theta}). \tag{18}$$

Thus, we sampled successively from the transition densities to obtain a full movement trajectory. We obtained samples from the subprocess $\boldsymbol{x}_n = (x_1, x_{n+1}, \ldots)$ consisting of every nth location by subsampling the full trajectories. These subsamples represent samples from the model with transition densities being the n-step densities $p_n(\cdot|\cdot,\boldsymbol{\theta})$.

Because the models rely on environmental data, we simulated resource landscapes as realizations of Gaussian random fields with exponential covariance model (Haran 2011; Schlather et al 2013). This resulted in spatially correlated resource landscapes, thus ensuring realism; compare Fig. 1 in Online Resource 1. The sampled movement trajectories were based on these simulated landscapes. To avoid confounding effects and to keep results as clear as possible, we assumed that the weighting function was based on only one resource r, thus we have $w_{\boldsymbol{\theta}}(r(x))$. With the exponential covariance model, we assumed that the covariance of resource values at two different locations is given by

$$\operatorname{Cov}(r(x),r(y)) = \exp\Big(\frac{|x-y|}{s}\Big), \tag{19}$$

where s affects the decrease of the spatial autocorrelation with increasing distance.

We sampled trajectories for varying parameter values. We used $\sigma\in\{5,6,7\}$ and $\beta\in\{0.5,1,1.5,2\}$ in all combinations. In the model with logistic RSF $w_{\log}$, we further combined the values $\alpha\in\{-1,-0.5,0,0.5,1\}$ with all other parameters. For each parameter combination, we sampled 16 trajectories for 15,000 time steps each; compare Fig. 2,3 in Online Resource 1. For each of the 16 trajectories, we used a different resource landscape, repeating the same set of resource landscapes across different parameter combinations. The 16 landscapes were generated with varying spatial autocorrelation, s ranging between 200–500. This led to a total of 192 sampled trajectories for the model with exponential RSF and 960 trajectories for the model with logistic RSF. We subsample every trajectory at levels $n = 1,\ldots,15$, leaving 1000 steps for the coarsest time series. The subsample for $n = 1$ is the original trajectory.

3.2.3 Analyzing parameters

While the simulated trajectories represent samples from the original model with transition densities $p_1(\cdot|\cdot,\boldsymbol{\theta})$, the subsamples of the full trajectories provide us with samples from the sub-models with n-step densities $p_n(\cdot|\cdot,\boldsymbol{\theta})$. To learn about the model's robustness properties, we need to test whether the subsamples reconcile with the parameter-adjusted one-step densities $p_1(\cdot|\cdot, g_n(\boldsymbol{\theta}))$ for some parameter transformation g_n. For a given parameter transformation, we can achieve this by analyzing the fit of the model with transitions $p_1(\cdot|\cdot, g_n(\boldsymbol{\theta}))$ with the subsamples. When g_n is unknown, or when the fit for a hypothesized g_n is poor, we first need to investigate the behaviour of the parameters under subsampling to see whether we can find a function g_n as required by our robustness definitions.

Here, we both tested a priori expectations on the parameter transformation and searched for better alternatives. We estimated parameters for all trajectories and their subsamples using maximum likelihood optimization. The likelihood function for the full trajectories is given in equation (18). For subsamples, we applied the same model, although we did not know whether subsamples of trajectories followed the same (parameter-adjusted) process as full trajectories. We expected parameter estimates for the full trajectories to be close to the values that we used during the simulations. We call these the ‘true values’, although deviations in the simulations are possible, because simulated trajectories are realizations of stochastic processes. Our main interest are parameter estimates for the subsamples. To distinguish estimates from underlying true parameters, we denote the estimate with a hat, e.g. $\hat{\sigma}$. Ideally, the parameters of the subsamples should follow some function $g_n(\sigma, \alpha, \beta)$, and so should the estimates. To see whether such a function exists, we fitted non-linear regression models to the relationship of parameter estimates of subsamples and the subsampling amount n. For each parameter, we fitted two models. One model was more restrictive and represented a priori expectations, whereas the other model had an additional free parameter that allowed more flexibility for the parameter transformation.

The general movement kernel k has one parameter, the standard deviation σ of the Gaussian distribution. This kernel describes the general movement tendencies of the animal, and σ influences the distance covered in each step. With increasing subsampling, the temporal resolution of the movement path becomes coarser, and we thus expected the standard deviation of the kernel to increase. Each step in a subsample is in fact the accumulated result of one or several steps in the full trajectory. If the kernel is the only force driving the movement, the linearity of the Gaussian distribution caused us to expect the standard deviation of the kernel to increase as $\sqrt{n}\sigma$; compare section 3.1. With additional resource selection, however, there may be deviations from this behaviour.

For the resource selection parameters α and β, an ideal behaviour would be that they remain unaffected by the subsampling, analogously to our assumptions in section 3.1. In our model, we assume that each step is influenced by the RSF. One of the underlying assumptions of a traditional RSF is that it gives weights to locations independently of the values of other locations, which means each location is weighted by its present resource only, without consideration of alternative locations. Therefore, resource selection parameters should be independent of the temporal resolution of the data. However, within the spatially-explicit movement framework, resource selection always occurs in the context of the current location and the available surrounding area as defined by the general movement kernel. Therefore, a change in the movement kernel due to increased subsampling may be accompanied by a change in resource selection parameters.

We fitted the non-linear regression models to the parameter estimates separately for each parameter combination. This means that in each regression, we fitted estimates of 16 trajectories and their subsamples. Because of our previous considerations about the kernel parameter σ, we assumed a power relationship between the estimate $\hat{\sigma}$ and the subsampling amount n, stratified by trajectories. We chose the stratification because trajectories were simulated on different landscapes. Also, for the resource selection parameters, especially when their true values were close to zero, estimates could vary between being positive and negative. In these cases, the stratification allowed for flexibility. The model for the estimate of the nth subsample of trajectory i is

$$\hat{\sigma}_{i,n} = \hat{\sigma}_{i,1} \cdot n^b + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16, \tag{20}$$

where the error term ε is normally distributed with mean zero and positive standard deviation ζ. The maximum likelihood estimate of b should ideally be close to 0.5, however as noted above, it may deviate from this value because of resource-selection mechanisms. To test whether b differs from 0.5, we used model selection via AIC between the model in equation (20) and the model in which we fixed $b = 0.5$.

Model choice for the resource selection parameters was less clear. Visual inspection of the estimates, preliminary fits with varying models and inspection of residuals suggested a power law for the parameter β as well. We thus fitted the following model,

$$\hat{\beta}_{i,n} = \hat{\beta}_{i,1} \cdot n^b + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16. \tag{21}$$

We compared the fit of this model with the model in which we assumed that subsampling does not change the estimate by setting $b = 0$.

For the intercept parameter α in the logistic form of the resource selection function, we chose a linear model,

$$\hat{\alpha}_{i,n} = \hat{\alpha}_{i,1} + b\,(n-1) + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16. \tag{22}$$

Inspection of residuals suggested that in some cases the relationship between $\hat{\alpha}$ and n was non-linear. However, a power-law model or other non-linear relationships were not consistently more suitable either. Therefore we remained with the simpler, the linear, model, noting that this is a mainly illustrative analysis.

3.2.4 Calculating approximate robustness

To accompany the simulation analysis, we examined approximate robustness properties of the two models with exponential and logistic RSF. We focused on approximate robustness of degree 2, and we tested the ideal parameter transformations $g_2(\sigma,\beta) = (\sqrt{2}\sigma,\beta)$ and $g_2(\sigma,\alpha,\beta) = (\sqrt{2}\sigma,\alpha,\beta)$ for w_{exp} and w_{log}, respectively. We numerically calculated a magnitude $\delta = \max_{x,y}(|v(x,y)-1|)$ for every possible scenario that we used in the previous section. This means that we calculated a magnitude for each combination of the parameters σ, β, and α (in case of the logistic RSF) and for each of the 16 simulated resource landscapes. We may therefore think of δ as $\delta(\sigma,\alpha,\beta,i)$, for $1 \leq i \leq 16$; compare Fig. 2 We examined whether magnitudes were influenced by parameter values and specific characteristics of the landscapes, such as their spatial autocorrelation and their overall variation $\mathrm{Var}(r(x))$ over the spatial domain. We further calculated an overall maximum $\max_{\sigma,\alpha,\beta,i} \delta(\sigma,\alpha,\beta,i)$. We compared results between the model with exponential RSF, $w_{\exp}$, and logistic RSF, $w_{\log}$.

4 Results

4.1 Analytical and numerical results

We found few special cases of weighting functions $w_{\boldsymbol{\theta}}$ that, together with the Gaussian kernel k_σ, resulted in a robust movement model according to Definition 1.

The simplest case was a constant weighting function. Such a weighting function reduces equation (14) to the case of a homogeneous environment, where only general movement tendencies play a role, but no environmental information. The model is then a simple random walk with normally distributed steps between locations. Because of the linearity of the normal distribution, the model is robust of degree n for all $n \in \mathbb{N}$ for the assumed parameter transformation $g_n(\sigma) = \sqrt{n}\sigma$; compare also Theorem 2 for parameters $a = b = 0$.

A natural next step was to consider a linear weighting function. However, a linear weighting function violates the assumption of being strictly positive everywhere. If in equation (14) the current location y is the point at which w becomes zero, the normalization integral vanishes. Also, equation (14) can become negative and thus cease to be a valid density function. Still, we could draw on the linearity of the expectation of a random variable to look into this further. The normalization constant in the transition density (14) can be viewed as an expectation of the form $E(w(Z))$ for a normally distributed random variable Z with mean y. Therefore, if the function w is linear, the normalization constant reduces to $w(y)$. Equation (14) then becomes

$$p_1(x|y,\sigma,\boldsymbol{\theta}) = k_\sigma(x;y)\frac{w_{\boldsymbol{\theta}}(x)}{w_{\boldsymbol{\theta}}(y)}. \tag{23}$$

The right-hand side of the equation is positive whenever x and y are either both negative or both positive. If movement only occurs in the domain where the weighting function is positive the model is robustness within this domain. The details of the proof can be found in Appendix A.

Theorem 1 (Linear weighting function) *Let w be a linear function $w(x) = ax + b$, for $a, b \in \mathbb{R}$. Let $\mathscr{I} \subset \mathbb{R}$ be the interval where $w > 0$. For the restricted domain $\mathscr{I}$, the movement model with transition densities* (14) *is robust of degree n for all $n \in \mathbb{N}$. The parameter transformation is given by $g_n(\sigma, a, b) = (\sqrt{n}\sigma, a, b)$.*

We found another special case to be given by an exponential weighting function. Here, no restriction on the domain is necessary. Again, see Appendix A for details of the proof.

Theorem 2 (Exponential weighting function) *Let w be an exponential function of the form $w(x) = Ce^{ax+b}$ for $C, a, b \in \mathbb{R}$. Then the movement model with transition* *densities* (14) *is robust of degree n for all $n \in \mathbb{N}$ with parameter transformation $g_n(\sigma, C, a, b) = (\sqrt{n}\sigma, C, a, b)$.*

The above two Theorems show that it is possible to verify exact robustness with the ideal parameter transformation $g_n(\sigma, \boldsymbol{\theta}) = (\sqrt{n}\sigma, \boldsymbol{\theta})$ for certain weighting functions. However, the cases are very restrictive, and robustness will fail for many other, and especially more complex, weighting functions.

We could additionally establish asymptotic robustness for more general conditions on the weighting function. The main result is summarized in the following theorem. For a detailed proof of the theorem, see Appendix B.

Theorem 3 (Asymptotic robustness of degree 2) *Let $w_{\boldsymbol{\theta}}$ be continuous and bounded away from zero. Let $w_{\boldsymbol{\theta}}$ further be twice differentiable with bounded second derivative. Then the model with transition densities* (14) *is asymptotically robust of degree 2 with parameter transformation $g_2(\sigma, \boldsymbol{\theta}) = (\sqrt{2}\sigma, \boldsymbol{\theta})$.*

Thus, if the weighting function is well-behaved according to the theorem, we can place a bound on the factor by which the one- and two-step density vary; compare equation (6). This bound is of order τ, such that the discrepancy between one- and two-step density decreases with the time interval.

Example 1 (Asymptotic robustness of degree 2) As a simple example, consider the weighting function $w(x) = \sin(\alpha x) + \beta$ for $\alpha > 0$ and $\beta > 1$. The choice of β guarantees that the weighting function is positive everywhere. The function w is bounded between $0 < \beta - 1 \leq w(x) \leq \beta + 1$ for all $x \in \mathbb{R}$, and its second derivative is bounded by $|w''(x)| = \alpha^2$. Therefore, Theorem 3 holds.

The proof of Theorem 3 is constructive in the sense that it provides us with a constant C for equation (6) in terms of the bounds on w and w''. However, this constant may be rather large and does not necessarily provide the closest bound on the function v. Therefore, it can be informative to calculate approximate robustness numerically.

Example 2 (Approximate robustness of degree 2) We continue the above example with weighting function $w(x) = \sin(\alpha x) + \beta$ for $\alpha > 0$ and $\beta > 1$. We calculated the function $v(x, y; \sigma, \alpha, \beta)$ from Definition 3 numerically, using different values of α and β (Fig. 3a). From this, we obtained $\delta_{\alpha,\beta}(\sigma)$ (Fig. 3b), which is the magnitude of approximate robustness (degree 2) for the model with specific weighting function (i.e. with specific parameters); compare Fig. 2. In each case, after reaching a maximum the function vanishes for increasing σ. Therefore it appears that we can find $\delta_{\alpha,\beta} := \max_\sigma \delta_{\alpha,\beta}(\sigma)$. The wavelength of the sine curve, determined by α, and the intercept β have different effects on the function $\delta_{\alpha,\beta}(\sigma)$. While α shifts the curve, β changes the height of the peak (Fig. 3b). Therefore, it appears that $\delta_{\alpha,\beta}$ is independent of α and decreases for larger β. For the weighting function to be positive, β needs to be larger than one. For $\beta = 1$, the function $\delta_{\alpha,\beta}$ has a maximum at one. From these considerations, we can conclude that $\max_{\alpha,\beta} \delta_{\alpha,\beta} = 1$. This is the overall magnitude of approximate robustness (degree 2) for the family of weighting functions $w(x) = \sin(\alpha x) + \beta$, $\alpha > 0$, $\beta > 1$; compare Fig. 2 As a word of caution, we note that we only calculated $\delta_{\alpha,\beta}$ for a fixed number of parameter values and only within finite intervals for x and y, and therefore results may be limited to these ranges.

In the region where $\delta(\sigma)$ peaks, the approximation of the parameter-adjusted one-step density $p_1(x|y,\sqrt{2}\sigma,\alpha,\beta)$ to the actual two-step density $p_2(x|y,\sigma,\alpha,\beta)$ is only rough. However, for larger values of σ, and independent of α and β, the function $\delta_{\alpha,\beta}(\sigma)$ seems to vanish, which means that the approximation is good and the discrepancy between two- and one-step densities may be neglected. From Theorem 3, we would have been able to conclude that $\delta_{\alpha,\beta}(\sigma(\tau))$ is bounded by $C\tau$, for a constant $C > 0$, for all $\alpha > 0$ and $\beta > 1$. As we can see from the steep initial slope of $\delta_{\alpha,\beta}(\sigma)$, especially for higher values of α, the constant C would need to be rather large (Fig. 3b). The calculations of approximate robustness could additionally show that the bound on $v(x,y)$ is in fact much smaller.

4.2 Simulation results

4.2.1 Results for parameter estimates

When analyzing parameter estimates from the simulated trajectories and their subsamples, we found a difference in the behaviour of parameters between the exponential and the logistic form of the RSF. Generally, subsampling had less effect on the value of parameter estimates using the logistic form, and the behaviour of estimates agreed closer with our expectations.

For both RSF, estimates $\hat{\sigma}$ showed a good fit with the power-law model. When we used the exponential RSF, the estimated power b ranged from 0.45 to 0.5 for varying parameter combinations, thus deviating from expected behaviour for some parameter combinations (Fig. 4a). For small selection parameter β, the estimate $\hat{\sigma}$ showed the expected increase as $\hat{\sigma}\sqrt{n}$. With increasingly strong selection, i.e. higher value of β, estimates $\hat{\sigma}$ became smaller with increased subsampling relative to the ideal relationship. An increase in σ did not influence the fit other than leading to a larger residual standard error $\hat{\zeta}$, which is to be expected because of the overall larger values of the dependant variable. In contrast, when using the logistic RSF, the estimated power b differed only very slightly from 0.5 and in some cases, the simpler model with fixed b was preferred by model selection right away (Fig. 4b).

The behaviour of the resource-selection parameter β also differed between exponential and logistic RSF. For the exponential RSF, $\hat{\beta}$ showed a clear increase with increased subsampling, fitted well by our power-law model (Fig. 5a). The power b remained similar (ranging 0.105–0.124) across parameter combinations, increasing slightly with larger σ (Fig. 5b). For the logistic RSF, estimates $\hat{\beta}$ generally remained closer to the original values for $n = 1$ (Fig. 5c,d). In most cases, model selection via AIC preferred the power-law model to the ideal constant relationship, however, the estimated values of the power b are small, with 53 out of 60 values being below 0.1 (total range 0–0.156, with one exceptional negative value $b = -0.041$). There was a tendency of b to be smaller and more concentrated under stronger selection (Fig. 5d).

Estimates of the intercept α in the logistic RSF showed a slight decline with increased subsampling in most cases (Fig. 6). This decreasing trend existed no matter whether α was positive, negative, or zero. In general, slopes of the linear fit were all close to zero (ranging -0.047–0.058), and in a few cases the null model with $b = 0$ was chosen. We found a trend in the realized intercept values in the simulated trajectories. With stronger effect of selection (larger β), the intercept estimate $\hat{\alpha}$ of original trajectories ($n = 1$) was stronger concentrated around the true underlying value, which subsequently lead to a stronger concentration of estimates of subsamples (Fig. 6).

4.2.2 Results about approximate robustness

When comparing magnitudes $\delta(\sigma, \alpha, \beta, i)$ of approximate robustness (degree 2) between the two models with exponential and logistic RSF, we found lower magnitudes for the model with logistic function $w_{\log}$. Magnitudes for the model with exponential RSF ranged between 0.067 and 1.82, whereas those for the model with logistic RSF ranged between 0.02 and 1.19. The 5% quantile, the median and the 0.95% quantile were $[0.092, 0.34, 0.97]$ (exponential RSF) and $[0.046, 0.21, 0.64]$ (logistic RSF).

We found that especially the selection parameter β had a strong influence on magnitudes, higher values of β leading to higher magnitudes (Fig. 7). For the model with exponential RSF, there was a tendency that weighting functions whose underlying landscapes had higher variation $\mathrm{Var}(r(x))$ lead to smaller magnitudes (Fig. 7a). However, we did not find an effect of the parameter s that was used in the simulations to influence the spatial autocorrelation of the landscapes. The model with logistic RSF did not show such an effect of landscape variation. The logistic model had the additional intercept parameter α. We found that higher values of α tended to result in lower magnitudes (Fig. 7b).

5 Discussion

We have proposed a new rigorous framework for analyzing movement models' capacities to compensate for varying temporal discretization of data. Our framework comprises three definitions of varying strength for robustness of discrete-time movement models. Generally, if a model is robust, it can overcome problems of mismatching temporal scales between different data sets or between data and biological questions. Because our robustness is a very strong condition that holds only for very few and generally more simple models, we have introduced the additional concepts of asymptotic and, most importantly, approximate robustness. While for many movement models it is difficult, or even impossible, to examine the transition densities and their marginals analytically, approximate robustness properties of a model can be calculated numerically also for analytically intractable models. Therefore, we believe that especially approximate robustness will prove a useful new concept for movement analyses.

We have formulated our robustness definitions in terms of the transition densities of Markov models, because these models are often fitted to movement data with likelihood-based methods of statistical inference. For the considered models, we can obtain the likelihood function by multiplying the transition densities of subsequent steps. If a model is robust, the transition densities keep their functional form across varying temporal scales, and parameters are transformed via a well-defined function g_n. The likelihood function therefore remains the same but will yield different parameter estimates. However, if the parameter transformation is known, estimates from one scale can be translated to estimates at other scales. If a model is only approximately robust, the likelihood function will not remain exactly but at least approximately the same under a change of scale. Depending on the magnitude of the approximate robustness, the approximation of the likelihood function may be sufficiently good to allow parameter estimates to be reasonably comparable for different scales, especially if the difference in scales is small.

5.1 Relationship of the framework to statistical robustness

Our concept of robustness for discrete-time movement models is related to the formal concept of robustness in statistics. Generally speaking, robust methods in statistics acknowledge that models are approximations to reality and seek to protect outcomes of statistical procedures (e.g. hypothesis testing, estimation) against deviations from the underlying model assumptions. Classic examples are the arithmetic mean and median as estimates of a population mean: while the median is robust against outliers the mean is not (e.g. Hampel 1986). Often, robustness is viewed in the context of deviations from assumed probability distributions (distributional robustness; e.g. Huber and Ronchetti 2009). For example, data may be contaminated by few observations with heavier tailed distribution than the majority of the observations. In regression analyses, robustness may also relate to the homoscedasticity assumption or the functional form of the response function (Wiens 2000; Wilcox 2012). Additionally, robustness has been considered when the assumption of independence is violated and instead observations are correlated (Hampel 1986; Wiens and Zhou 1996). In our paper, we consider robustness in the context of discrete-time movement models with respect to assumptions about the temporal discretization. In view of statistical robustness, we study violations against the assumption that the temporal resolution of our movement model, a stochastic process, matches the resolution of the data, when in fact the data is only a subsample of the assumed process.

There is also a difference between our robustness of movement models and the well-established robustness in statistics. In our framework, robustness is a direct property of a model. In contrast, classical robustness in statistics is defined for objects such as estimators, test-statistics, or more generally, functionals (real-valued functions of distributions) (Hampel 1971, 1986). For the type of models we have considered here, parameter estimates cannot be obtained analytically but through numerical optimization of the likelihood function. The likelihood function is build by the model's transition densities, and thus we have defined robustness at a very basic level. A possibility for future research is to investigate whether some of the formal concepts of statistical robustness can be applied to our framework to add further insight. With our paper, we provide a new perspective for studying effects of temporal discretization of movement processes, and we hope to encourage further research.

5.2 Discussion of analytical results

Our analytical investigations indicate that robustness is a rare property among movement models, especially when behavioural mechanisms such as resource selection are added. Therefore, if we apply models to data without considering this issue, we are in danger of misinterpreting results and drawing erroneous conclusions. However, our analysis also shows positive prospects with respect to approximate robustness. Theorem 1 suggests that in slowly varying environments that produce locally linear weighting functions we may find some robustness. Theorem 3 and the following examples show that certain smoothness and boundedness conditions on the weighting function can lead to approximate robustness. In addition, Example 2 further demonstrates that approximate robustness can be investigated numerically on a case-by-case basis. We have illustrated this with a smooth weighting function $w(x)$ that is a direct function of space. In data applications, an animal's preferences for locations usually do not depend on space per se but rather through the type of habitat and available resources, and the weighting function will be less regular. In our simulation study, we have therefore presented a case with a more realistic model.

5.3 Discussion of resource selection simulation study

While it is difficult to analyze the transition densities and resulting n-step densities with analytical calculations, we have demonstrated with the simulation approach how we can still investigate robustness properties of complex models. Sampling from probability distributions instead of calculating them is the key idea of Monte Carlo methods. We have used this method to examine sub-models that have the n-step densities as transition densities. With this we obtained the parameter transformation g_n. Our approach differs from previous studies that have used subsamples of fine-scale data to establish an empirical relationship between sampling interval and movement characteristics (Pépin et al 2004; Ryan et al 2004; Rowcliffe et al 2012). When using data, it can be difficult to tease apart effects that result from the methodology and effects of actual behavioural changes at different scales. Analyzing model properties as we have proposed here allows us to identify those effects of temporal discretization that are attributable to the methodology.

In our demonstration of the simulation approach, we analyzed spatially-explicit resource selection models. These models have an advantage over traditional resource-selection and step-selection functions. In the traditional, regression-type approach, observed movement steps are compared to potential steps that are obtained separately from an empirical movement kernel (Fortin et al 2005; Forester et al 2009). In this approach, movement and resource-selection are treated independently, although it is highly likely that both influence each other. In contrast, when fitting the full random walk with resource selection to data by using the likelihood function (18), we can simultaneously estimate parameters both of the general movement kernel and the weighting function, that is the RSF.

In our analysis of the resource-selection model, we observed systematic trends in values of parameter estimates with changing temporal discretization of movement trajectories. The main purpose was not to analyze these relationships in full detail but to illustrate that they occur and thus must not be neglected. Comparing the exponential and logistic form of the spatially-explicit resource selection model, we found that estimates varied more with increased subsampling when the exponential RSF was used, compared to the logistic RSF. Using the exponential RSF, estimates of the kernel standard deviation σ decreased with increased subsampling compared to the ideal relationship $\sqrt{n}\sigma$. On the other hand, using the logistic RSF, σ followed the ideal relationship that would occur for a purely Gaussian process very closely, even under additional influence of resource selection. The estimated strength of resource selection, indicated by β, increased with the subsampling amount. While this effect was strongly pronounced for the model with exponential RSF, it was only weak for the logistic RSF. Therefore, if using the logistic RSF, one may expect to obtain similar inference results across varying temporal discretization.

When we calculated the magnitudes of approximate robustness for the models used in the simulations, we found that those were in line with the results for the parameter estimates. Overall, the model with logistic RSF had better robustness properties than the model with exponential RSF. We also found a matching trend for the movement parameter σ with varying true values of β. Estimates of σ were closer to the expected behaviour for weaker resource-selection parameters. This was also reflected in magnitudes of approximate robustness. If selection was weaker in the original model, the model exhibited better robustness properties. These results suggest that numerical calculations of approximate robustness can assist our expectations about changes in parameter estimates. On the other hand, although parameter estimates of the weighting function showed a clear difference in behaviour when comparing between the exponential and logistic RSF, differences within one model between different parameter combinations were less clear. More analyses would be required to entangle more detailed effects.

Overall, the results from the simulations suggest that depending on the resolution of movement data, we may misinterpret animals' movement tendencies and also may overestimate resource selection effects. It is therefore important that we are aware of the changes to statistical inference that can arise merely from the methodology. Here, we have seen that changes in inference results were stronger for the resource selection model with exponential RSF compared to the logistic RSF. A possible explanation may be the additional intercept in the logistic RSF. With increased subsampling, estimates of α tended to decrease, possibly counteracting the increase in estimates $\hat{\beta}$. This could have led to more stability for the parameter σ of the general movement kernel. However, this may not explain why resource selection parameters generally varied less themselves compared to the exponential RSF. Another possibility is that the different form of the RSFs causes their different behaviour. While the exponential form of the RSF greatly enhances differences in landscape values, the logistic RSF is restricted to values in the interval $(0,1)$. Theorem 3 suggests that variation in the rate of change of the weighting function influences robustness properties. Thus the logistic RSF may produce more stable inference results for varying temporal resolutions. Lele and Keim (2006) suggested several alternatives to the exponential RSF. Our study case showed that the choice of resource selection functions can have implications for statistical inference and we encourage to choose resource selection functions more deliberately.

5.4 Concluding remarks

With our study we have illustrated that the concept of robustness and its parameter transformation g_n can help to bridge the gap between different temporal resolutions of data. For example, in the resource-selection model with exponential RSF, we found that with increased subsampling estimates of the resource selection parameter β deviated strongly from the original values, however, the increase in $\hat{\beta}$ could be fitted with a power-relationship. Thus, using the idea of Monte Carlo sampling, we were able to obtain a parameter transformation g_n that links parameter values between different temporal resolutions. Using such transformations when comparing results obtained from data with different temporal resolutions could greatly improve our statistical inference, leading to a better understanding of movement behaviour.

At the same time, we would like to reiterate that robustness of movement models cannot replace careful design of movement studies and data collection. To obtain reliable results, it is crucial to acquire movement data with a resolution that is fine enough to hold information about the behavioural process of interest. Our robustness concept can then be used to mitigate between different resolutions within this temporal scale of interest. Additionally, robustness considerations should not trump biologically meaningful model properties. For example, in many situations a scale-free random walk may not be a suitable model (James et al 2011; Pyke 2015) although it is robust (Schlägel and Lewis 2016). We therefore emphasize the importance of careful model choice while adding the framework of movement model robustness as a new tool to evaluate models' sensitivity to temporal discretization. With our study, we hope to deepen our insight into the problem and to encourage further research.

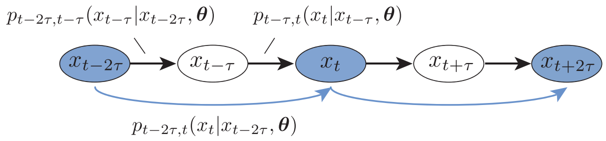

Fig. 1 The second sub-model consists of every second location. The transition densities of the sub-model, which we refer to as 2-step densities, are the marginals over the two intermediate one-step densities of the original model

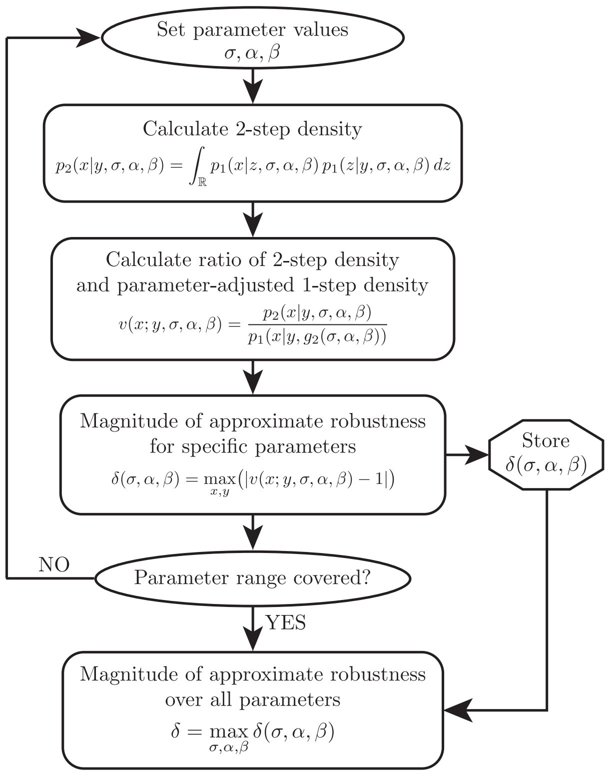

Fig. 2 Steps for calculating the magnitude of approximate robustness of degree 2 for a given model, where σ is the parameter of the movement kernel, and α and β are parameters of the weighting function. The one-step density p_1 can, for example, be equation (14) with the weighting function from Example 2, or the resource selection model (15) with weighting function (16) or (17). When the resource selection model is used, the flowchart shows the calculation of the magnitude for one specific resource landscape $r(x)$. When calculating an overall magnitude, practically we do this for a subset of the parameter space

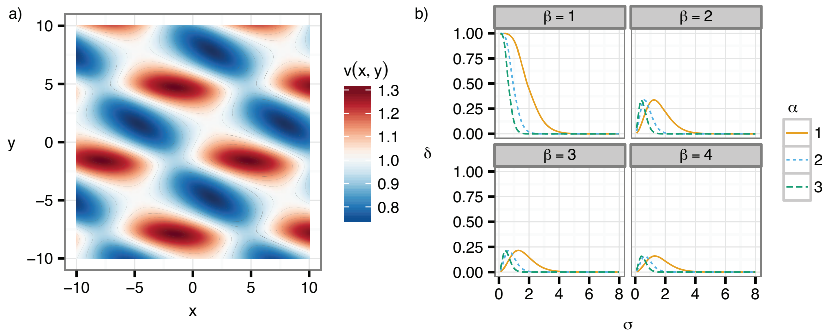

Fig. 3 Panel a): Numerical calculation of the function $v(x,y)$, which is the ratio of two-step density $p_{t-2\tau,t}(x|y,\sigma,\alpha,\beta)$ and one-step density $p_{t-\tau,t}(x|y,g_2(\sigma,\alpha,\beta))$, for the weighting function $w(x) = \beta + \sin(\alpha x)$. Parameter values are $\sigma = 1$, $\alpha = 1$, $\beta = 2$. The function $v(x,y)$ varies roughly between 0.72 and 1.31. Panel b): Numerical calculation of $\delta(\sigma) := \max_{x,y} |v(x,y;\sigma) - 1|$ for the weighting function $w(x) = \beta + \sin(\alpha x)$ for varying values of α and β. The parameter α, which determines the wavelength of the sine, shifts the curve $\delta(\sigma)$ and varies the skewing, while retaining the height of the maximum. The parameter β in contrast changes height of the maximum, which is the magnitude δ of the approximate robustness

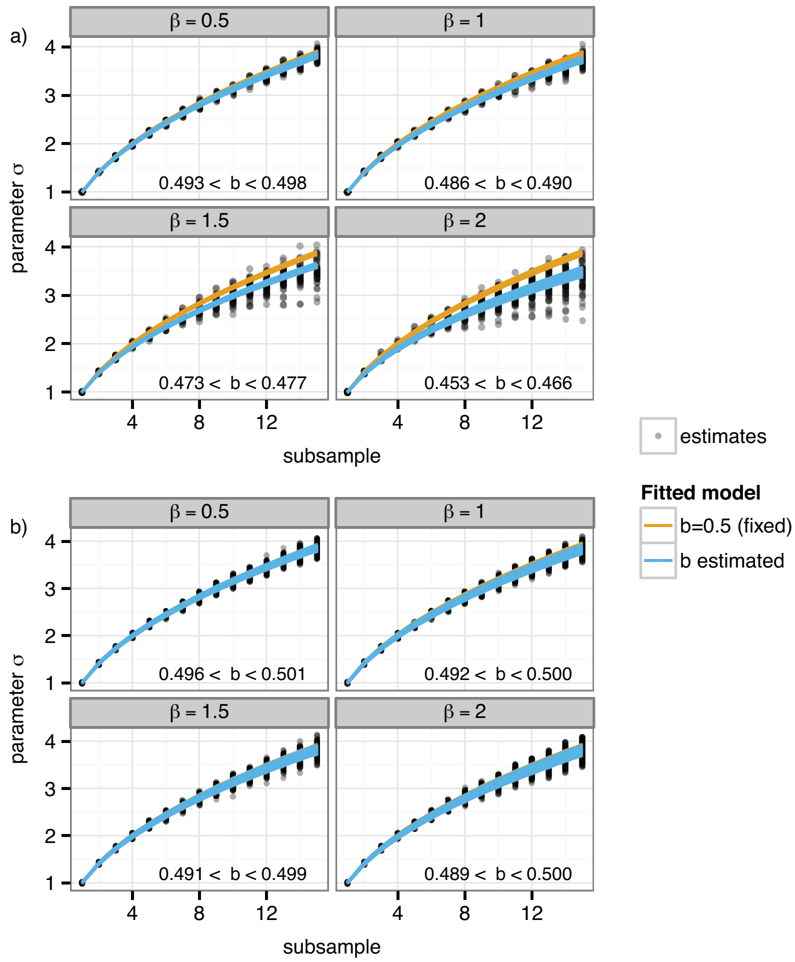

Fig. 4 Values of σ against increasing subsampling amount n. Estimates $\hat{\sigma}$ (gray dots) were fitted with a power-relationship, stratified by trajectories, and separately for several combinations of true parameter values (σ, β, and α for the model with logistic RSF). The power b was either fixed at 0.5 (ideal relationship; upper orange lines) or flexible and estimated (lower blue lines). The noted range of b refers to variation for different parameter combinations. Estimates and predictions are standardized by the corresponding true value. Panel a): Model with exponential RSF. With increasing value of β, estimates $\hat{\sigma}$ tended to increase less with subsampling compared to the ideal relationship. Panel b): Model with logistic RSF. The fitted power-relationship was very close to the ideal relationship, such that lines indicating the ideal relationship are overlaid by lines showing the fitted relationship

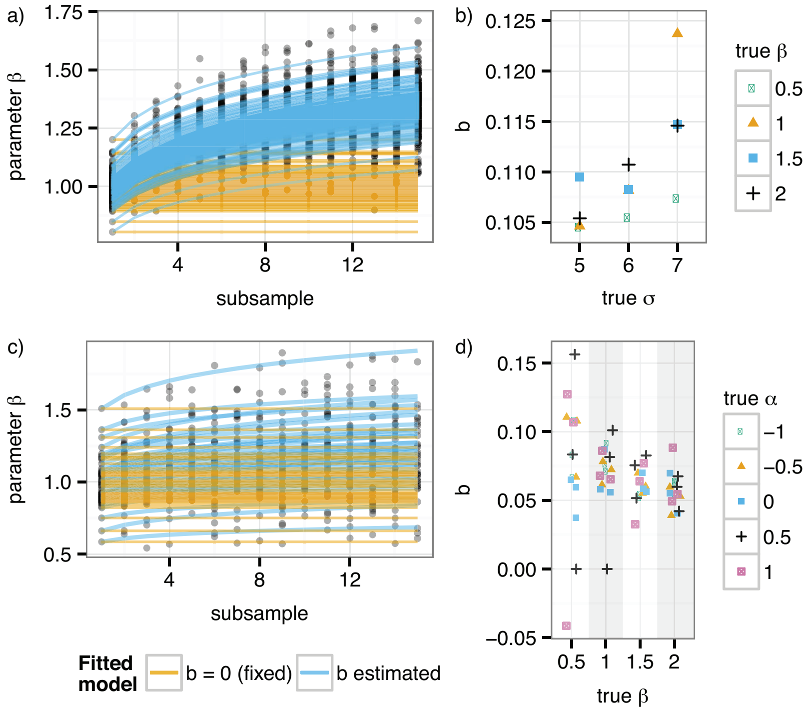

Fig. 5 Simulation results for the resource selection parameter β for the model with exponential RSF (panels a,b) and logistic RSF (panels c,d). Panels a) and c): Estimates $\hat{\beta}$ (gray dots) for increasing subsampling amount n were fitted with a power-relationship, stratified by trajectories, and separately for several combinations of true parameter values (σ, β, and α for the model with logistic RSF). The power b was either fixed at zero, representing the assumption that resource-selection parameters do not change with changing temporal resolution (ideal relationship; straight orange lines), or flexible and estimated (curved blue lines). Estimates and predictions are standardized by the corresponding true value. In panel c), only estimates and predictions for $\alpha = 0$, $\beta = 1$ are shown. Panel b): For the exponential RSF, the estimated power b was always above 0.1 and tended to increase with σ. Panel d): For the logistic RSF, the estimated power b was mainly below 0.1 and tended to decrease and concentrate more for increasing β

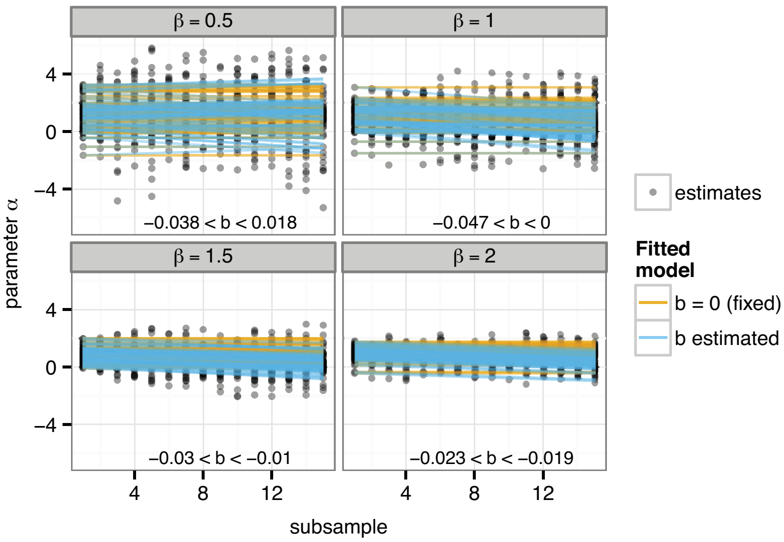

Fig. 6 For the model with logistic RSF, values of α against increasing subsampling amount n. Estimates were fitted with a linear relationship, stratified by trajectories, and separately for several combinations of true parameter values (σ, β, and α for the model with logistic RSF). The slope b was either fixed at zero, representing the assumption that resource-selection parameters do not change with changing temporal resolution (ideal relationship; straight orange lines), or flexible and estimated (blue lines). Estimates and predictions are standardized by the corresponding true value and only shown for $\alpha = 0.5$. The noted range of b refers to variation for different parameter combinations

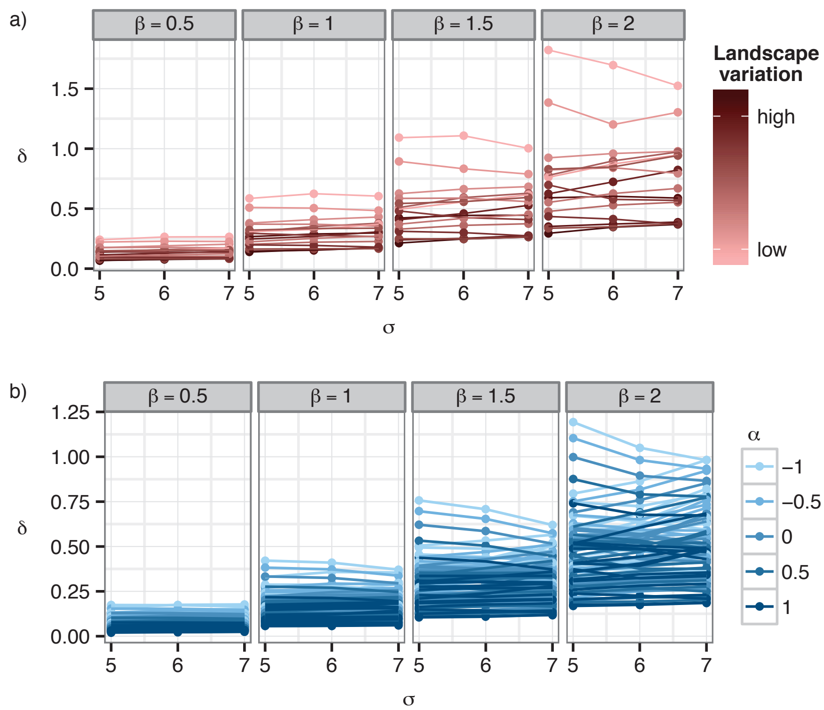

Fig. 7 Magnitudes of approximate robustness for the study case models with resource selection. The plots depict δ for varying values of σ and selection parameter β (dots). Lines join values for the same landscape i, $1 \leq i \leq 16$. Panel a): Magnitudes for the model exponential RSF. Values of δ tend to be lower for landscapes with less variation $\mathrm{Var}(r(x))$. Panel b): Magnitudes for the model with logistic RSF. Values of δ tend to be lower for higher values of the additional intercept parameter α

Acknowledgements We thank members of the Lewis research group for fruitful discussions and an anonymous reviewer for helpful comments on the manuscript. UES was supported by a scholarship from iCORE, now part of Alberta Innovates-Technology Futures and funding from the University of Alberta. MAL gratefully acknowledges Natural Sciences and Engineering Research Council Discovery and Accelerator grants, a Canada Research Chair and a Killam Research Fellowship.

Compliance with Ethical Standards Conflict of Interest: The authors declare that they have no conflict of interest.

References

Aarts G, Fieberg J, Matthiopoulos J (2011) Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol Evol 3(1):177–187
Benhamou S (2013) Of scales and stationarity in animal movements. Ecol Lett 17(3):261–272
Benhamou S, Sudre J, Bourjea J, Ciccione S, De Santis A, Luschi P (2011) The role of geomagnetic cues in green turtle open sea navigation. PLoS ONE 6(10):e26,672
Borger L, Dalziel BD, Fryxell JM (2008) Are there general mechanisms of animal home range behaviour? A review and prospects for future research. Ecol Lett 11(6):637–650
Boyce M, Vernier P, Nielsen S, Schmiegelow F (2002) Evaluating resource selection functions. Ecol Modell 157(2):281–300
Breed GA, Costa DP, Goebel ME, Robinson PW (2011) Electronic tracking tag programming is critical to data collection for behavioral time-series analysis. Ecosphere 2(1):art10
Codling EA, Hill NA (2005) Sampling rate effects on measurements of correlated and biased random walks. J Theor Biol 233(4):573
Colchero F, Conde DA, Manterola C, Chávez C, Rivera A, Ceballos G (2010) Jaguars on the move: modeling movement to mitigate fragmentation from road expansion in the Mayan Forest. Anim Conserv 14(2):158–166
Côrtes MC, Uriarte M (2013) Integrating frugivory and animal movement: a review of the evidence and implications for scaling seed dispersal. Biol Rev 88(2):255–272
Costa DP, Breed GA, Robinson PW (2012) New insights into pelagic migrations: implications for ecology and conservation. Annu Rev Ecol Evol Syst 43(1):73–96
Courbin N, Fortin D, Dussault C, Fargeot V, Courtois R (2013) Multi-trophic resource selection function enlightens the behavioural game between wolves and their prey. J Anim Ecol 82(5):1062–1071
Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF (2014) From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. Am Nat 183(5):E154–E167
Forester JD, Im H, Rathouz P (2009) Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. Ecology 90(12):3554–3565
Fortin D, Beyer HL, Boyce M, Smith D, Duchesne T, Mao J (2005) Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. Ecology 86(5):1320–1330
Frair JL, Fieberg J, Hebblewhite M, Cagnacci F, DeCesare NJ, Pedrotti LA (2010) Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. Philos Trans R Soc B 365(1550):2187–2200
Giuggioli L, Kenkre VM (2014) Consequences of animal interactions on their dynamics: emergence of home ranges and territoriality. Movement Ecology 2:2–20
Hampel FR (1971) A general qualitative definition of robustness. Ann Math Statist 42(6):1887–1896
Hampel FR (1986) Robust Statistics: The approach based on influence functions. Wiley, New York
Haran M (2011) Gaussian random field models for spatial data. In: Brooks S, Gelman A, Jones GL, Meng XL (eds) Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, pp 449–478
Hebblewhite M, Merrill E (2008) Modelling wildlife–human relationships for social species with mixed-effects resource selection models. J Appl Ecol 45(3):834–844
Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, N.J.
Ito TY, Lhagvasuren B, Tsunekawa A, Shinoda M, Takatsuki S, Buuveibaatar B, Chimeddorj B (2013) Fragmentation of the habitat of wild ungulates by anthropogenic barriers in Mongolia. PLoS ONE 8(2):e56,995

James A, Plank MJ, Edwards A (2011) Assessing Lévy walks as models of animal foraging. J R Soc Interface 8(62):1233–1247

Jerde CL, Visscher DR (2005) GPS measurement error influences on movement model parameterization. Ecol Appl 15(3):806–810

Johnson CJ, Parker KL, Heard DC, Gillingham MP (2002) Movement parameters of ungulates and scale-specific responses to the environment. J Anim Ecol 71(2):225–235

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, Morales JM (2013) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology 93(11):2336–2342

Latham ADM, Latham MC, Boyce M, Boutin S (2011) Movement responses by wolves to industrial linear features and their effect on woodland caribou in northeastern Alberta. Ecol Appl 21(8):2854–2865

Lele SR, Keim JL (2006) Weighted distributions and estimation of resource selection probability functions. Ecology 87(12):3021–3028

Lele SR, Merrill EH, Keim J, Boyce MS (2013) Selection, use, choice and occupancy: clarifying concepts in resource selection studies. J Anim Ecol 82(6):1183–1191

Manly BF, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals: statical design and analysis for field studies, 2nd edn. Kluwer Academic Publishers, Dordrecht

Masden EA, Reeve R, Desholm M, Fox AD, Furness RW, Haydon DT (2012) Assessing the impact of marine wind farms on birds through movement modelling. J R Soc Interface 9(74):2120–2130

McClintock BT, Johnson DS, Hooten MB, Ver Hoef JM, Morales JM (2014) When to be discrete: the importance of time formulation in understanding animal movement. Movement Ecology 2(1):334

Mills KJ, Patterson BR, Murray DL (2006) Effects of variable sampling frequencies on GPS transmitter efficiency and estimated wolf home range size and movement distance. Wildl Soc Bull 34(5):1463–1469

Moorcroft PR, Barnett A (2008) Mechanistic home range models and resource selection analysis: a reconciliation and unification. Ecology 89(4):1112–1119

Mueller T, Lenz J, Caprano T, Fiedler W, Böhning-Gaese K (2014) Large frugivorous birds facilitate functional connectivity of fragmented landscapes. J Appl Ecol 51(3):684–692

Pépin D, Adrados C, Mann C, Janeau G (2004) Assessing real daily distance traveled by ungulates using differential GPS locations. J Mammal 85(4):774–780

Postlethwaite CM, Dennis TE (2013) Effects of temporal resolution on an inferential model of animal movement. PLoS ONE 8(5):e57,640

Potts JR, Lewis MA (2014) How do animal territories form and change? Lessons from 20 years of mechanistic modelling. Proc R Soc B 281(1784):20140,231

Potts JR, Bastille-Rousseau G, Murray DL, Schaefer JA, Lewis MA (2014) Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. Methods Ecol Evol 5(3):253–262

Pyke GH (2015) Understanding movements of organisms: it's time to abandon the Lévy foraging hypothesis. Methods Ecol Evol 6(1):1–16

Rhodes JR, McAlpine CA, Lunney D, Possingham HP (2005) A spatially explicit habitat selection model incorporating home range behavior. Ecology 86(5):1199–1205

Robert CP, Casella G (2000) Monte Carlo statistical methods, corrected 2. print. edn. Springer texts in statistics, Springer, New York

Robinson WD, Bowlin MS, Bisson I, Shamoun-Baranes J, Thorup K, Diehl RH, Kunz TH, Mabey S, Winkler DW (2009) Integrating concepts and technologies to advance the study of bird migration. Frontiers in Ecology and the Environment 8(7):354–361

Rosser G, Fletcher AG, Maini PK, Baker RE (2013) The effect of sampling rate on observed statistics in a correlated random walk. J R Soc Interface 10(85):20130,273

Rowcliffe MJ, Carbone C, Kays R, Kranstauber B, Jansen PA (2012) Bias in estimating animal travel distance: the effect of sampling frequency. Methods Ecol Evol 3(4):653–662

Ryan PG, Petersen SL, Peters G, Gremillet D (2004) GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins. Mar Biol 145(2)

Sawyer H, Kauffman M, Nielson R, Horne J (2009) Identifying and prioritizing ungulate migration routes for landscape-level conservation. Ecol Appl 19(8):2016–2025

Schick RS, Loarie SR, Colchero F, Best BD, Boustany A, Conde DA, Halpin PN, Joppa LN, McClellan CM, Clark JS (2008) Understanding movement data and movement processes: current and emerging directions. Ecol Lett 11(12):1338–1350

Schlägel UE, Lewis MA (2016) A framework for analyzing the robustness of movement models to variable step discretization. J Math Biol pp 1–31, DOI 10.1007/s00285-016-0969-5

Schlather M, Menck P, Singleton R, Pfaff B, R Core team (2013) RandomFields: Simulation and Analysis of Random Fields

Smouse PE, Focardi S, Moorcroft PR, Kie JG, Forester JD, Morales JM (2010) Stochastic modelling of animal movement. Philos Trans R Soc B 365(1550):2201–2211

Squires JR, DeCesare NJ, Olson LE, Kolbe JA, Hebblewhite M, Parks SA (2013) Combining resource selection and movement behavior to predict corridors for Canada lynx at their southern range periphery. Biol Conserv 157(0):187–195

Tanferna A, López-Jiménez L, Blas J, Hiraldo F, Sergio F (2012) Different location sampling frequencies by satellite tags yield different estimates of migration performance: pooling data requires a common protocol. PLoS ONE 7(11):e49,659

Tsoar A, Nathan R, Bartan Y, Vyssotski A, Dell'Omo G, Ulanovsky N (2011) Large-scale navigational map in a mammal. Proc Natl Acad Sci USA 108(37):E718–E724

Turchin P (1998) Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. Sinauer Associates, Sunderland, Mass.

Wiens DP (2000) Bias constrained minimax robust designs for misspecified regression models. In: Balakrishnan N, Melas VB, Ermakov S (eds) Statistics for Industry and Technology, Birkhäuser, Boston, MA, pp 117–133

Wiens DP, Zhou J (1996) Minimax regression designs for approximately linear models with autocorrelated errors. J Statist Plann Inference 55(1):95–106

Wilcox R (2012) Introduction to robust estimation and hypothesis testing, 3rd edn. Academic Press, Boston

A Proofs of results about exact robustness

Proof (Theorem 1) First, note that for any standard deviation of the kernel, σ, the integral $\int_{\mathbb{R}} k_\sigma(y;x)w(y)\,\mathrm{d}y$ reduces to the weighting function evaluated at the kernel's mean,

$$\begin{aligned}\int_{\mathbb{R}} k_\sigma(y;x)w(y)\,\mathrm{d}y &= \int_{\mathbb{R}} k_\sigma(y;x)(ay+b)\,\mathrm{d}y = \int_{\mathbb{R}} k_\sigma(y;x)(a(y-x+x)+b)\,\mathrm{d}y \\ &= (ax+b)\int_{\mathbb{R}} k_\sigma(y;x)\,\mathrm{d}y + a\int_{\mathbb{R}} k_\sigma(y;x)(y-x)\,\mathrm{d}y = ax+b = w(x), \qquad (24)\end{aligned}$$

because $k_\sigma(\cdot|y)$ is a Gaussian density integrating to 1 and with vanishing first central moment. If we consider w as a linear transformation of a Normally distributed random variable with mean x, then equation (24) reflects a special case of Jensen's inequality, in which equality holds.

We now show robustness of degree n with parameter transformation $g_n(\sigma,a,b) = (\sqrt{n}\sigma,a,b)$ by induction. For $n=1$, we have the trivial transformation $g_1(\sigma,a,b) = (\sigma,a,b)$, and there is nothing to show for robustness of degree 1.

We assume that robustness or degree n holds, that is we have the relationship

$$p_n(x_n|x_0,\sigma,a,b) = p_1(x_n|x_0,\sqrt{n}\sigma,a,b). \qquad (25)$$

for all $x_n, x_0 \in \mathbb{R}$. For $n+1$, we use the Chapman-Kolmogorov equation and Markov property and obtain

$$\begin{aligned}p_{n+1}(x_{n+1}|x_0,\sigma,a,b) &= \int_{\mathbb{R}^n} \prod_{k=1}^{n+1} p_1(x_k|x_{k-1},\sigma,a,b)\,\mathrm{d}x_1\ldots\mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b)\left(\int_{\mathbb{R}^{n-1}} \prod_{k=1}^{n} p_1(x_k|x_{k-1},\sigma,a,b)\,\mathrm{d}x_1\ldots\mathrm{d}x_{n-1}\right)\mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b)\,p_n(x_n|x_0,\sigma,a,b)\,\mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b)\,p_1(x_n|x_0,\sqrt{n}\sigma,a,b)\,\mathrm{d}x_n, \qquad (26)\end{aligned}$$

where the last step follows by induction. We can now insert the model's step probabilities and use equation (24) to further calculate,

$$\begin{aligned} p_{n+1}(x_{n+1}|x_0,\sigma,a,b) &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\,w(x_{n+1})}{\int_{\mathbb{R}} k_\sigma(y;x_n)w(y)\,dy}\;\frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\,w(x_n)}{\int_{\mathbb{R}} k_{\sqrt{n}\,\sigma}(y;x_0)w(y)\,dy}\,\mathrm{d}x_n \\ &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\,w(x_{n+1})}{w(x_n)}\;\frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\,w(x_n)}{w(x_0)}\,\mathrm{d}x_n \\ &= \frac{w(x_{n+1})}{w(x_0)}\int_{\mathbb{R}} k_\sigma(x_{n+1};x_n)\,k_{\sqrt{n}\sigma}(x_n;x_0)\,\mathrm{d}z. \end{aligned} \tag{27}$$

Note that we have assumed that all movement steps are within the domain $\mathscr{I}$, where the weighting function is positive. Since $k_\sigma(x_{n+1};x_n) = k_\sigma(x_{n+1}-x_n;0)$, the integral in the last expression is the convolution of two Gaussian densities with variances σ^2 and $n\sigma^2$ and with means 0 and x_0, respectively. Because of the linearity of Gaussian random variables, this is again a Gaussian density with mean x_0 and variance $(n+1)\sigma^2$. Because equation (24) holds for the kernel with any standard deviation, we can rewrite the denominator as $w(x_0) = \int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y;x_0)w(y)\,dy$. Thus,

$$p_{n+1}(x_{n+1}|x_0,\sigma,a,b) = \frac{k_{\sqrt{n+1}\,\sigma}(x_{n+1};x_0)w(x_{n+1})}{\int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y;x_0)w(y)\,dy} = p_1(x_{n+1}|x_0,\sqrt{n+1}\,\sigma,a,b). \tag{28}$$

□

Proof (Theorem 2) We proceed analogously to the previous proof. The integral of weighting function and kernel with arbitrary standard deviation σ and mean x is here given by

$$\begin{aligned} \int_{\mathbb{R}} k_\sigma(y;x)\,w(y)\,\mathrm{d}y &= \int_{\mathbb{R}} k_\sigma(y;x)\,Ce^{ay+b}\,\mathrm{d}y \\ &= \frac{C}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}+ay+b\right)\mathrm{d}y. \end{aligned}$$

By completing the square and using substitution $u = \frac{1}{\sqrt{2}\sigma}(y-x-a\sigma^2)$ we obtain

$$\begin{aligned} \int_{\mathbb{R}} k_\sigma(y;x)\,w(y)\,\mathrm{d}y &= \frac{C}{\sqrt{2\pi}\sigma}e^{\frac{a^2\sigma^2}{2}+ax+b}\int_{\mathbb{R}} \exp\left(-\left(\frac{y-x-a\sigma^2}{\sqrt{2}\sigma}\right)^2\right)\mathrm{d}y \\ &= \frac{C}{\sqrt{2\pi}\sigma}e^{\frac{a^2\sigma^2}{2}+ax+b}\int_{\mathbb{R}} \exp\left(-u^2\right)\sqrt{2}\sigma\,\mathrm{d}u. \end{aligned}$$

The final integral reduces to $\sqrt{2\pi}\sigma$, and therefore,

$$\int_{\mathbb{R}} k_\sigma(y;x)\,w(y)\,\mathrm{d}y = C\,e^{\frac{a^2\sigma^2}{2}+ax+b}. \tag{29}$$

Again, we prove robustness of degree n by induction, using parameter transformation $g_n(\sigma, C, a, b) = (\sqrt{n}\sigma, C, a, b)$. In the induction step, we obtain, with help of equation (29),

$$
\begin{aligned}
p_{n+1}(x_{n+1}|x_0, \sigma, a, b) &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1}; x_n)\, Ce^{ax_{n+1}+b}}{\int_{\mathbb{R}} k_\sigma(y; x_n) Ce^{ay+b}\, dy} \frac{k_{\sqrt{n}\sigma}(x_n; x_0)\, Ce^{ax_n+b}}{\int_{\mathbb{R}} k_{\sqrt{n}\sigma}(y; x_0) Ce^{ay+b}\, dy} \,\mathrm{d}x_n \\
&= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1}; x_n)\, Ce^{ax_{n+1}+b}}{Ce^{\frac{a^2\sigma^2}{2}+ax_n+b}} \frac{k_{\sqrt{n}\sigma}(x_n; x_0)\, Ce^{ax_n+b}}{Ce^{\frac{na^2\sigma^2}{2}+ax_0+b}} \,\mathrm{d}x_n \\
&= \frac{e^{x_{n+1}}}{e^{\frac{(n+1)a^2\sigma^2}{2}+ax_0}} \int_{\mathbb{R}} k_\sigma(x_{n+1}; x_n)\, k_{\sqrt{n}\sigma}(x_n; x_0) \,\mathrm{d}z \\
&= \frac{e^{x_{n+1}}}{e^{\frac{(n+1)a^2\sigma^2}{2}+ax_0}} k_{\sqrt{n+1}\,\sigma}(x_{n+1}; x_0). \\
&= \frac{k_{\sqrt{n+1}\,\sigma}(x_{n+1}; x_0)\, Ce^{ax_{n+1}+b}}{\int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y; x_0)\, Ce^{ay+b}\,\mathrm{d}y} \\
&= p_1(x_{n+1}|x_0, \sqrt{n+1}\,\sigma, a, b)
\end{aligned}
\tag{30}
$$

□

B Proof of result about asymptotic robustness

To highlight the main steps necessary to prove Theorem 3, we establish a series of intermediate results. As a first step, we show that the 2-step transition density can be broken up into a product of the form (5) in Definition 2.

Proposition 1 *The 2-step transition density of model with transitions* (14) *can be written as*

$$p_2(x_t|x_{t-2\tau}, \sigma, \boldsymbol{\theta}) = p_1(x_t|x_{t-2\tau}, \sqrt{2}\sigma, \boldsymbol{\theta}) \cdot v(x_t, x_{t-2\tau}; \tau), \tag{31}$$

where the function v is given by

$$v(x_t, x_{t-2\tau}; \tau) = \frac{\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y; x) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}{\int_{\mathbb{R}} k_\sigma(y; x) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z; \frac{1}{2}(x_t + x_{t-2\tau})\right) \frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \,\mathrm{d}z. \tag{32}$$

Note that v depends on τ through σ. For later convenience, we define

$$Q(x; \tau) := \frac{\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y; x) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}{\int_{\mathbb{R}} k_\sigma(y; x) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \tag{33}$$

$$I(x_1, x_2; \tau) := \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z; \frac{1}{2}(x_1 + x_2)\right) \frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \,\mathrm{d}z. \tag{34}$$

Proof The proposition can be shown with a straightforward calculation. The 2-step transition density is given by

$$p_2(x_t|x_{t-2\tau}, \sigma, \boldsymbol{\theta}) \tag{35}$$

$$= \int_{\mathbb{R}} \frac{k_\sigma(x_t; z)\, w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \frac{k_\sigma(z; x_{t-2\tau})\, w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; x_{t-2\tau}) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \,\mathrm{d}z \tag{36}$$

$$= \frac{w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y; x_{t-2\tau}) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \int_{\mathbb{R}} k_\sigma(x_t; z)\, k_\sigma(z; x_{t-2\tau}) \frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \,\mathrm{d}z. \tag{37}$$

Tthe product of the two Gaussian densities in the integrand can be transformed as follows

$$k_\sigma(x_t;z)\,k_\sigma(z;x_{t-2\tau}) = k_{\sqrt{2}\sigma}(x_t;x_{t-2\tau})\,k_{\frac{\sigma}{\sqrt{2}}}\Big(z;\frac{1}{2}(x_t+x_{t-2\tau})\Big). \tag{38}$$

The two-step density therefore becomes

$$p_2(x_t|x_{t-2\tau},\sigma,\boldsymbol{\theta}) = \frac{k_{\sqrt{2}\sigma}(x_t;x_{t-2\tau})\,w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y;x_{t-2\tau})w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z;\frac{1}{2}(x_t+x_{t-2\tau})\Big)\,\frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y;z)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z. \tag{39}$$

The numerator of the first factor is the desired one-step density up to appropriate normalization. If we extend by the required normalization constant $\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y;x_{t-2\tau})w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y$, we obtain equations (31) and (32). □

We are now left to show that the function $v-1$ is in the order of τ on its entire domain $\mathbb{R}^2\times\mathbb{R}^+$. In particular, this means that for any fixed τ^*, the function $v(x_1,x_2;\tau^*)-1$ is bounded on $\mathbb{R}^2$ via $c\tau^*$ for a constant c. It turns out to be helpful to analyze v separately on $\mathbb{R}^2\times(0,\tau_0)$ and $\mathbb{R}^2\times[\tau_0,\infty)$ for some τ_0. Because the proof is simpler for large τ, we present this result first.

Lemma 1 *Let w be continuous and bounded away from zero, that is there exist L and U such that* $0<L\le w_{\boldsymbol{\theta}}(x)\le U$ *for all* $x\in\mathbb{R}$. *Let w further be twice differentiable on* $\mathbb{R}$ *with* $|w''(x)|<M$ *for some M and all* $x\in\mathbb{R}$. *For any* $\tau_0>0$, *we have* $v(x_1,x_2,;\tau)-1=\mathscr{O}(\tau)$ *on* $\mathbb{R}^2\times[\tau_0,\infty)$.

Proof Let τ_0 be a number away from zero and fixed. Our goal is to establish bounds on the functions Q and I, as defined in (33) and (34), and to use these to place a bound on $v-1$. Because w is twice differentiable we can apply Taylor's theorem to obtain a linear approximation for w using any point $x\in\mathbb{R}$,

$$w_{\boldsymbol{\theta}}(y) = w_{\boldsymbol{\theta}}(x)+w'(x)(y-x)+R(y), \tag{40}$$

where $R(y)$ is the remainder term. This leads to

$$\int_{\mathbb{R}} k_\sigma(y;x)\,w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y = w_{\boldsymbol{\theta}}(x)\int_{\mathbb{R}} k_\sigma(y;x)\,\mathrm{d}y + w'(x)\int_{\mathbb{R}} k_\sigma(y;x)\,(y-x)\,\mathrm{d}y + \int_{\mathbb{R}} k_\sigma(y;x)\,R(y)\,\mathrm{d}y, \tag{41}$$

where the first term on the RHS becomes $w_{\boldsymbol{\theta}}(x)$, because the kernel integrates to 1, and the integral in the second term is the first central moment of the kernel, hence vanishes. The remainder $R(y)$, using the Lagrange form, is given by $R(y)=\frac{w''(\xi)}{2}(y-x)^2$, for some ξ between x_2 and y. Since the second derivative of w is assumed to be globally bounded, we have $|R(y)|\le\frac{M}{2}(y-x)^2$. We use this to place bounds on the third term, recognizing that the remaining integral $\int_{\mathbb{R}} k_\sigma(y;x)\,(y-x)^2\,\mathrm{d}y$ is the second central moment of the Gaussian kernel k_σ, which is given by its variance $\sigma^2=\omega^2\tau$. Therefore,

$$w_{\boldsymbol{\theta}}(x)-\frac{M}{2}\omega^2\tau \le \int_{\mathbb{R}} k_\sigma(y;x)\,w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y \le w_{\boldsymbol{\theta}}(x)+\frac{M}{2}\omega^2\tau. \tag{42}$$

In general, the lower bound can be arbitrarily close to zero, therefore we cannot simply invert this inequality to obtain an estimate on the inverse of the integral. Instead, we use the bounds on w and again the fact $\int_{\mathbb{R}} k_\sigma(y;x)\,\mathrm{d}y=1$ for any σ and any $x\in\mathbb{R}$ to establish

$$0<L\le\int_{\mathbb{R}} k_\sigma(y;x)\,w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y\le U, \tag{43}$$

which can be inverted. Since inequalities (42) and (43) hold for any σ and any $x\in\mathbb{R}$, they allow us to place bounds on both Q and I. For Q, we obtain

$$\frac{1}{U}\big(w_{\boldsymbol{\theta}}(x)-M\omega^2\tau\big)\le Q(x;\tau)\le\frac{1}{L}\big(w_{\boldsymbol{\theta}}(x)+M\omega^2\tau\big) \tag{44}$$

for all $x\in\mathbb{R}$, $\tau\in\mathbb{R}^+$. We can avoid the dependency of the bounds on x by again invoking the bounds on w,

$$\frac{1}{U}\big(L-M\omega^2\tau\big)\le Q(x)\le\frac{1}{L}\big(U+M\omega^2\tau\big). \tag{45}$$

For the function I, we only make use of the bounds on w and inequality (43) and get

$$0 < \frac{L}{U} \leq I(x_1,x_2;\tau) \leq \frac{U}{L} \tag{46}$$

for all $x_1,x_2 \in \mathbb{R}$, $\tau \in \mathbb{R}^+$. We can now continue to calculate $v-1$. An upper bound is immediately given by

$$v(x_1,x_2;\tau) - 1 = Q(x_1;\tau)\, I(x_1,x_2;\tau) - 1 \leq \frac{U^2-L^2}{L^2} + \frac{MU}{L^2}\,\omega^2\tau. \tag{47}$$

With only few more additional steps, we obtain a lower bound by simply drawing upon $L \leq U$, its squared version and its inverse,

$$-\big(v(x_1,x_2;\tau)-1\big) \leq \frac{U^2-L^2}{U^2} + \frac{ML}{U^2}\,\omega^2\tau \leq \frac{U^2-L^2}{L^2} + \frac{MU}{L^2}\,\omega^2\tau. \tag{48}$$

Define $C := \frac{U^2-L^2}{L^2\tau_0} + \frac{MU}{L^2}\omega^2$ for the τ_0 chosen up front. Then,

$$\begin{aligned}
|v(x_1,x_2;\tau)-1| &\leq \frac{U^2-L^2}{L^2} + \frac{MU}{L^2}\,\omega^2\tau - C\tau + C\tau &(49)\\
&= \frac{U^2-L^2}{L^2} - \frac{U^2-L^2}{L^2\tau_0}\,\tau + C\tau &(50)\\
&= \left(1-\frac{\tau}{\tau_0}\right)\frac{U^2-L^2}{L^2} + C\tau. &(51)
\end{aligned}$$

The product on the RHS is non-positive for $\tau \geq \tau_0$, and hence $|v(x_1,x_2;\tau)-1| \leq C\tau$ for all $\mathbb{R}^2 \times [\tau_0,\infty)$. □

The bounds on Q and I, and thus $v-1$, established in the preceding proof are not sufficient to conclude the result as $\tau \to 0$, unless $L = U$, which is the trivial case of a constant weighting function. More suitable bounds, however, can be found if inequality (42) can be inverted. This can be achieved by assuming τ to be small enough.

Lemma 2 *Let w be continuous and bounded away from zero, that is there exist L and U such that $0 < L \leq w_{\boldsymbol\theta}(x) \leq U$ for all $x \in \mathbb{R}$. Let w further be twice differentiable on $\mathbb{R}$ with $|w''(x)| < M$ for some M and all $x \in \mathbb{R}$. Let $\tau_0 = \frac{2L}{M\omega^2}$. Then $v(x_1,x_2,;\tau) - 1 = \mathscr{O}(\tau)$ on $\mathbb{R}^2 \times (0,\tau_0)$.*

Proof Here we develop bounds on Q and I such that both $Q-1$ and $I-1$ are in the order of τ. Let $\tau \leq \tau_0$ for τ_0 as defined in the lemma. Then the lower bound of equation (42) is bounded away from zero,

$$w_{\boldsymbol\theta}(x) - \frac{M}{2}\,\omega^2\tau \geq w_{\boldsymbol\theta}(x) - \frac{M}{2}\,\omega^2\tau_0 > w_{\boldsymbol\theta}(x) - \frac{M}{2}\,\omega^2\frac{2L}{M\omega^2} = w_{\boldsymbol\theta}(x) - L \geq 0. \tag{52}$$

Hence we can invert the inequality (42) and obtain

$$\frac{w_{\boldsymbol\theta}(x) - M\omega^2\tau}{w_{\boldsymbol\theta}(x) + \frac{M}{2}\omega^2\tau} \leq Q(x;\tau) \leq \frac{w_{\boldsymbol\theta}(x) + M\omega^2\tau}{w_{\boldsymbol\theta}(x) - \frac{M}{2}\omega^2\tau}. \tag{53}$$

Note that the values in the numerators and denominators differ slightly because the variances of the kernel k in the numerator and denominator of Q differ by a factor of 2.
Since $2w_{\boldsymbol\theta}(x) - M\omega^2\tau \geq 2L - M\omega^2\tau_0 > 0$, we can conclude

$$Q(x;\tau) - 1 \leq \frac{w_{\boldsymbol\theta}(x) + M\omega^2\tau - w_{\boldsymbol\theta}(x) - \frac{M}{2}\omega^2\tau}{w_{\boldsymbol\theta}(x) - \frac{M}{2}\omega^2\tau} = \frac{M\omega^2\tau}{2w_{\boldsymbol\theta}(x) - M\omega^2\tau} \leq \frac{M\omega^2\tau}{2L - M\omega^2\tau_0}, \tag{54}$$

for all $x \in \mathbb{R}$ and $\tau < \tau_0$. Using $2w_{\boldsymbol\theta}(x) + M\omega^2\tau \geq 2w_{\boldsymbol\theta}(x) \geq 2L$, we similarly obtain,

$$-(Q(x;\tau)-1) \leq \frac{3M\omega^2\tau}{2w_{\boldsymbol\theta}(x) + M\omega^2\tau} \leq \frac{3M}{2L}\,\omega^2\tau \tag{55}$$

for all $x \in \mathbb{R}$ and $\tau < \tau_0$. If we set $C_1 := \max\left(\frac{M\omega^2}{2L-2\omega^2\tau_0}, \frac{3M\omega^2}{2L}\right)$, it follows that $|Q(x;\tau)-1| \leq C_1\tau$ on $\mathbb{R}^2 \times (0,\tau_0)$.

Using analogous arguments as before, we can fine an find an upper bound on I,

$$I(x_1,x_2;\tau) = \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y;z)w_{\boldsymbol{\theta}}(y)\,dy}\,\mathrm{d}z \tag{56}$$

$$\leq \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{w_{\boldsymbol{\theta}}(z)}{w_{\boldsymbol{\theta}}(z)-\frac{M}{2}\omega^2\tau}\,\mathrm{d}z \tag{57}$$

$$= \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{w_{\boldsymbol{\theta}}(z)-\frac{M}{2}\omega^2\tau+\frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z)-\frac{M}{2}\omega^2\tau}\,\mathrm{d}z \tag{58}$$

$$= \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\mathrm{d}z + \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{\frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z)-\frac{M}{2}\omega^2\tau}\,\mathrm{d}z \tag{59}$$

$$\leq 1 + \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{\frac{M}{2}\omega^2\tau}{L-\frac{M}{2}\omega^2\tau_0}\,\mathrm{d}z = 1 + \frac{M\omega^2\tau}{2L-M\omega^2\tau_0}. \tag{60}$$

A lower bound is given by

$$I(x_1,x_2;\tau) \geq \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{w_{\boldsymbol{\theta}}(z)}{w_{\boldsymbol{\theta}}(z)+\frac{M}{2}\omega^2\tau}\,\mathrm{d}z \tag{61}$$

$$= 1 - \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\left(z;\frac{1}{2}(x_1+x_2)\right)\frac{\frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z)+\frac{M}{2}\omega^2\tau}\,\mathrm{d}z \geq 1 - \frac{M\omega^2\tau}{2L}. \tag{62}$$

Setting $C_2 := \frac{M\omega^2\tau}{2L-M\omega^2\tau_0}$, we obtain $|I(x_1,x_2;\tau)-1| \leq C_2\tau$ on $\mathbb{R}^2 \times (0,\tau_0)$.

We can now estimate $v-1$ as follows,

$$|v(x_1,x_2;\tau)-1| = |Q_\tau I_\tau - 1| \leq |Q_\tau - 1|\,|I_\tau - 1| + |Q_\tau - 1| + |I_\tau - 1| \tag{63}$$

$$\leq C_1 C_2 \tau^2 + (C_1+C_2)\,\tau \leq \left(C_1 C_2 \tau_0 + C_1 + C_2\right)\tau, \tag{64}$$

for all $x_1, x_2 \in \mathbb{R}$ and all $\tau < \tau_0$. □

Lemmata 1 and 2, together with proposition 1 prove Theorem 3.