Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2014, **5,** 1012–1022 doi: 10.1111/2041-210X.12253

A generalized residual technique for analysing complex movement models using earth mover's distance

Jonathan R. Potts[1,2]*, Marie Auger-Méthé[1,3], Karl Mokross[4,5] and Mark A. Lewis[1,3]

[1]*Centre for Mathematical Biology, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada;* [2]*School of Mathematics and Statistics, University of Sheffield, Sheffield, UK;* [3]*Department of Biological Sciences, University of Alberta, Edmonton, AB,Canada;* [4]*School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA; and* [5]*Projeto Dinâmica Biológica de Fragmentos Florestais, INPA, Av. André Araújo 2936, Petrópolis, Manaus 69083-000, Brazil*

Summary

1. Complex systems of moving and interacting objects are ubiquitous in the natural and social sciences. Predicting their behaviour often requires models that mimic these systems with sufficient accuracy, while accounting for their inherent stochasticity. Although tools exist to determine which of a set of candidate models is best relative to the others, there is currently no generic goodness-of-fit framework for testing how close the best model is to the real complex stochastic system.
2. We propose such a framework, using a novel application of the *Earth mover's distance*, also known as the *Wasserstein metric*. It is applicable to any stochastic process where the probability of the model's state at time t is a function of the state at previous times. It generalizes the concept of a residual, often used to analyse 1D summary statistics, to situations where the complexity of the underlying model's probability distribution makes standard residual analysis too imprecise for practical use.
3. We give a scheme for testing the hypothesis that a model is an accurate description of a data set. We demonstrate the tractability and usefulness of our approach by application to animal movement models in complex, heterogeneous environments. We detail methods for visualizing results and extracting a variety of information on a given model's quality, such as whether there is any inherent bias in the model, or in which situations, it is most accurate. We demonstrate our techniques by application to data on multispecies flocks of insectivore birds in the Amazon rain forest.
4. This work provides a usable toolkit to assess the quality of generic movement models of complex systems, in an absolute rather than a relative sense.

Key-words: spatial or time-series, statistics, modelling, population ecology

Introduction

How good is a model at describing reality? This fundamental question, ubiquitous across the quantitative sciences, has troubled and intrigued scientists for over 200 years (Legendre 1805; Gauss 1809). A variety of techniques have been discovered to address the problem in certain situations. Residual analysis is one example that has a long history of useful application in various areas (Zuur *et al.* 2009; Gordon 2010). However, it is only usable when the underlying model, or a summary statistic arising from the model, can be framed as a simple deterministic function.

Despite this, our world is infused with complex, multidimensional, stochastic systems. These range from biological systems, such as ant colonies, bird flocks and slime mould aggregation (Camazine *et al.* 2003), to crowd movement psychology in social sciences (Helbing, Johansson & Al-Abideen 2007), to protein dynamics (Berendsen & Hayward 2000). Such systems are typically high dimensional and can rarely be described in an accurate way without taking into account underlying randomness in movements of constituent objects. The aim of this paper was to generalize the technique of residual analysis so that it can be used for generic stochastic systems of moving and interacting objects.

The type of models that are analysable by residual analysis can be characterized as *deterministic* models. These are models of the form $\mathbf{a} = f(\mathbf{b})$, where $\mathbf{a}$ is the prediction, $\mathbf{b}$ is a vector of independent input variables and f is a deterministic function. The *residual*, $\mathbf{a}_{\text{obs}} - f(\mathbf{b})$, where $\mathbf{a}_{\text{obs}}$ is an observation, measures the closeness of the model to the data. Residual analysis is well developed and often used for assessing the quality of models arising from techniques such as regression (Zuur *et al.* 2009; Gordon 2010). However, when the function f is replaced by a probability distribution, $P(\mathbf{b})$, residuals are no longer well defined. If the distribution is sufficiently close to a Gaussian, such as if $P(\mathbf{b}) = f(\mathbf{b}) + \xi$, where ξ is a zero-mean noise term and f is deterministic, one can simply define the residual to be the distance between the data point $\mathbf{a}_{\text{obs}}$ and the mean of $P(\mathbf{b})$.

*Correspondence author. E-mail: j.potts@sheffield.ac.uk

However, this fails to be reasonable if the distribution is more complex, for example multimodal or long-tailed.

Typical stochastic movement-and-interaction models often depend on heterogeneous properties of either the environment (Forester, Im & Rathouz 2009; Van Moorter *et al*. 2009; Potts *et al*. 2014a) or surrounding agents (Camazine *et al*. 2003), frequently making the probability distribution of state transitions complex and multipeaked. While methods exist for selecting the *relative* quality between competing models of these complex systems, such as Likelihood Ratio (Potts *et al*. 2014a), Akaike Information Criteria (AIC), Deviance Information Criteria (Morales *et al*. 2004) or Bayesian methods (Jonsen, Flemming & Myers 2005), the current suite of goodness-of-fit tests fails to provide sufficient techniques for assessing the *absolute* quality of such a model, that is its closeness to the data. This has led to researchers either ignoring the question and solely performing model selection (Moorcroft, Lewis & Crabtree 2006), or performing *ad hoc* tests on 1D summary statistics (Grimm *et al*. 2005; Gautestad, Loe & Mysterud 2013). For example, a search for the 20 highest cited papers that fit animal movement models to data reveals that *none* test the absolute fit of the best model to the data (methods in Appendix S3).

In the animal movement literature in particular, this tendency to ignore the absolute quality of a model has been partially responsible for various controversies regarding the detection of underlying movement processes (Auger-Méthé *et al*. 2011). This has led to criticism of many papers for appearing to draw strong conclusions about animal behaviour by selecting the best of a small number of simple models, all of which may be very poor at reflecting data. For example, the results of Viswanathan *et al*. (1996) were later overturned by Edwards *et al*. (2007), and de Jager *et al*. (2011) were criticized by Jansen, Mashanova & Petrovskii (2012) for drawing possibly incorrect conclusions by only examining very simplistic models.

Recent work (Auger-Méthé *et al*. 2011, 2014) demonstrates that these issues may sometimes be resolved by examining the residuals of the respective models' step length and turning angle distributions. However, this technique is only applicable to a specific set of models, which have relatively simple distributions, and cannot easily incorporate the effects of heterogeneous surroundings on movements. Increasingly, it is proving necessary to factor such effects into movement models. Recent developments in both the step selection literature (Fortin *et al*. 2005; Rhodes *et al*. 2005; Forester, Im & Rathouz 2009; Latombe, Fortin & Parrott 2013; Vanaka *et al*. 2013; Potts *et al*. 2014a) and collective behavioural studies (Deneubourg *et al*. 1989; Couzin *et al*. 2002; Hoare *et al*. 2004; Guttal & Couzin 2010) amply demonstrate the importance of incorporating often heterogeneous surroundings into the understanding and modelling of animal movement. It is therefore necessary to construct tools similar to residual analysis, yet applicable to these more complex models, to avoid repeating the sort of problems that have already plagued the field of movement ecology regarding simpler models, often caused by choosing between a limited set of possibly poor models (Plank & Codling 2009; de Jager *et al*. 2011; Gautestad, Loe & Mysterud 2013; Auger-Méthé *et al*. 2014).

With these issues in mind, we construct a generalized residual method that can be applied to any stochastic system of moving and interacting objects. The particular types of models that we are concerned with are one-step Markovian, describing the probability $P_\tau(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$ of a system being in state $\mathbf{X}_{t+\tau}$ at time $t+\tau$ having been in state $\mathbf{X}_t$ at time t, as this eases notation and explanation. However, generalizing to non-Markovian situations merely involves rewriting the probability function so that it is dependent on several previous steps rather than just one. The states $\mathbf{X}_{t+\tau}$ and $\mathbf{X}_t$ could consist of a variety of information about the system, for example the positions of the agents, their directions, environmental information perceived by the agents and so forth; whatever is appropriate for the scientific questions being addressed.

Large classes of complex systems models in the movement literature can be described in this way, as so-called *coupled step selection functions* (Potts, Mokross & Lewis 2014b). These include models of collective behaviour, which are often applied to both human systems and inter- and intracellular systems (Berendsen & Hayward 2000; Camazine *et al*. 2003; Helbing *et al*. 2005). Therefore, our technique fills a gap in the increasingly important field of complex systems science, important for ecological applications and beyond.

EARTH MOVER'S DISTANCE: A MEASURE OF ABSOLUTE FIT

Suppose that a complex system is described in 'reality' by a function $P^R(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$ but that the best model of the system is given by $P^M(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$. In other words, if the system is currently in state $\mathbf{X}_t$, then the probability distribution function of it being at state $\mathbf{X}_{t+\tau}$ after a time of τ has elapsed is $P^R(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$. However, the best model constructed so far predicts that the system will have probability distribution function $P^M(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$. To assess how well this model reflects reality requires a measure of the distance between the two probability functions P^R and P^M. (Note that P^R and P^M depend upon the time interval τ between successive states of the system, but τ remains fixed throughout the paper so we do not include it in the notation.)

Mathematicians have developed such a distance function, called the *Wasserstein metric* (Vasershtein 1969), a special case of which has recently reemerged in the visual biometrics literature as the *earth mover's distance* (Rubner, Tomasi & Guibas 2000). Although the general measure-theoretic definition is rather formal and technical (Appendix S2), the distance has an intuitive explanation. Imagine that one of the probability distributions describes a pile of earth (e.g. sand, soil, etc.) that you have in front of you, and the other describes the shape of a pile of earth that you want to construct. Intuitively, the earth mover's distance is the minimum average distance that each particle of earth has to move to change the pile from what you have to what you want (Fig. 1a).

Although simple to state, this distance can be computationally complex due to an inherent minimization procedure (Pele & Werman 2009). However, in practice, we are often interested in how close a movement model is to a *data set*, rather than a probability distribution that reflects *reality*. It turns out that

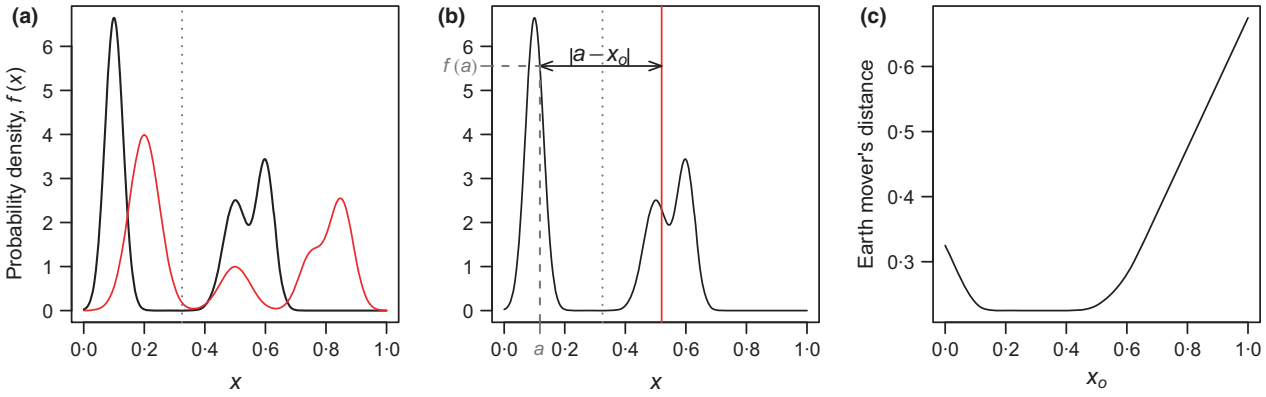

Fig. 1. The Earth Mover's Distance (EMD). Panel (a) shows two probability distributions. Imagine that the black one is a pile of earth and we want to construct from this the red distribution in the most efficient way possible. The EMD gives the average distance that we would need to move each particle of earth when performing this transformation. Panel (b) shows the situation relevant to the present study, where we have a model probability distribution of possible states that a complex system might be in at time τ in the future (black curve), together with an observation of the state at which the system actually ended up in, denoted by point $x = x_o$. This observation translates to a Dirac delta probability density function $\delta(x-x_o)$, assuming that the observation has negligible error. The EMD in this instance is the average distance each part of the probability density function has to move to end up at x_o. For example, at point $x = a$, we need to move $f(a) = 5{\cdot}5$ amount of probability distribution a distance of $|a-x_o|$. By integrating the product $|a-x_o|\ f(a)$ over all such a, we obtain the EMD between the model and data, for a single data point (1). The dotted line denotes the mean of the black distribution. Although we illustrate this in 1 dimension for ease of explanation, typically complex movement models may have states in much higher dimensions. Panel (c) shows the EMD as a function of the observation x_o.

the earth mover's distance between a model and a data set is considerably easier to compute than that between two probability distributions, as it obviates the need for minimization (see Appendix S2).

Suppose we have data on a complex system saying that it is in states $\mathbf{S}_0$, $\mathbf{S}_1$,..., $\mathbf{S}_N$ at times 0, τ, 2τ, ..., $N\tau$ respectively. Then the probability density function describing the transition between data point $n-1$ and n is just a Dirac delta function $P^R(\mathbf{X}|\mathbf{S}_{n-1}) = \delta(\mathbf{X}-\mathbf{S}_n)$. In other words, the probability of the system transitioning to any state other than $\mathbf{S}_n$ is zero, and the integral of the probability density function is equal to 1.

Suppose also that the best model we have so far constructed for these data is $P^M(\mathbf{X}_{t+\tau}|\mathbf{X}_t)$. Then the *earth mover's distance* (EMD) between this model and a data point $\mathbf{S}_n$, given a previous data point $\mathbf{S}_{n-1}$, is

$$\mathrm{EMD}(P^M;\mathbf{S}_n) = \int_\Omega d(\mathbf{X},\mathbf{S}_n)P^M(\mathbf{X}|\mathbf{S}_{n-1})\mathrm{d}\mathbf{X}, \qquad \text{eqn 1}$$

where d is a distance metric between the states of the system and Ω is the space of all system states. For example, d could be the Euclidean distance D_E between two points in space and Ω could be a subset of a 2-dimensional plane, if modelling a single terrestrial animal's movement. As another example, for a collective system with K animals, d could represent the mean Euclidean distance between pairs of points for each animal, $d(\mathbf{x}^1,\ldots,\mathbf{x}^K|\mathbf{y}^1,\ldots,\mathbf{y}^K) = \frac{1}{K}\sum_{k=1}^K D_E(\mathbf{x}^k,\mathbf{y}^k)$, where $\mathbf{x}^k$, $\mathbf{y}^k$ are points in Ω. If the model is in discrete space, so that Ω is a finite set of points, one simply replaces the integral in eqn 1 with a sum and divide by the number of points in Ω. An illustration of eqn 1, in the simplest case of one agent moving in one dimension, is given in Fig. 1b. Notice that the Kullback–Liebler distance [see e.g. Burnham & Anderson (2002)] only gives information based on the value of $f(x_o)$, whereas EMD takes into account the shape of the entire distribution $f(x)$.

The definition in eqn 1 implicitly assumes that the noise in the data is negligible. This is often reasonable for movement models constructed from GPS data of animals, as the error in GPS trackers is very highly correlated (Severns & Breed 2014). However, if it is necessary to take into account of noise, this can be done by replacing 1 with the general definition of EMD, given in Appendix S2 equation (1) with $p = 1$. Although analytically simple, the EMD in this case is much more intensive to compute than eqn 1. See Appendix S2 for more details.

Notice that if the model were deterministic then the state $\mathbf{X}$ of the model at time $t+\tau$ given that it was at state $\mathbf{X}_t$ at time t is $\mathbf{X}_{t+\tau} = f(\mathbf{X}_t)$ for some function f. Writing this in the notation of probability distributions, we have $P^M(\mathbf{X}_{t+\tau}|\mathbf{X}_t) = \delta[f(\mathbf{X}_t)]$ so that the earth mover's distance for each data point $\mathbf{S}_n$ is precisely the absolute residual $|\mathbf{S}_n - f(\mathbf{S}_{n-1})|$ of the model $\mathbf{X}_{t+\tau} = f(\mathbf{X}_t)$. In other words, eqn 1 generalizes the concept of a residual, rationalizing the choice of this particular metric over the others available (Gibbs & Su 2002).

The EMD between a model and the whole data set $\mathbf{S}_0,\mathbf{S}_1,\ldots,\mathbf{S}_N$ is the mean of eqn 1 over all the data points

$$\mathrm{EMD}(P^M;\mathbf{S}_0,\ldots,\mathbf{S}_N) = \frac{1}{N}\sum_{n=1}^N \mathrm{EMD}(P^M;\mathbf{S}_n). \qquad \text{eqn 2}$$

One drawback of the EMD is that it gives more weight to distributions with higher variance. Also, it is a dimensional quantity, with units of space. To mitigate against these issues, we use the dimensionless *standardized EMD* (SEMD), $E_S(P^M;\mathbf{S}_0,\ldots,\mathbf{S}_N)$, which is defined by dividing the EMD by the standard deviation s_n of the model. For a single data point, this is

$$\mathrm{EMD}_S(P^M;\mathbf{S}_n) = \frac{\mathrm{EMD}(P^M;\mathbf{S}_n)}{s_n}, \qquad \text{eqn 3}$$

and s_n is the standard deviation of the model for moving from position $\mathbf{S}_{n-1}$. In other words,

$$s_n^2 = \left[\int_\Omega \mathbf{X}^2 P^M(\mathbf{X}|\mathbf{S}_{n-1})\mathrm{d}\mathbf{X} - \left(\int_\Omega \mathbf{X} P^M(\mathbf{X}|\mathbf{S}_{n-1})\mathrm{d}\mathbf{X}\right)^2\right]. \quad \text{eqn 4}$$

The SEMD between a model and a sequence of data points $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_N$ is

$$\mathrm{EMD}_S(P^M; \mathbf{S}_0, \ldots, \mathbf{S}_N) = \frac{1}{N}\sum_{n1}^{N} \mathrm{EMD}_S(P^M; \mathbf{S}_n). \quad \text{eqn 5}$$

Any method described in this paper using EMD can equally be performed using SEMD, so we explain everything just using EMD, for simplicity. However, as the Results show, either SEMD or EMD may be preferable depending on the situation.

Further information about the model can be gained by looking at the mean directions from the model to the data, given by $\hat{\mathbf{v}}_n = \mathbf{v}_n/|\mathbf{v}_n|$ where

$$\mathbf{v}_n = \int_\Omega (\mathbf{S}_n - \mathbf{X}) P^M(\mathbf{X}|\mathbf{S}_{n-1})\mathrm{d}\mathbf{X}, \quad \text{eqn 6}$$

so that $\hat{\mathbf{v}}_n$ is a unit vector in the direction from the data point $\mathbf{S}_n$ to the mean of the distribution $P^M(\mathbf{X}|\mathbf{S}_{n-1})$ that predicts where $\mathbf{S}_n$ is likely to be. When the system state is given by positions on a 2D plane, this information can be visualized by plotting each line from the origin to the position given by $\hat{\mathbf{v}}_n EMD(P^M; \mathbf{S}_n)$, giving a *wagon wheel* of directional EMDs (Fig. 2a,b). However, for a large data set, this can be somewhat messy. Instead, we bin the directions into eight equal sections, constructing what we call a *dharma wheel* (Fig. 2c–f), for its resemblance to the Buddhist symbol for the noble eightfold path (Beer 2005). The smaller the dharma wheel, the more accurate the model.

Dharma wheels are examples of the classical concept of a polar area diagram (Friendly 2008). The choice of eight sections is quite arbitrary, and, depending on the situation, it may be valuable to use a different number. Alternatively, one could obtain a smoother wheel by fitting the wagon wheel to a mixture of wrapped normal distributions. However, for simplicity of explanation, we use eight sections throughout this paper.

Dharma wheels also detect bias in data (Fig. 2e,f), in analogy with residual analysis for linear models (Zuur *et al*. 2009). In addition to binning by direction, insight can be gained by constructing histograms of EMD against specific properties of the system (see Results).

An important use of EMD is to investigate goodness-of-fit statistically, by testing the null hypothesis 'the model could have given rise to the data' against the alternative that it fails in this regard. We assume the modeller has already used some form of selection technique (e.g. AIC, BIC) to find and parametrize the best of the models so far considered. Then the following sequence of steps enables the modeller to find out whether this best model reflects the data well:

1. suppose there are N data points (henceforth *the data*) and a best candidate model (henceforth, *the model*)
2. simulate the model for N steps and repeat M times, where M is as big as is computationally feasible
3. for each simulation, generate the EMDs between each of the simulated paths and the model used to simulate them, to give M distances $\mathrm{EMD}_1, \ldots, \mathrm{EMD}_M$
4. find the EMD between the data and the model, $\mathrm{EMD}_{\mathrm{data}}$
5. we can then test whether $\mathrm{EMD}_{\mathrm{data}}$ is likely to be a sample from the distribution given by $\mathrm{EMD}_1, \ldots, \mathrm{EMD}_M$. We use a 5% significance level, so that if $\mathrm{EMD}_{\mathrm{data}}$ is greater than the 97·5 percentile of $\mathrm{EMD}_1, \ldots, \mathrm{EMD}_M$, or less than the 2·5 percentile, then we reject the null hypothesis.

Notice that Step 5 is precisely equivalent to testing whether the Bayesian P-value $\mathrm{Prob}(\mathrm{EMD}_i \neq \mathrm{EMD}_{\mathrm{data}}|$ the model) is less than 5% (Agresti & Hitchcock 2005).

Methods

To show the practicality of our approach, we demonstrate how to use the earth mover's distance (2) for models of animal movement in heterogeneous environments. We use a simulated data set to test the efficacy of our model, based on an animal moving in an environment with two resource layers in a 1000 by 1000 square lattice (Fig. 3). These can be thought of, for example, as Geographic Information System (GIS) layers or resource distributions (Bolstad 2005). The layers are Gaussian random fields, generated by the R function `GaussRF()` from the `RandomFields` package (Schlather *et al*. 2013), using the `exponential` model. Both layers have mean =0, variance =1 and nugget =0. Layer 1 has scale =10 so varies rapidly through space. For the sake of intuition, this might be thought of as denoting the amount of food available throughout the terrain. For Layer 2, scale =1000, thus varies much more slowly than layer 1. This layer could represent the topography or another large geographical constraint to movement, for example. Disregarding the effect of the layers, animals move as random walkers with exponentially distributed step lengths which have a mean length of 5 lattice points. Then the effect of the layers on the animal's movement follows the concept of a step selection function, so that the probability $f(\mathbf{x}|\mathbf{y})$ of moving to position $\mathbf{x}$ from position $\mathbf{y}$ in a time interval τ is given by

$$f(\mathbf{x}|\mathbf{y}) = K\exp[\underbrace{-\lambda|x-y|}_{\text{step length distribution}} \underbrace{+\alpha w_1(x) + \beta w_2(x)}_{\text{effect of resource layers}}], \quad \text{eqn 7}$$

where $\lambda = 1/5$, $w_i(\mathbf{x})$ is a function taking the value of Layer i at position $\mathbf{x}$, α and β are the model parameters and K is a constant that ensures that the integral of f with respect to $\mathbf{x}$ is 1, so that f is a probability distribution. This is a simple version of a step selection function, or movement kernel, often used for modelling animal movement [e.g. Fortin *et al*. (2005); Forester, Im & Rathouz (2009); Potts, Mokross & Lewis (2014b)].

We generate two different simulated data sets. One is of 100 different animals, starting at random locations in the grid, for 1000 'steps' (by which we mean 'movements between successive location fixes') each. The other is of 10 animals, again with random starting points, simulated for 500 steps each. This

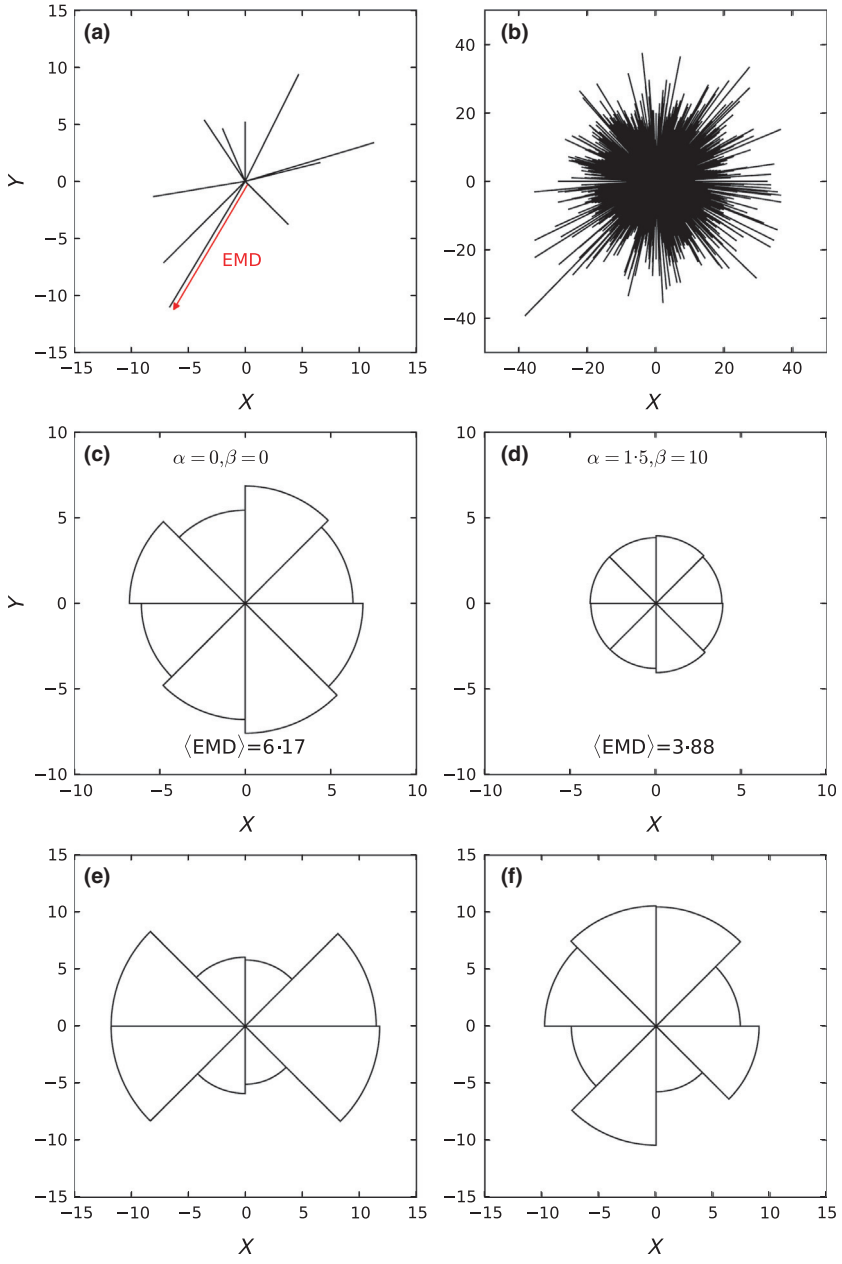

Fig. 2. Wagon wheels and dharma wheels. Panel (a) shows a wagon wheel for a hypothetical data set of 10 points. The length of each line from the origin is the earth mover's distance (EMD) for a single data point. The direction of the line is the mean direction from the model to the data point. This becomes rather messy when there are many data points, as panel (b) shows, where there are 100 000 simulated points. Instead, we bin the spokes up into eight segments, to construct a dharma wheel (Panels c,d). The dharma wheels in Panels (c,d) were created using a simulated data set of the model in eqn 7 with $\alpha = 1{\cdot}5$ and $\beta = 10$. The dharma wheel obtained by calculating the EMD from this data set to two different models of the form in eqn 7 is shown here. The mean EMDs, denoted ⟨EMD⟩, are given within the panels, together with the parameter values used. The latter correspond to models f_1, f_4 from the main text for panels (c,d), respectively. Panel (b) was also constructed from f_1. Panels (e,f) were constructed using simulated data from a random walk with a tendency to move faster in the east–west than north–south direction. Panel (e) shows the dharma wheel using the EMD from this data set to an unbiased random walk model. Panel (f) shows the EMD to the model from which the data were simulated.

enables us to demonstrate the relative effectiveness of our methods as applied to different sizes of data set. The simulated data have $\alpha = 1{\cdot}5$ and $\beta = 10$ (eqn 7). We analysed the simulated data $\{\mathbf{x}_0, \ldots, \mathbf{x}_N\}$ (which can be thought of as a special case of the arbitrary data $\{\mathbf{S}_0, \ldots, \mathbf{S}_N\}$ above) by finding the EMD from $\{\mathbf{x}_0, \ldots, \mathbf{x}_N\}$ to four different models, $f_j(\mathbf{x}|\mathbf{y})$, $j = 1,2,3,4$. Each model is of the form in eqn 7. Model f_1 has $\alpha = 0, \beta = 0, f_2$ has $\alpha = 1{\cdot}5, \beta = 0, f_3$ has $\alpha = 0, \beta = 10$ and for f_4, $\alpha = 1{\cdot}5, \beta = 10$.

To show that dharma wheels can detect bias in a process that may not be evident in the underlying model, we simulated 100 animals on a 1000 by 1000 square lattice for 1000 time steps each, performing a random walk with step length distribution $(1/5) \exp[-x/5g(\theta)]$ where $g(\theta) = (8/5)[(9/5)\cos^6(\theta) + (1/5)$

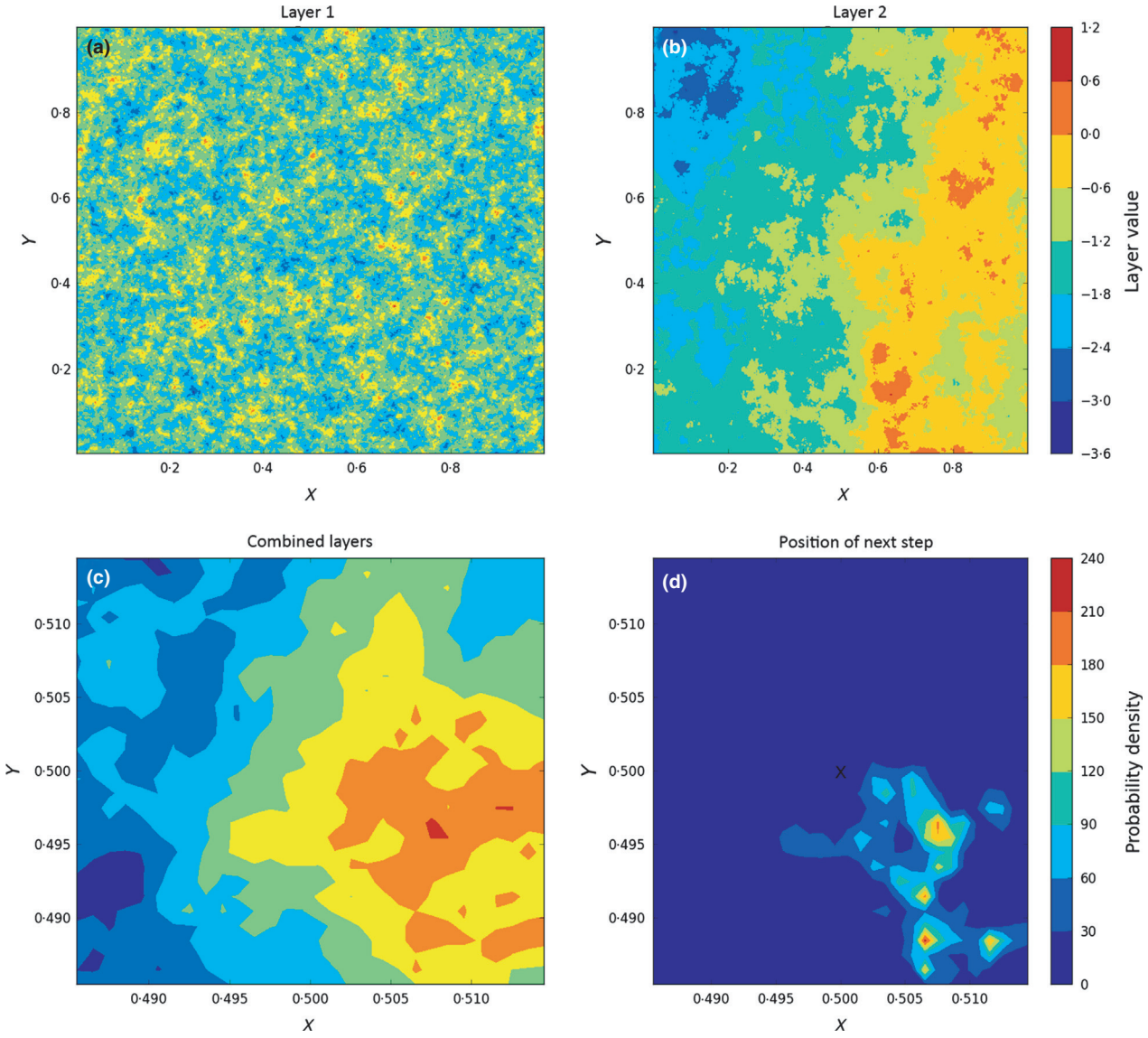

Fig. 3. Example scenario of a complex movement model. An animal moves in a heterogeneous environment, with some randomness but also a tendency to move towards articular regions of space. Panels (a) and (b) are simulated Geographic Information System (GIS) layers. The higher the value of the layers at a given point, the more that each movement the animal undergoes is biased towards that point. Panel (c) shows the combined biasing effect of the two layers in a region close to the centre of the simulated study area (using $\alpha = 1{\cdot}5$, $\beta = 10$ in the notation of the Methods section). Panel (d) shows the probability distribution of where an animal, starting at the centre, will move to after a time τ has elapsed (see Methods for details).

$\sin^6(\theta)]$ and θ is the animal's bearing. This means the animal tends to move more in the east–west direction than north–south. For example, this could be due to a confining valley running from east to west. We found the EMD between these simulated data and a random walk with a uniform exponential step length distribution, mean 5 units. We compared it to the EMD between these data and the model from which they were generated.

As a demonstration of the hypothesis test explained in the 'Earth mover's distance' section, we use simulated data sets with $\alpha = 1{\cdot}5$ and all integer values of β ranging from 0 to 10. We imagine that someone has gathered these data but only knows about, or has data on, Layer 1. Therefore, the best model that this person can construct has $\alpha = 1{\cdot}5$ and $\beta = 0$. We test the hypothesis that this latter model is an accurate description of the various simulated data sets, using the above test. This mimics a situation where two layers (Layer 1 and Layer 2) are affecting animal movement but the data gatherer has only thought to test one of them (Layer 1). It tests how well the technique does at informing the user that there is something missing from the model.

For each pair of parameter values (α,β), we simulate 500 data sets. We test the null hypothesis that 'a model with $\alpha = 1{\cdot}5$ and $\beta = 0$ accurately describes the data' using the above test on each of the 5500 data sets (500 for each value of β). Then, for each value of β, a certain percentage of the tests accept the null hypothesis, while the rest reject it, so we can plot this percentage against β to give a power curve. A better test

would have more hypothesis tests rejected for values of $\beta > 0$, whilst having the same or less hypothesis tests rejected for $\beta = 0$. Thus, we can assess the relative power of the test using EMD and that using SEMD, by examining the respective areas under the power curves. We use this to test whether the SEMD improves the power of our hypothesis testing as compared with ordinary EMD. The power is also likely to be affected by the size of the data set. To test this, we performed the same power test but for 100 animals moving 1000 steps each. As this is highly computationally intensive, we used $M = 100$ and simulated only 100 data sets, rather than 500. Simulations are performed in the C programming language and data analysis in Python. The Python code has also been translated into R, and the C code can be run from R. The code can be downloaded from the Data Dryad Repository (doi:10.5061/dryad.9h42f). We include an instruction manual in Appendix S1.

To test the applicability of our technique for models on a real data set, we use a recent model of bird flock movements and territorial interactions in the Amazon rain forest. The flocks are multispecies, with the cinereous antshrike (*Thamnomanes caesius*) playing a nuclear role in flock cohesion and movement (Munn 1986). Details of the data collection methods, justification for them, the rationale behind the model construction and the model selection techniques are all given in Potts, Mokross & Lewis (2014b). Although we do not duplicate these specifics here, we give a brief summary of the model.

The model treats each flock as a single, moving unit. This reflects the nature of the data which were gathered using the cinereous antshrike's position, where possible, to infer the flock's central location. The antshrike was usually conspicuous in the centre of the flock. Data were gathered at 30 s intervals. The flocks tend to move from tree to tree approximately once every 1–2 min. Model selection techniques reveal that a 1-min timescale is the best timescale to model these flocks' movement (Potts *et al.* 2014d) so the model we use has $\tau = 1$ min. An exponentially decaying distribution of movement lengths between successive 1-min location fixes is used to model the bird's movement. The model also includes the intrinsic persistence in the birds' movement. In addition to this, the birds are modelled as having a preference for higher tree canopies and lower ground. Finally, the birds know where other flocks have been in the recent past, due to vocalizations, so they are modelled as moving away from areas that have been visited by other flocks.

We use the EMD testing procedure, given at the end of the previous section, with $M = 1000$, to test the hypothesis that the model detailed in Potts, Mokross & Lewis (2014b) could have given rise to the data observed in the same study. We examine the resulting dharma wheel as well as the histogram of EMD against canopy height, topography, change in canopy height over a step and change in topography over a step.

Results

The example process we use is of a hypothetical animal in a heterogeneous environment consisting of two layers that affect movement (Fig. 3). As expected, the dharma wheel for the EMD between simulated data and a model that accurately reflects the simulations (Fig. 2c) has a smaller area than one with certain parameters suppressed (Figs 2d and S2). Visually, this is not so apparent with SEMD, suggesting that there is value in using EMD for such qualitative tests.

Figure 2e,f shows what happens when there is a bias in the movement process. If the model fails to take this into account, then the resulting dharma wheel is skewed in the direction of the bias, in this case parallel to the x-axis. A model that does take this bias into account results in a more symmetric dharma wheel (Fig. 2f), albeit with some random variation.

By constructing histograms of EMD against the value of the layer where each step ends, Fig. 4 shows that model f_4 is better at predicting the observations when animals are in environments where the value of layer 1 is relatively high (see Methods for descriptions of models f_i). However, if it is low, then model f_3 performs marginally better. This is important if we wish to

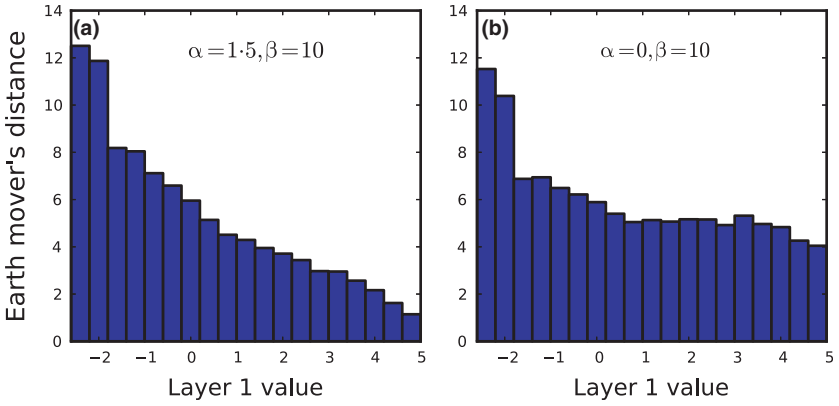

Fig. 4. EMD binned by value of layer 1. Using 100 000 simulated data points from the model 7 with $\alpha = 1{\cdot}5$ and $\beta = 10$, the EMDs from these data to the model with (a) $\alpha = 1{\cdot}5$ and $\beta = 10$, and (b) $\alpha = 0$ and $\beta = 10$. EMDs for each step are binned according to the value of Layer 1 (Fig. 3a) at the point where the step ends. Unless this value is very low, the model with $\alpha = 1{\cdot}5$ is better, otherwise a model excluding the effect of Layer 1 is better. This shows how EMD can be used to ascertain which environments a model may prove to be good at predicting movements, and where it is likely to fail.

use a model parametrized in one study area to predict movement in another. In the example here, if we imagine layer 1 denotes food availability, then Fig. 4 tells us that our overall best model, f_4, will be good at predicting movement in a food-rich environment but may be quite bad if we try to use it in a food-poor area.

Performing a power analysis of the hypothesis test detailed in the previous section by varying β in the simulations (see Methods), SEMD performed far better than ordinary EMD (Fig. 5a,b). Ordinary EMD turned out to be quite poor when used for testing the hypothesis that a model accurately reflects the data, almost always failing to reject the null hypothesis when it should, thus being highly susceptible to type II errors (Casella & Berger 2002). Replacing SEMD with log-likelihood in the hypothesis test also yielded worse results. Therefore, we would recommend using SEMD for testing this type of hypothesis.

The power of the test depends upon both the actual underlying processes and the size of the data set. Figure 5 demonstrates that using 100 000 data points (Panels b and d) rather than

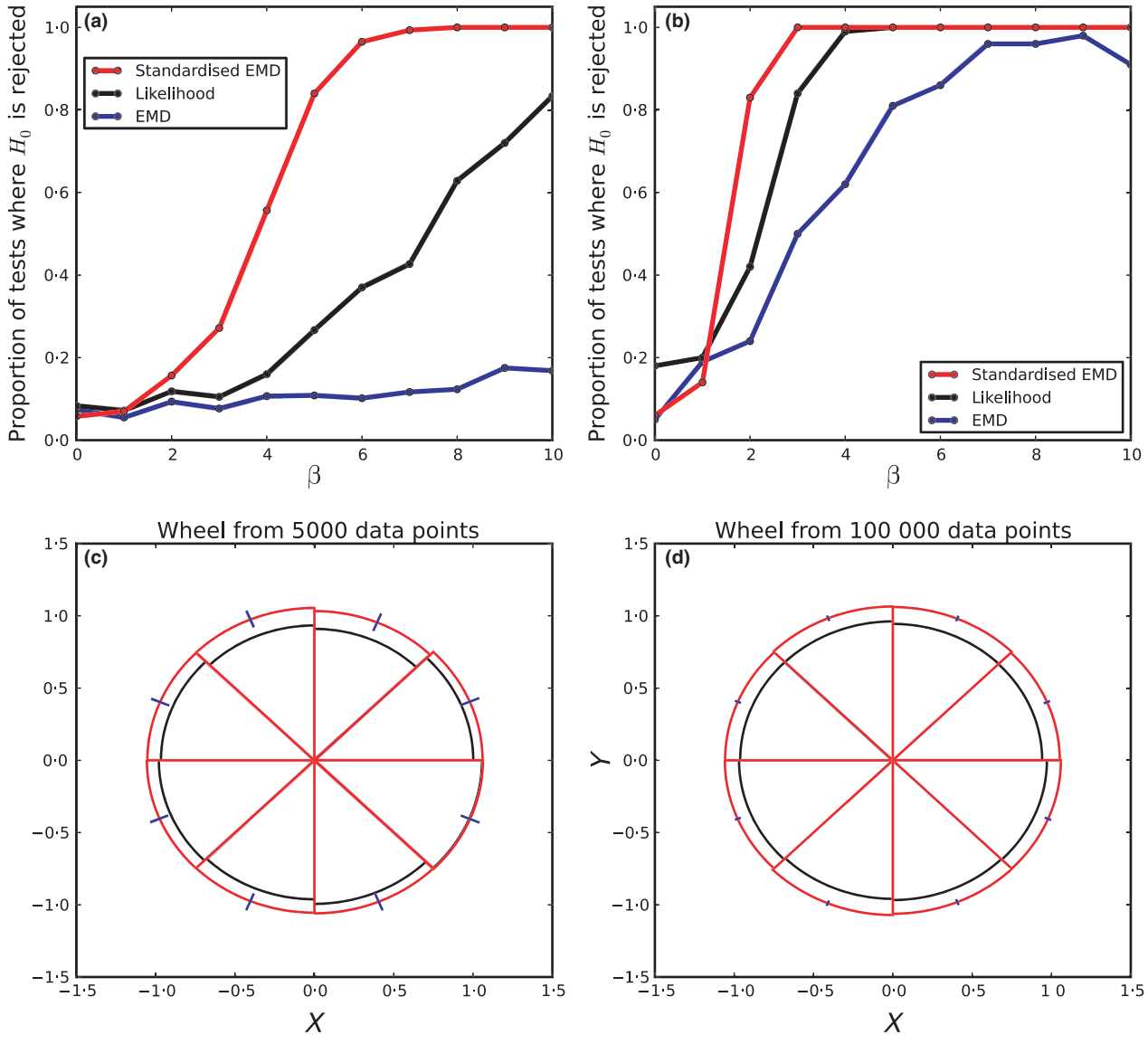

Fig. 5. Power of the EMD hypothesis test. Panel (a) shows the proportion of simulated data sets for which the hypothesis that a model with α = 1·5 and β = 0 accurately describes the data were rejected, for each value of β used. Using SEMD proves to be far preferable to both ordinary EMD and likelihood. In all situations, approximately 5% of data sets with β = 0 result in type I errors, as expected due to the use of 95% confidence intervals. As β is increased, the number of type II errors decreases, to the point where zero out of 500 data sets exhibited type II errors occurred when using SEMD, if β≥8. Panel (b) shows a similar plot, but using 100 000 data points for each simulated set, rather than the 5000 used in panel (a), showing that the larger the data set, the stronger the power of the test. Panels (c) and (d) represent the hypothesis test visually. The black curves show dharma wheels of simulated data with α = 1·5 and β = 10 tested against a model with α = 1·5 and β = 0. The red curves show dharma wheels of the mean of 1000 simulated data sets with α = 1·5 and β = 0 tested against a model with α = 1·5 and β = 0 (red curves). The blue lines show 95% confidence intervals. Each simulated data set for constructing Panel (c) had 5000 points, while Panel (d) used 100 000 points for each data set.

merely 5000 (Panel a and c) gives much smaller error bars on the EMD (Panels c and d) and rejects false null hypotheses more frequently (Panels a and b).

Using the EMD procedure to test the quality of the model of Amazonian bird flock movement from Potts, Mokross & Lewis (2014b) against the data described there revealed that the model is insufficient to describe the data in full accuracy. This rejection occured at both the 5% and 1% significance levels. Using notation from the end of the section 'Earth mover's distance: a measure of absolute fit', $EMD_{data} = 1{\cdot}737$, whereas the 0·5 and 99·5 percentiles of $EMD_1,\ldots,EMD_M$ are 1·690 and 1·691 respectively. Using SEMD, the corresponding values are $EMD_{data} = 1{\cdot}153$ and 0·5 and 99·5 percentiles are 1·129 and 1·134 respectively.

The resulting dharma wheel is very round (Fig. S3), suggesting that the model is unlikely to be missing a directional bias. Histogramming the EMDs by canopy height and topography also reveals no clear trend (Fig. S4a,b). When we use the difference in canopy height from the start to the end of the step, there is also no clear trend in EMD (Fig. S4c). However, when we histogram by difference in topography, the EMD is lower for steps where the difference in topography is lower (Fig. S4d). This suggests that there may be some additional trigger that causes the particular decisions to move to higher or lower ground. These observations are unchanged when we use standardized EMD (Fig. S5)

Discussion

We have constructed a generalized version of a residual that is usable with complex stochastic movement models. We have given a method for using this to test the validity of a model in an absolute rather than relative sense, as well as showing how this concept can give visual insight into the strengths and shortcomings of a stochastic model. By testing our techniques on simulated data, where we have control over the mechanisms underlying movement decisions, we have demonstrated that our techniques are tractable and can give useful insight into realistic situations.

While many techniques exist for comparing models within a limited class, for example AIC and BIC (Burnham & Anderson 2002), our techniques can be used for showing whether there is an important model parameter missing from outside that class. This would help mitigate against scientists making bold conclusions after having only fitted data to a small number of poor models, only later to have those conclusions refuted when a more realistic one is used (Edwards *et al.* 2007; Jansen, Mashanova & Petrovskii 2012; Auger-Méthé *et al.* 2014).

Incorporating the correct level of complexity is also of great importance if we want to construct models that are truly predictive, as such models must include all necessary mechanisms to make the predictions accurate (Evans *et al.* 2013). To this end, we believe that combining our EMD techniques with cross-validation may prove to be useful (Seymour 1993). One might use model selection on one subset Σ of a data set, then analyse the rest of the data, Σ^c, by finding the EMD between Σ^c and the best model. If this EMD is very different from the EMD between the best model and Σ, then this suggests that the model has weak predictive power. Turning this into a rigorous statistical test would be a useful extension of our approach.

The main aim of our hypothesis test procedure is to evaluate goodness-of-fit by rejecting models that do not capture the data well. Although this worked reasonably well in the scenarios we examined, if the data set is too small or the effect of a certain covariate too mild, then the procedure can be prone to Type II errors, that is failing to reject an incorrect null hypothesis (Fig. 5a). The chance of making Type I errors, that is rejecting a correct null hypothesis, is simply the same as the confidence interval used in the testing procedure (Casella & Berger 2002). As we used 95% in our study, this will happen 5% of the time, as confirmed by simulations (Fig. 5a). However, there is no reason in principle why a user of our methods could not use a different confidence interval, or search for the exact *P*-value of the test for their data set, which may be very small if the data set is large.

As well as being a natural generalization of a residual, a strength of our method is that it takes into account multimodal probability distributions, giving a low EMD to data points that are on any one of many peaks. Methods such as posterior predictive checks (PPC) (Gelman *et al.* 2004) typically examine the mean of a probability distribution, or samples thereof. Suppose, for example, we have two peaks with a low probability area between them. Then PPC would penalize an observation on one of the peaks more than in the low area between the two peaks. EMD, on the other hand, would penalize these observations roughly equally (e.g. Fig. 1b). Therefore, one could use EMD as an alternative criterion within a PPC. Generalized forms of 2, discussed in Appendix S2, can have rich properties that allow the user to decide how to penalize areas of a distribution where there are multiple peaks.

Our technique revealed that a recent model of Amazonian bird movement (Potts, Mokross & Lewis 2014b) is insufficient to describe the data it models, inasmuch as we were forced to reject the hypothesis that the data arose purely from processes described in the model. This is despite the fact that the model includes five different factors (step length, turning angle, attraction to high canopies, bias towards lower ground and repulsion away from other flocks' territories) that were all shown in the previous study to have significant impacts on bird movement. In other words, it is the best model from a variety of different hypothesized models, but it is not good enough for accurate description or prediction of movement. Consequently, we might be missing a key process in understanding the movement of these animals.

This corroborates qualitative findings from the previous study (Potts, Mokross & Lewis 2014b), which showed pictorially that this model appeared to give slightly inaccurate territorial patterns when simulated. However, quantitative confirmation of this is far better than relying on pictures, and it demonstrates that we need to think of further covariates that may be affecting bird movement if we are to build an accurate, predictive model. As the dharma wheel for this model is very

round (Fig. S2), it is likely that these covariates do not include directional biases, so we should look for other, non-directional effects. The reader is referred to the previous study for a discussion of hypothesized candidates for these effects, which include extra confinement due to memory and territorial displays (Potts, Mokross & Lewis 2014b).

The EMD analysis also revealed that the model does not capture well any moves that flocks make where there is a big difference in topography. This suggests that there is some additional trigger causing birds to move to higher or lower ground beyond those examined so far. Flocks display movements akin to patrolling areas on edges of territories on higher ground, possibly to reinforce cues for demarcation of territorial limits. In such areas, when not engaged in territorial interactions, flocks display more ballistic paths compared to the area restricted behaviour seen inside drainage valleys, indicating that intensive area restricted foraging may not be a primary motivator (K. Mokross pers. obs.). Therefore, movements to much higher ground may indicate a switch from foraging to territorial defence, whereas movements to much lower ground may indicate a switch back to foraging.

Although we have focused on animal movement in this paper, the techniques we propose can, in principle, be used for analysing any stochastic movement model. For example, these could include collective motion of humans, cancer growth or complex systems of intracellular proteins (Berendsen & Hayward 2000; Friedl & Wolf 2003; Helbing, Johansson & Al-Abideen 2007). We imagine that the techniques proposed here are merely the tip of the iceberg of possible uses for EMD in analysing movement models. We have already suggested some possible extensions, such as using the generalized form in Appendix S2, or combining EMD with cross-validation. Analysis of these situations and others, while beyond the scope of the present paper, would doubtless provide further important techniques for understanding how best to model complex systems. This paper provides an introduction to the concept of residuals generalized to stochastic movement models. We hope that future work, by us and others, will find many more riches that come from this fundamental idea.

Acknowledgements

This study was partly funded by NSERC Discovery and Accelerator grants (MAL, JRP). MAL also gratefully acknowledges a Canada Research Chair and a Killam Research Fellowship. MAM gratefully acknowledges Alberta Innovates-Technology futures, the Killam trusts, NSERC and the University of Alberta for graduate student scholarships. KM would like to acknowledge the Biological Dynamics of Forest Fragments Project (BDFFP) staff for providing logistic support; J. Lopes, E.L. Retroz, P. Hendrigo, A. C. Vilela, A. Nunes, B. Souza, M. Campos for field assistance; M. Cohn-Haft for valuable discussions. Funding for the research was provided by US National Science Foundation grant LTREB 0545491 awarded to Phil Stouffer, which helped fund KM's work. This article represents publication no. 648 in the BDFFP Technical Series and contribution no. 34 in the Amazonian Ornithology Technical Series of the INPA Zoological Collections Program. This manuscript was approved for publication by the Director of the Louisiana Agricultural Experiment Station as manuscript 2014-241-16702. We are grateful to Mike Bryniarski for compiling our code for use with Mac OS, to Phil Stouffer for helpful comments regarding data collection, to members of the Lewis Lab for helpful discussions and suggestions, and for three anonymous reviewers and an associate editor for comments that have helped improve the manuscript.

Data accessibility

Data used in this study can be obtained from Potts, Mokross & Lewis (2014c). Code for calculating EMD and generating simulated data can be found from the Data Dryad Repository (doi:10.5061/dryad.9h42f).

References

Agresti, A. & Hitchcock, D.B. (2005) Bayesian inference for categorical data analysis. *Statistical Methods and Applications*, **14**, 297–330.

Auger-Méthé, M., St. Clair, C.C., Lewis, M.A. & Derocher, A.E. (2011) Sampling rate and misidentification of Lévy and non-Lévy movement paths: comment. *Ecology*, **92**, 1699–1701.

Auger-Méthé, M., Derocher, A.E., Plank, M.J., Codling, E.A. & Lewis, M.A. (2014) Differentiating between the Lévy and the area-restricted search strategies. http://arxiv.org/abs/1406.4355.

Beer, R. (2005) *The Handbook of Tibetan Buddhist Symbols*. Shambhala, Boston, MA.

Berendsen, H.J. & Hayward S. (2000) Collective protein dynamics in relation to function. *Current Opinion in Structural Biology*, **10**, 165–169.

Bolstad, P. (2005) *GIS Fundamentals: A First Text on Geographic Information Systems*. Eider Press, White Bear Lake, MN.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference. A Practical Information-theoretic Approach*, 2nd edn. Springer, New York.

Camazine, S., Deneubourg, J.-L., Franks, N.R., Sneyd, J., Theraulaz, G. & Bonabeau, E. (2003) *Self-Organization in Biological Systems*. Princeton University Press, Princeton, NJ.

Casella, G. & Berger, R.L. (2002) *Statistical Inference*. Duxbury, Pacific Grove, CA.

Couzin I.D., Krauze, J., James, R., Ruxton, G.D. & Franks, N.R. (2002) Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, **218**, 1–11.

Deneubourg, J.L., Goss, S., Franks N. & Pasteels, J.M. (1989) The blind leading the blind: Modeling chemically mediated army ant raid patterns. *Journal of Insect Behavior*, **2**, 719–725.

Edwards, A.M., Phillips, R.A., Watkins, N.W., Freeman, M.P., Murphy, E.J., Afanasyev, V., Buldyrev, S.V., da Luz, M.G.E., Raposo, E.P., Stanley, H.E. & Viswanathan, G.M. (2007) Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, **449**, 1044–1048.

Evans, M.R. et al. (2013) Predictive systems ecology. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 1471–2954.

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

Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005) Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, **86**, 1320–1330.

Friedl, P. & Wolf, K. (2003) Tumour-cell invasion and migration: diversity and escape mechanisms. *Nature Reviews*, **3**, 362–374.

Friendly, M. (2008) The golden age of statistical graphics. *Statistical Science*, **23**, 502–535.

Gauss, F. (1809) *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Frid. Perthes & I.H. Besser, Hamburg.

Gautestad, A.O., Loe, L.E. & Mysterud, A.J. (2013) Inferring spatial memory and spatiotemporal scaling from GPS data: comparing red deer *Cervus elaphus* movements with simulation models. *Journal of Animal Ecology*, **82**, 572–586.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. & Rubin, D. (2004) *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL.

Gibbs, A.L. & Su, F.E. (2002) On choosing and bounding probability metrics. *International Statistical Review*, **70**, 419–435.

Gordon, R.A. (2010) *Regression Analysis for the Social Sciences*. Routledge, Oxford.

Grimm, V., et al. (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, **310**, 987–991.

Guttal, V. & Couzin, I.D. (2010) Social interactions, information use, and the evolution of collective migration. *Proceedings of the National Academy of Sciences*, **107**, 16172–16177.

Helbing, D., Buzna, L., Johansson, A. & Werner, T. (2005) Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions. *Transportation Science*, **39**, 1–24.

Helbing, D., Johansson, A & Al-Abideen, H.Z. (2007) Dynamics of crowd disasters: an empirical study. *Physical Review E*, **75**, 046109.

Hoare, D.J., Couzin, I.D., Godin, J.-G.J. & Krause, J. (2004) Context-dependent group size choice in fish. *Animal Behaviour*, **67**, 155–164.

de Jager, M., Weissing, F.J., Herman, P.M.J., Nolet, B.A. & van de Koppel, J. (2011) Lévy walks evolve through interaction between movement and environmental complexity. *Science*, **332**, 1551–1553.

Jansen, V.A.A., Mashanova, A. & Petrovskii, S. (2012) Comment on 'Lévy Walks Evolve Through Interaction Between Movement and Environmental Complexity'. *Science*, **335**, 918.

Jonsen, I.D., Flemming, J.M., & Myers, R.A. (2005) Robust state-space modeling of animal movement data. *Ecology*, **86**, 2874–2880.

Latombe, G., Fortin, D. & Parrott, L. (2013). Spatio-temporal dynamics in the response of woodland caribou and moose to the passage of gray wolf. *Journal of Animal Ecology*, **83**, 185–198.

Legendre, A.-M. (1805) *Nouvelles Méthodes Pour la Détermination des Orbites des Comètes*. F. Didot, Paris.

Moorcroft, P.R., Lewis, M.A. & Crabtree, R.L. (2006) Mechanistic home range models capture spatial patterns and dynamics of coyote territories in Yellowstone. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1651–1659.

Morales, J.M., Haydon, D.T., Friar, J., Holsinger, K.E. & Fryxell, J.M. (2004) Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85**, 2436–2445.

Munn, C.A. (1986) Birds that cry wolf. *Nature*, **319**, 143–145.

Pele, O. & Werman, M. (2009) Fast and robust Earth Mover's Distances. *IEEE 12th International Conference on Computer Vision*, 460–467.

Plank, M.J. & Codling, E.A. (2009) Sampling rate and misidentification of Lévy and non-Lévy movement paths. *Ecology*, **90**, 3546–3553.

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

Potts, J.R., Mokross, K., & Lewis, M.A. (2014b) A unifying framework for quantifying the nature of animal interactions. *Journal of the Royal Society Interface*, **11**, 20140333. doi:10.1098/rsif.2014.0333.

Potts, J.R., Mokross, K. & Lewis, M.A. (2014c) Data from: A unifying framework for quantifying the nature of animal interactions. *Dryad Digital Repository*, doi: 10.5061/dryad.47jh1.

Potts, J.R., Mokross, K., Stouffer, P.C., Lewis, M.A. (2014d) Step selection techniques uncover the environmental predictors of space use patterns in flocks of Amazonian birds. (http://arxiv.org/abs/1403.6869).

Rhodes, J.R., McAlpine, C.A., Lunney, D. & Possingham, H.P. (2005) A spatially explicit habitat selection model incorporating home range behavior. *Ecology*, **86**, 1199–1205.

Rubner, Y., Tomasi, C. & Guibas, L.J. (2000) The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, **40**, 99–121.

Schlather, M., Menck, P., Singleton, R., Pfaff, B., & R Core team. (2013) RandomFields: Simulation and Analysis of Random Fields. R package version 2.0.66. http://CRAN.R-project.org/package=RandomFields.

Severns, P.M. & Breed, G.A. (2014) Behavioral consequences of exotic host plant adoption and the differing roles of male harassment on female movement in two checkerspot butterflies. *Behavioral Ecology and Sociobiology*, **68**, 805–814 DOI:10.1007/s00265-014-1693-z

Seymour, G. (1993) *Predictive Inference*. Chapman and Hall, New York, NY.

Van Moorter, B., Visscher, D., Benhamou, S., Börger, L., Boyce, M.S. & Gaillard, J.-M. (2009) Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, **118**, 641–652.

Vanaka, A.T., Fortin, D., Thakera, M., Ogdene, M., Owena, C., Greatwood, S. & Slotow, R. (2013) Moving to stay in place – behavioral mechanisms for coexistence of African large carnivores. *Ecology*, **94**, 2619–2631.

Vasershtein, L.N. (1969) Markov processes over denumerable products of spaces describing large system of automata. *Problemy Peredachi Informatsii*, **5**, 47–52.

Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Murphy, E.J., Prince, P.A. & Stanley, H.E. (1996) Lévy flight search patterns of wandering albatrosses. *Nature*, **381**, 413–415.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, NY.

Received 12 May 2014; accepted 11 August 2014
Handling Editor: Robert B. O'Hara

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Instructions for running code.

Appendix S2. Generalizing the Earth Mover's Distance (EMD).

Appendix S3. Literature search for movement models.