The Annals of Applied Statistics
2019, Vol. 13, No. 4, 2483–2508
https://doi.org/10.1214/19-AOAS1289

NEW FORMULATION OF THE LOGISTIC-GAUSSIAN PROCESS TO ANALYZE TRAJECTORY TRACKING DATA[1]

BY GIANLUCA MASTRANTONIO[*,2], CLARA GRAZIAN[†,‡,2], SARA MANCINELLI[§] AND ENRICO BIBBONA[¶]

Politecnico di Torino, Università degli Studi "Gabriele d'Annunzio"*[†], *University of New South Wales*[‡], *Università di Roma "La Sapienza"*[§] *and Politecnico di Torino*[¶]

Improved communication systems, shrinking battery sizes and the price drop of tracking devices have led to an increasing availability of trajectory tracking data. These data are often analyzed to understand animal behavior.

In this work, we propose a new model for interpreting the animal movent as a mixture of characteristic patterns, that we interpret as different behaviors. The probability that the animal is behaving according to a specific pattern, at each time instant, is nonparametrically estimated using the Logistic-Gaussian process. Owing to a new formalization and the way we specify the coregionalization matrix of the associated multivariate Gaussian process, our model is invariant with respect to the choice of the reference element and of the ordering of the probability vector components. We fit the model under a Bayesian framework, and show that the Markov chain Monte Carlo algorithm we propose is straightforward to implement.

We perform a simulation study with the aim of showing the ability of the estimation procedure to retrieve the model parameters. We also test the performance of the information criterion we used to select the number of behaviors. The model is then applied to a real dataset where a wolf has been observed before and after procreation. The results are easy to interpret, and clear differences emerge in the two phases.

1. Introduction. Global Positioning System (GPS) telemetry is currently the main tool used to remotely determine the position of an animal with high precision, and at time intervals that can be programmed by the researcher (Cagnacci et al. (2010)). These data, which are often defined as "trajectory tracking data", take the form of a time series of coordinates. The large amount of data, gathered

Received April 2019; revised July 2019.

[1]Support by the Italian Ministry of Education, University and Research (MIUR) grant *Dipartimenti di Eccellenza*, CUP: E11G18000350001, conferred to the Dipartimento di Scienze Matematiche—DISMA, Politecnico di Torino.

[2]The work of the first two authors has been developed in part under the PRIN2015 supported-project Environmental processes and human activities: capturing their interactions via statistical methods (EPHAStat), funded by MIUR (Italian Ministry of Education, University and Scientific Research) (20154X8K23-SH3).

Key words and phrases. Wolf data, invariance, coregionalization, continuous-time hidden Markov model.

from GPS collars, facilitates a greater resolution in the study of habitat selection (Hebblewhite and Merrill (2008)), spatio-temporal movements (Frair et al. (2010), Morales et al. (2004)) and behavior (Anderson and Lindzey (2003), Merrill and David Mech (2000)). For recent reviews, the reader may refer to Patterson et al. (2017) and Hooten et al. (2017)

In this work, we propose a new statistical model for the analysis of trajectory tracking data. The model can detect patterns in an animal's movement, that are here interpreted as consequence of different behaviors. The motivation for this study arises from a dataset that was collected to study the behavior of a female wolf, that was observed in the Abruzzo, Lazio and Molise National Park in the central Apennines, Italy (Mancinelli, Boitani and Ciucci (2018)), before and after reproduction.

Animal movement modeling has a long history and, starting from the diffusion model of Brownlee (1911), a wide range of different approaches have been proposed. These can be grouped into three categories: (i) point process models (Brost et al. (2015), Johnson, Hooten and Kuhn (2013), Johnson et al. (2008b)); (ii) continuous-time dynamic models (Blackwell (1997), Brillinger (2010), Fleming et al. (2014b), Johnson et al. (2008a)); (iii) discrete-time dynamic models (Jonsen, Flemming and Myers (2005), McClintock et al. (2012), Morales et al. (2004)). Point process models require a numerical integration to compute the likelihood (Cressie (1993)) that limit their use, while the approaches based on continuous-time dynamics generally fail to model heterogeneity in the trajectory over time; notable exceptions can be found in Hooten and Johnson (2017), Hooten et al. (2018) and Blackwell et al. (2016). Our proposal is part of the discrete-time dynamic framework.

In a discrete-time dynamic model, the joint distribution of the coordinates, or of their appropriate transformation, is seen as a mixture process. Each component (or regime) represents a different movement pattern that may be interpreted as a particular kind of behavior. Switching between behaviors is often assumed to be temporally structured (Houston and Mcnamara (1999), Morales et al. (2004)) and sometimes even spatially structured, as in Blackwell (2003), often ruled by a nonobserved Markov process that leads to the class of hidden Markov models (HMMs) in discrete (DT-HMM) (Zucchini and MacDonald (2009)) and continuous-time (CT-HMM) (Blackwell (2018)).

Although HMMs have been widely adopted (see, e.g., Franke et al. (2006), Langrock et al. (2012), Maruotti et al. (2015)), they require knowledge or assumptions on the switching process between states, needed to define the rate functions; see, for example, Morales et al. (2004), Patterson et al. (2009) or Blackwell (2018), among others. However, we believe that this approach may be too restrictive to model animal trajectories. In this work, we present a discrete-time dynamic model, where the latent temporal probability structure is estimated via a nonparametric approach.

We define a probability vector, also called compositional vector, for each time. The elements of such vectors represent the probability that the animal is currently adopting a specific behavior. These vectors are modeled in a nonparametric fashion, which does not require prior knowledge of the possible temporal evolution, it can easily accommodate covariate information, it allows dependence to be achieved between and within the probability vectors, and it is easy to implement.

We assume that the compositional vectors follow, marginally, a Logistic-normal distribution (LogitN). The LogitN has a representation in terms of normally distributed variables. Then, to introduce both serial dependence among the compositional vectors at different times, and internal dependence among the components of each vector, we envision these variables as a realization of a multivariate Gaussian process (GP) based on the coregionalization approach (Gelfand et al. (2010)). The Logistic-Gaussian process (LogitGP) is then defined.

This is not the first construction that has used GPs to introduce dependence over compositional vectors. Nevertheless, our formalization allows inference to be performed in such a way that it enjoys two properties, which we believe are essential in this setting: (i) invariance with respect to the choice of the reference element; (ii) invariance with respect to the reordering of the labels that identify the behaviors. These two points have often been overlooked in the literature (Brunsdon and Smith (1998), Paci and Finazzi (2018)) while, in other cases, oversimplified hypotheses have been imposed and this has resulted in a reduced flexibility of the model (Martins, Bonat and Ribeiro (2016), Tjelmeland and Lund (2003)).

In the recent literature, the flexibility of GPs has often been exploited to model trajectory tracking data with different aims. We mention, for example, Johnson et al. (2008a), Fleming et al. (2014a, 2014b), Blackwell et al. (2016), Calabrese, Fleming and Gurarie (2016), Fleming and Calabrese (2019), Gurarie et al. (2017), Scharf et al. (2018). However, none of these references have any direct relationship with the methodology proposed here, since they use a GP to directly model the observed locations.

We fit the model under a Bayesian framework, and propose a Markov chain Monte Carlo (MCMC) algorithm that is straightforward to implement. In order to fit the model, the number of latent behaviors must be specified, and an offline procedure should then be used to select the best value. We suggest to use the integrated classification likelihood (ICL) (Biernacki, Celeux and Govaert (2000)) and we test its performance in a simulation study, and show that it retrieves the right number of latent states most of the time.

The model is applied to the dataset that motivated the study. It consists of a series of GPS positions of a female wolf over two separate time windows, that correspond to different phases of the animal's life. In the first period, movements are more erratic, while in the second, right after reproduction, movements are more regular, and exhibit the classic star-shaped pattern around the den (Mech and Boitani (2003)).

Three behaviors are detected: one slow-speed behavior, which we call R (for Resting), and two high speed behaviors with different characteristics (HE1 and HE2, after Hunting/Exploring). Only R and HE1 are observed in the first time window, while only R and HE2 are observed in the second time window. The temporal characterization of the R behavior is also different for the two time windows, since in the first it is more likely during daylight and in the second during the early hours of the night. A biological interpretation of these behaviors is investigated.

The paper is organized as follows. Section 2 presents the motivating example and a full description of the dataset that is analyzed. Section 3 describes the model, while Section 4 gives more details on the LogitGP, its properties and its connections and differences from previous LogitGP-based models and CT-HMM. Section 5 presents the results of the model fitted on the wolf data, and Section 6 concludes the paper. We provide implementation details in Appendix A, while Appendix B presents the simulation study.

2. Data. We have recorded a time series of spatial locations following the female wolf called F24 (2–3 years old), which was live-trapped in May 2009 in the Abruzzo, Lazio and Molise National Park (central Apennines, Italy) and equipped with a Vectronic Pro Light-1 collar (Vectronic Aerospace GmbH, Berlin, Germany). Details on the capturing and handling of the wolf are provided in Mancinelli, Boitani and Ciucci (2018). During winter months (January–April), the recording rate was scheduled for every 30 minutes, while it was every 3 hours for the rest of the year. The higher acquisition rate during winter months was necessary to estimate wolf kill rates through field investigations of GPS clusters (Sand et al. (2010)).

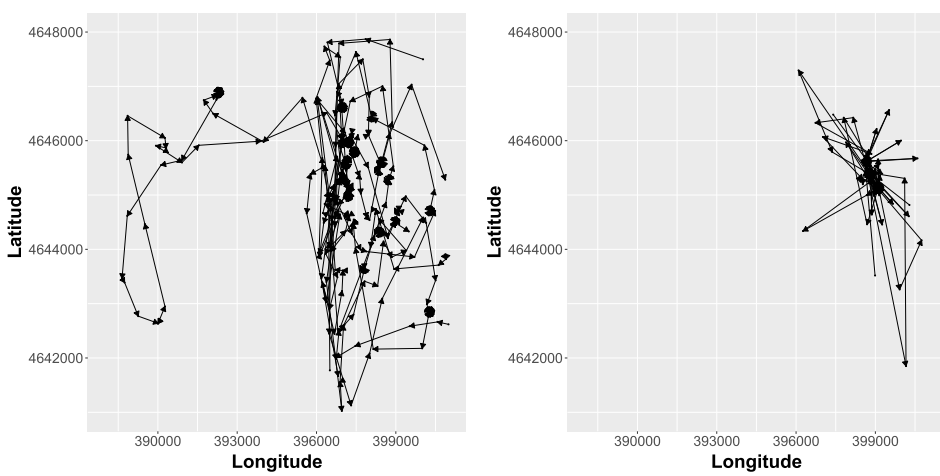
When first captured, F24 was a member of the Villa pack, where it remained for 7.9 months before leaving and establishing a new pack (the Bisegna pack) in January 2010. The wolf gave birth in May 2010 and, using information derived from its GPS locations, it was possible to determine the actual position of the den. Between May 28 and June 4, F24 restricted its movements in the proximity to the den. Therefore, we assumed that F24 entered the den on May 28 and gave birth in the following days. Until collar failure occurred, on the 16th of June, it was observed that F24 systematically revisited the den to feed and attend the cubs. In order to detect the changes in behavior before and after wolf reproduction, we use a subset of the data relative to the final period in which F24 settled down with the Bisegna pack. The dataset is divided into two time windows, before and after reproduction, which are described in Table 1. The trajectories are shown in Figure 1, where it is possible to observe that the movements of F24 in the first period are more nomadic, while the classic star-shaped pattern around the den, which characterizes the wolf reproductive period (Mech and Boitani (2003)), can be observed in the second.

TABLE 1
Observed time windows—Temporal intervals, sample rates, number of observations and missing data

Start	End	Interval	Number of obs.	Missing
03/01 00:00	03/10 23:30	30 minutes	473	7
05/28 00:00	06/16 09:00	3 hours	138	18

2.1. *Preliminaries and notation.* It should be noted that the two time windows have different recording rates and there is a temporal gap between the two. We denote $\mathcal{T}_j = (t_1^j, \dots, t_{T_j}^j)$ the set of equally spaced time instants, separated by 30 minutes, which covers the jth time window, and $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. The set of spatial locations is denoted by $\mathbf{s} = \{\mathbf{s}_t\}_{t \in \mathcal{T}}$, where each element $\mathbf{s}_t = (s_{t,1}, s_{t,2})'$ is a longitude and latitude vector. The elements of $\mathbf{s}$ that are not observed are considered missing.

We define $t+$ to be the element that follows t in $\mathcal{T}_j$, for every $t \in \mathcal{T}_j$ but the last one. Similarly, we define $t-$ to be the element that precedes t in $\mathcal{T}_j$, for every $t \in \mathcal{T}_j$ but the first one. We define $\mathbf{y}_t = (y_{t,1}, y_{t,2})'$, for $t \in \mathcal{T}_j \backslash \{t_1^j, t_{T_j}^j\}$, as the rotated vector of the coordinate increments, expressed in Km, between times t and $t+$, so that $y_{t,1}$ is the projection of $\mathbf{s}_{t+} - \mathbf{s}_t$ onto the direction of the previous increment $\mathbf{s}_t - \mathbf{s}_{t-}$ (see Figure 2). The values of $\mathbf{y}_t$ with $t \in \{t_1^j, t_{T_j}^j\}$ are considered to be missing. A classic approach for animal behavior modeling involves the analysis of the following variables: the *step-length* $r_t \in \mathbb{R}^+$, which is a proxy of the speed



(a) First time window (b) Second time window

FIG. 1. *Trajectories in the two time windows.*

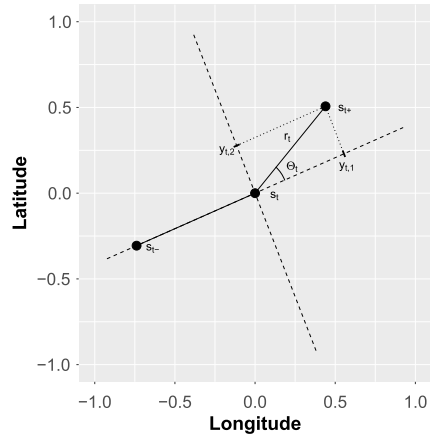


FIG. 2. *A graphical example of the relation between* $\mathbf{s}_{t_i}$, $(\mathbf{r}_{t_i}, \theta_{t_i})$ *and* $\mathbf{y}_{t_i}$.

of the animal, and the *turning-angle* $\theta_t \in [0, 2\pi)$, which describes how the animal turns. These variables are transformations of $\mathbf{y}_t$, as described in Figure 2. The explicit relation is $r_t = \|\mathbf{y}_t\|$ and $\theta_t = \operatorname{atan}^*(y_{t,2}, y_{t,1})$, where atan^* is the two-arguments arctangent, which is a quadrant-specific inverse of the tangent function (Mardia and Jupp (2000)).

The trajectory, within each time window, can then be described as

$$\mathbf{s}_{t+} = \mathbf{s}_t + \mathbf{R}_{t-,t}\mathbf{y}_t \quad \Rightarrow \quad \mathbf{y}_t = \mathbf{R}_{t-,t}^{-1}(\mathbf{s}_{t+} - \mathbf{s}_t), \tag{1}$$

where $\mathbf{R}_{t-,t}$ is the rotation matrix based on the angle θ_{t-}. The variable $\mathbf{y}_t$ contains all the information needed to describe the trajectory without losing any relevant properties.

3. The model. The coordinate increments $\mathbf{y} = \{\mathbf{y}_t\}_{t\in\mathcal{T}}$ are hypothesized to follow a mixture-type model with conditionally independent data following a bivariate normal distribution, also called *emission-distribution* in this context, with parameters that depend on the mixture component; each component of the mixture represents a different kind of behavior. The clustering is encoded into a discrete (latent) random variable $\mathbf{z} = \{z_t\}_{t\in\mathcal{T}}$, a membership variable that takes values in $\{1, 2, \dots, K\} \equiv \mathcal{K}$, where $z_t = k$ indicates that the behavior at time t is the one identified as k. The latent variable z_t is distributed according to a discrete distribution with probabilities $\boldsymbol{\pi}_t = \{\pi_{t,1}, \dots, \pi_{t,K}\}$. These probabilities are realizations of a (latent) LogitGP, which is the main contribution of this paper. The latent $\boldsymbol{\pi}_t$ vary over time and are temporally correlated. Their distribution is described separately in Section 4. Schematically, the model is defined in the following way:

$$f(\mathbf{y}|\mathbf{z}, \{\boldsymbol{\xi}_k, \boldsymbol{\Omega}_k\}_{k\in\mathcal{K}}) = \prod_{t\in\mathcal{T}} \prod_{k\in\mathcal{K}} \phi_2(\mathbf{y}_t|\boldsymbol{\xi}_k, \boldsymbol{\Omega}_k)^{\mathbb{I}_k(z_t)}, \tag{2}$$

$$z_t \sim \text{Discrete}(\boldsymbol{\pi}_t),$$
$$\boldsymbol{\pi}(t) \sim \text{LogitGP}\big(\mathbf{A}, \boldsymbol{\mu}(t), \boldsymbol{C}(h)\big), \tag{3}$$

where $\boldsymbol{\xi}_k$ and $\boldsymbol{\Omega}_k$ are the mean vector and covariance matrix of the bivariate normal density $\phi_2(\cdot)$ for component k, $\mathbb{I}_b(a)$ is the indicator function, which assumes a value of one when $a = b$, and zero otherwise. The probabilities $\boldsymbol{\pi}_t = \{\pi_{t,k}\}_{k\in\mathcal{K}}$ are discrete-time observations of an underlying and nonobserved continuous-time process $\boldsymbol{\pi}(t)$, which is described in Section 4. For the moment, it is enough to say that $\boldsymbol{\pi}(t)$ is constructed from a K-dimensional GP with a coregionalization matrix $\mathbf{A}$, a mean function $\boldsymbol{\mu}(t)$ and a vector of correlation functions $\boldsymbol{C}(h)$, where h is a temporal distance.

In practice, given the latent variables z_t (the behavior), the observations $\mathbf{y}_t$ are independent and normally distributed, with parameters that depend on the current behavior (z_t). The model described in (2) assumes that all the observations ($\mathbf{y}_t$) follow the same distribution, which is only reasonable if the temporal distances between consecutive observations are fixed. This is the reason why we consider a time grid with fixed increments, as stated in Section 2. The fixed time difference between the element of $\mathcal{T}$ forces us to estimate a large number of missing observations. Although this increases the computational time, it is necessary to model the two time windows together and compare them.

It should be noted that a bivariate normal distribution on $\mathbf{y}_t$ induces a projected normal distribution on the turning-angle (Wang and Gelfand (2013)), which is one of the most flexible distributions for circular data (see, e.g., Mastrantonio, Jona Lasinio and Gelfand (2016), Mastrantonio, Maruotti and Jona-Lasinio (2015)). Unfortunately, no closed form is available for the step-length.

4. LogitN distribution and LogitGP. Aitchison (1986) proposed the LogitN distribution to model compositional data as an alternative to the Dirichlet distribution. Although it has a closed form, we prefer to introduce it by using its constructive definition.

The vector $\boldsymbol{\pi}_t$ is defined as

$$\pi_{t,k} = \frac{e^{\omega_{t,k}}}{\sum_{j=1}^{K} e^{\omega_{t,j}}}, \quad k \in 1, \dots, K, \tag{4}$$

where $\omega_{t,k}$ are real valued variables. It should be noted that adding a constant to each $\omega_{t,k}$ produces the same vector of probabilities, and an identifiability constraint is therefore needed; without loss of generality, the Kth element is set to zero ($\omega_{t,K} = 0$) and is treated as the *reference element*. If $\boldsymbol{\omega}_t = \{\omega_{t,k}\}_{k=1}^{K-1}$ is normally distributed, then $\boldsymbol{\pi}_t$ is said to be LogitN distributed (Aitchison (1986)). It is possible to introduce dependence between the compositional vectors, while preserving the LogitN as the marginal distribution, by envisioning $\boldsymbol{\omega}_t$ as a realization of a $K-1$ dimensional GP $\boldsymbol{\omega}(t)$.

Several proposals, that directly define all the elements of ω_t as a $(K-1)$-variate GP exist (see, e.g., Martins, Bonat and Ribeiro (2016), Paci and Finazzi (2018), Pawlowsky and Burger (1992), Tjelmeland and Lund (2003)). Unlike those proposals, we introduce an auxiliary K-dimensional GP $\boldsymbol{\gamma}(t)$, with coregionalization matrix $\mathbf{A}$ which we require to be nonnegative definite and symmetric, with mean function $\boldsymbol{\mu}(t)$ and a vector of correlation functions $\boldsymbol{C}(h)$. From $\boldsymbol{\gamma}(t)$, we construct $\boldsymbol{\omega}(t)$ as

$$\omega_k(t) = \gamma_k(t) - \gamma_K(t), \tag{5}$$

$$\boldsymbol{\gamma}(t) = \boldsymbol{\mu}(t) + \mathbf{A}\boldsymbol{\gamma}^*(t), \tag{6}$$

$$\gamma_k^*(t) \sim \mathrm{GP}\big(0, C_k(h)\big). \tag{7}$$

Matrix $\mathbf{A}$ introduces dependence between the elements of $\boldsymbol{\gamma}(t)$, since $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ is the covariance of $\boldsymbol{\gamma}(t)$. The explicit relation between $\mathbf{A}$ and $\boldsymbol{\Sigma}$ is

$$\mathbf{A} = \boldsymbol{\Delta}\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Delta}', \tag{8}$$

where $\boldsymbol{\Delta}$ and $\boldsymbol{\Xi}$ are the matrix of the eigenvectors of $\boldsymbol{\Sigma}$ and the diagonal matrix of the eigenvalues, respectively. The above definition makes $\mathbf{A}$ the unique symmetric and nonnegative square root of $\boldsymbol{\Sigma}$. By stating that

$$\boldsymbol{\pi}(t) \sim \mathrm{LogitGP}\big(\mathbf{A}, \boldsymbol{\mu}(t), \boldsymbol{C}(h)\big), \tag{9}$$

as in equation (3), we mean that $\boldsymbol{\pi}(t)$ is constructed according to equations (4)–(8). We use this approach to avoid specifying a predefined functional form of $\pi(t)$ and to estimate $\pi(t)$ nonparametrically, using a GP, as is generally done for Bayesian nonparametric inference; see, for example, Muller (2017).

Equations (4)–(7) introduce a serial correlation between vectors $\boldsymbol{\pi}_t$, which is required if the aim is to make a behavior persist over time. It should be noted that $\boldsymbol{\omega}(t)$ is a linear combination of GPs and it is therefore a GP itself.

Equation (5) implicitly defines the reference element as the Kth since, by construction, $\omega_K(t) = \gamma_K(t) - \gamma_K(t) = 0$. From (5), it is also possible to see that the mean function $\boldsymbol{\mu}(t)$ is not identifiable, unless an identification constraint is introduced, and we set the component of $\boldsymbol{\mu}(t)$ relative to the reference element to zero: $\mu_K(t) = 0$.

It is important to highlight that $\boldsymbol{\gamma}(t)$ is not identifiable and any inference about $\boldsymbol{\pi}_t$ is in fact made by looking at $\boldsymbol{\omega}_t$ through equation (4). The advantage of introducing $\boldsymbol{\gamma}(t)$ can be seen by replacing (5) in (4):

$$\pi_{t,k} = \frac{e^{\gamma_{t,k}-\gamma_{t,K}}}{\sum_{j=1}^{K} e^{\gamma_{t,j}-\gamma_{t,K}}} = \frac{e^{\gamma_{t,k}}}{\sum_{j=1}^{K} e^{\gamma_{t,j}}}, \quad k \in 1, \dots, K. \tag{10}$$

With respect to (4), where $\omega_{t,K}$ must be set to zero, the right hand side of (10) has a more symmetric form, since all the components of $\boldsymbol{\pi}_t$ are written in terms of exponentials of $\gamma_k(t)$ and there is no reference element.

4.1. *Invariances*. A desirable property of a general model is that it should not depend on any arbitrary choice. In particular, it is desirable to have:

- invariance from the choice of the reference element;
- invariance with respect to the reordering of the labels.

In order to demonstrate the first property, it is sufficient to observe that the right hand side of equation (10) does not depend on the choice of the reference element.

The second property means that there exists a reparameterization of the model that maintains the likelihood invariant with respect to the reordering of the labels. It is well known that, when labels in $\boldsymbol{\pi}(t)$ are permuted, the likelihood function remains invariant, if the parameters are permuted accordingly. It is therefore sufficient to show that, when (9) holds and if $\mathbf{P}$ is a permutation matrix,

$$\mathbf{P}\boldsymbol{\pi}(t) \sim \mathrm{LogitGP}\big(\mathbf{PAP}', \mathbf{P}\boldsymbol{\mu}(t), \mathbf{P}\boldsymbol{C}(h)\big).$$

In order to understand this, let us consider the vector of the independent GPs $\tilde{\boldsymbol{\gamma}}^*(t)$ with the correlation vector $\mathbf{P}\boldsymbol{C}(h)$. The new vector $\tilde{\boldsymbol{\gamma}}^*(t)$ has the same distribution as the reordered vector $\mathbf{P}\boldsymbol{\gamma}^*(t)$ from equation (6). Moreover, if we apply the coregionalization matrix $\tilde{\mathbf{A}} = \mathbf{PAP}'$ to $\tilde{\boldsymbol{\gamma}}^*(t)$ and add the vector of the reordered means $\mathbf{P}\boldsymbol{\mu}(t)$, we obtain $\tilde{\boldsymbol{\gamma}} = \mathbf{P}\boldsymbol{\mu}(t) + \tilde{\mathbf{A}}\tilde{\boldsymbol{\gamma}}^*(t)$, which has the same distribution as $\mathbf{P}\boldsymbol{\gamma}(t)$. It should be pointed out that the covariance matrix of $\mathbf{P}\boldsymbol{\gamma}(t)$ is $\tilde{\boldsymbol{\Sigma}} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}' = \tilde{\mathbf{A}}\tilde{\mathbf{A}}'$, and that $\tilde{\mathbf{A}}$ is symmetric and nonnegative definite. In short, if we construct $\tilde{\boldsymbol{\pi}}(t)$ from $\mathbf{P}\boldsymbol{\gamma}(t)$ according to equation (10), we obtain $\mathbf{P}\boldsymbol{\pi}(t)$.

Finally, it should also be pointed out that the standard choice for a coregionalization matrix in geostatistics, that is, the Cholesky decomposition, suffers from some drawbacks in this setting: it is in fact well know that it introduces an ordering between the processes (Rothman, Levina and Zhu (2010)). The triangular structure of coregionalization matrix $\mathbf{A}$ is not preserved under a (nondiagonal) permutation matrix $\mathbf{P}$: if $\mathbf{A}$ is triangular, $\mathbf{PAP}'$ is no longer triangular, and it cannot therefore be the coregionalization matrix of the process $\mathbf{P}\boldsymbol{\gamma}(t)$.

4.2. *The log-ratio mean and variance*. The nature of compositional vectors makes the interpretability of the moments difficult. In fact, the sum-to-one constraint restricts the domain to the simplex and induces negative correlations among the variables (see Aitchison (1986)). More consistent definitions of the moments, based on the log-ratios between the components, were proposed in Aitchison (1986). These definitions are particularly easy to compute under our model specification:

$$\rho_{i,k}(t,t') = \mathbb{E}\left(\log\frac{\pi_{t,i}}{\pi_{t',k}}\right) = \mathbb{E}(\gamma_{t,i} - \gamma_{t',k}), \tag{11}$$

$$\begin{aligned}\tau_{ij,k}(t,t') &= \mathbb{C}\mathrm{ov}\left(\log\frac{\pi_{t,i}}{\pi_{t,k}}, \log\frac{\pi_{t',j}}{\pi_{t',k}}\right)\\ &= \mathbb{C}\mathrm{ov}(\gamma_{t,i} - \gamma_{t,k}, \gamma_{t',j} - \gamma_{t',k}).\end{aligned} \tag{12}$$

Let A_{ij} be the (i, j)-th element of matrix $\mathbf{A}$, it is then possible to write (11) and (12) in closed form:

$$\begin{aligned}\rho_{i,k}(t,t') &= \mu_{t,i} - \mu_{t',k},\\ \tau_{ij,k}(t,t') &= A_{ii}C_i(|t'-t|)(A_{ji}+A_{ki})\\ &\quad + A_{jj}C_j(|t'-t|)(A_{ij}+A_{kj})\\ &\quad + A_{kk}C_k(|t'-t|)(2A_{kk}-A_{ik}-A_{jk}).\end{aligned} \tag{13}$$

Important properties of the compositional vectors $\boldsymbol{\pi}(t)$ can be established by imposing analogous properties on GP $\boldsymbol{\gamma}(t)$. For example, if the components of $\boldsymbol{\gamma}(t)$ are independent ($\mathbf{A}$ is diagonal), then the components of $\boldsymbol{\pi}(t)$ are subcompositionally independent (see Aitchison (1986), Property 10.3), and the log-ratios are uncorrelated (see Aitchison (1986), Section 5.9), since $\tau_{ij,K}(t,t)$ takes the particular expression

$$\tau_{ij,K}(t,t) = \begin{cases} 2A_{ii}^2 + 2A_{KK}^2 & \text{for } i=j \text{ and } i \neq K,\\ 2A_{KK}^2 & \text{for } i \neq j \text{ and } i, j \neq K.\end{cases}$$

Moreover, if the components of vector $\boldsymbol{\gamma}(t)$ are independent and identically distributed, we obtain $A_{ii} = A_{jj}$ and $\rho_{i,j}(t,t) = 0$ for $i, j = 1, \dots, K$, which is the requirement of having $\boldsymbol{\pi}(t)$ subcompositionally independent and with elements identically distributed (Aitchison (1986)). The latter property can be seen immediately from (10).

4.3. *Relationship with other definitions of the* LogitGP. A Bayesian nonparametric approach, based on a GP, was also adopted in previous works. Most of the the existing constructions are derived from equation (4), where a GP has been used to directly model $\omega_k(t)$. As already mentioned, identifiability requires that a reference element, here the Kth, is chosen. The corresponding component $\omega_{t,K}$ is set to zero, and its definition is hence deterministic and the covariance function of the reference element therefore vanishes. Unless special care is taken, changing the reference element can affect the structure of the log-ratio covariance of $\pi(t)$, measured by (13). In fact, if any of the indices i, j, k is equal to K in equation (13), only two correlation functions appear in the expression instead of three. To compensate for this effect, and to keep the expression of $\tau_{ij,k}$ invariant, it would be necessary to correspondingly adjust the other correlation functions in a nontrivial way, or to make some other ad-hoc assumptions. Our method provides a way of keeping this invariance in a natural way and in a very general framework.

In Paci and Finazzi (2018), the GPs $\omega_k(t)$ with $k \in \{1 \cdots K-1\}$, are assumed to be independent. However a different choice of the reference element is equivalent to using another set of GP $\boldsymbol{\omega}'(t) = \mathbf{L}\boldsymbol{\omega}(t)$ for a suitable nondiagonal matrix $\mathbf{L}$ (Aitchison (1986)). The new processes $\omega'_k(t)$ are correlated and no longer satisfy the same assumptions. On the other hand, in Tjelmeland and Lund (2003) and

in Martins, Bonat and Ribeiro (2016), the problem is circumvented by additionally assuming that there is a single correlation function that is common to all the components. This makes the model invariant, but at the price of introducing a restrictive constraint. Only GPs of the VARMA family are considered in Brunsdon and Smith (1998). Changing the reference element is again equivalent to introducing the transformation $\boldsymbol{\omega}'(t) = \mathbf{L}\boldsymbol{\omega}(t)$ mentioned above. Such a transformation preserves the fact that the new processes are in the VARMA family. Nevertheless it does not preserve the structure of the log-ratio covariances and, therefore, it does not make the resulting inference about the $\pi(t)$ invariant. A completely different proposal, which is not based on a coregionalization but on the co-kriging approach, can be found in Pawlowsky and Burger (1992) and in Lark and Bishop (2007).

4.4. *Differences from and similarities with the CT-HMM.* One of the main competitors of the here proposed model is the CT-HMM of Blackwell et al. (2016). However, the two approaches are somewhat different. To make the comparison straightforward, we still consider equation (2) valid, but interpret the membership variable z_t as a discretely observed continuous-time Markov chain $z(t)$, whose transition rates is $\lambda_{ij}(t)$. In Blackwell et al. (2016), the number of behaviors and the transition rates are established in advance, and are considered as part of the model. The transition rates between behaviors depend on some parameters that are inferred from data. For example, a simplified version of the wild boar example in Blackwell et al. (2016) assumes that three behaviors exist: resting, foraging and returning. The animal repeats these behaviors each day in the given fixed sequence. There exists a preferred time for each transition, making the Markov chain nonhomogeneous. The transition rates follow a parametric function of time that needs to be specified in advance. We are in the framework of parametric inference.

On the other hand, in our approach, the framework is nonparametric and there is one more level in the hierarchical structure. Given a realization of $\pi(t)$, the z_t are independent of each other. Since we do not want to give a predefined functional form to $\pi(t)$, we interpret it as a realization of an unobserved LogitGP, in analogy with what is done, for example, in nonparametric regressions (see Muller (2017), Chapter 4.3.3). We then let Bayesian inference reveal the a posteriori distribution of $\pi(t)$.

5. Wolf data. We now present the application of the proposed method to the dataset described in Section 2.

In order to fully specify the model, we assume that each μ_k is a piecewise constant function whose value is $\mu_k(t) = \beta_{k,1} + \chi_2(t)\beta_{k,2}$, where χ_2 is a function that takes the value 0 if t belongs to the first time window and 1 otherwise. Moreover, we assume that the correlation functions C_k are all exponential: $C_k(h) = \mathrm{e}^{-\psi_k|h|}$. Under this assumption, the GP is Markovian and this makes computation much faster. Other choices are possible, but since the GP has to be evaluated over a large number of temporal points, a geostatistical approximation is needed to avoid the

so-called "big N problem" (Jona Lasinio, Mastrantonio and Pollice (2013)). For a general review of the methods that can be exploited, see Heaton et al. (2019).

The model parameters are therefore $(\boldsymbol{\Sigma}, \{\beta_{k,1}, \beta_{k,2}\}_{k=1}^{K-1}, \{\boldsymbol{\xi}_k, \boldsymbol{\Omega}_k, \psi_k\}_{k=1}^{K})$, where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$, and a prior distribution must be specified for each of them. We assume independence between the parameters: $\boldsymbol{\xi}_k \sim N_2(\mathbf{0}_2, 100\mathbf{I}_2)$ and $\boldsymbol{\Omega}_k \sim \mathrm{IW}(3, \mathbf{I}_2)$ are used for the likelihood parameters, $\psi_k \sim U(0.3, 6)$ for the temporal decays, $\beta_{k,j} \sim N(0, 100)$, $j = 1, 2$, for the regression coefficients, and $\boldsymbol{\Sigma} \sim \mathrm{IW}(K + 1, \mathbf{I}_K)$ for the variance parameter of the LogitN process, where $\mathrm{IW}(b, \mathbf{B})$ stands for the inverse Wishart distribution with b degrees of freedom and scale matrix $\mathbf{B}$.

We compare our proposal with models based on the same likelihood but with different dynamics for z_t, in other words with a DT-HMM and a CT-HMM. We follow Blackwell et al. (2016) for the CT-HMM, and the rate function from state i to j is defined as follows:

$$\lambda_{i,j} = \frac{v_{i,j}}{1 + e^{v_{i,j}(t \bmod 1 - t_{i,j}^0)}},$$

while the probability of switching from the ith state to the jth, in the DT-HMM, is given by $\pi_{i,j}$. As prior distributions, we use a $U(0, 1)$ for all the $t_{i,j}^0$, while, following Blackwell (2018), we set $v_{kj}/48 \sim \mathrm{Beta}(1, 1)$, where 48 is the number of time points in a day. The vector $(\pi_{i,1}, \dots \pi_{i,K})'$ follows a Dirichlet distribution, with all the parameters equal to 1.

We test $K \in \{2, \dots, 6\}$ for all the models. The choice of the number of components is essential in applications involving mixtures. Since the models are defined via a latent variable z_t, we use the ICL, proposed by Biernacki, Celeux and Govaert (2000). For a thorough review of the methods that can be used to perform model selection, the reader may refer to Pohle et al. (2017). The MCMC is implemented with 1,000,000 iterations, burnin 70,000 and thinning equal to 6, which results in 5000 posterior samples. The models are implemented on the Bari ReCaS Data Center (Antonacci et al. (2017)).

5.1. *The results*. Table 2 shows the ICL for all the tested models. The results of each model suggest the presence of three different behaviors ($K = 3$), and our

TABLE 2
ICLs of the tested models. The best model is in bold. Numbers are rounded to the closest integer

Model	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$
HMM	−12,527	−14,926	−13,984	−11,941	−10,312
CT- HMM	−13,511	−15,034	−14,871	−12,951	−10,801
LogitGP	−13,831	**−15,925**	−14,129	−12,010	−10,210

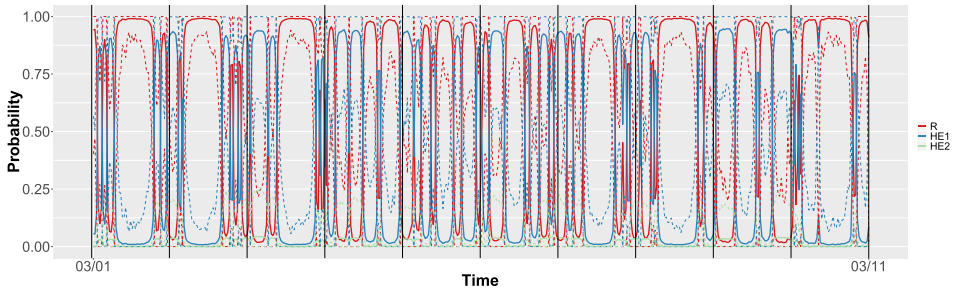
TABLE 3
Real data—Posterior means and 95% CIs of the parameters of the proposed model with $K = 3$

	$k = 1$	$k = 2$	$k = 3$
$[\boldsymbol{\xi}_k]_1$	−0.014	0.386	−0.543
(CI)	(−0.017 −0.004)	(0.251 0.531)	(−1.411 0.081)
$[\boldsymbol{\xi}_k]_2$	0.001	0.061	0.048
(CI)	(−0.005 0.008)	(−0.070 0.121)	(−1.291 1.235)
$[\boldsymbol{\Omega}_k]_{11}$	0.003	0.661	0.38
(CI)	(0.003 0.004)	(0.533 0.819)	(0.176 0.635)
$[\boldsymbol{\Omega}_k]_{12}$	0.000	0.052	0.001
(CI)	(0 0)	(−0.044 0.153)	(−0.032 0.030)
$[\boldsymbol{\Omega}_k]_{22}$	0.002	0.513	0.381
(CI)	(0.002 0.004)	(0.403 0.637)	(0.169 0.639)
β_{k1}	7.395	4.252	
(CI)	(6.764 8.94)	(1.848 6.932)	()
β_{k2}	−0.382	−7.476	
(CI)	(−0.879 0.153)	(−11.352 −4.796)	()
$\boldsymbol{\psi}_k$	6.746	12.276	5.221
(CI)	(4.887 9.187)	(6.135 14.862)	(0.909 11.179)
$[\boldsymbol{\Sigma}]_{kk}$	143.95	1.785	1.731
(CI)	(40.475 371.745)	(0.221 8.715)	(0.027 11.549)
	$[\boldsymbol{\Sigma}]_{12}$	$[\boldsymbol{\Sigma}]_{13}$	$[\boldsymbol{\Sigma}]_{23}$
	−4.418	−5.877	1.408
(CI)	(−38.881 19.974)	(−43.148 13.200)	(0.146 9.337)

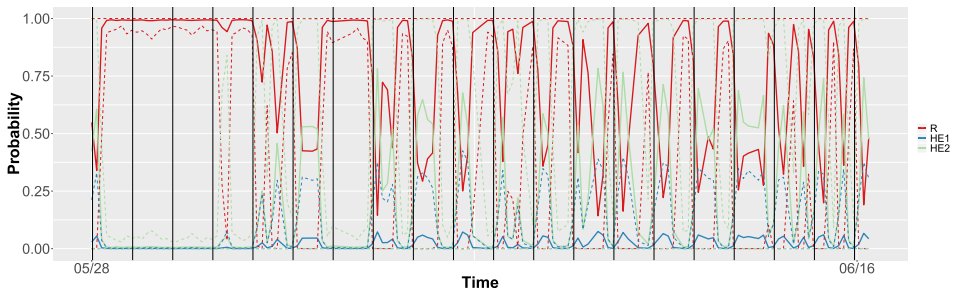
proposal is the one with the lowest ICL (the model with the best goodness-of-fit). The posterior estimates, means and 95% credible intervals (CIs) of the chosen model can be seen in Table 3, while Figure 3 shows the posterior estimates of the probability vector time series. Figure 4 shows the observed spatial locations with the associated classification and predictive densities of the step-length and turning-angle. As in the Introduction, we indicate the first behavior with R, the second one with HE1 and the third one with HR2.

5.2. *Behavior description.*

First behavior (*R*). From Figure 4(c) and (d), it is possible to observe that the speed in the first behavior is very close to zero, and the circular distribution, even though it has a mode at around π, has a high variability, thus showing that there is not a clear preferred direction. The means of the associated GP in both windows (Table 3) are larger then the ones of the other behaviors, thus indicating that it is the one with the highest (mean) probability, as it is confirmed in Figure 3, where it is evident that the probability values are often equal to 1. This behavior may be



(a) First time window



(b) Second time window

FIG. 3. *Posterior estimates of the probability vector time series. The solid lines represent the posterior means, while the dotted lines are the limits of the* 95% *CIs. The vertical lines indicate* 00:00 *hours.*

described as a slow-movement behavior, which represents a variety of activities, such as resting, feeding, social interacting and, in the second window, attending cubs during the reproduction period. It should be noted that the occurrences of this behavior, in the second time window, are spatially localized over a relatively small area, which can reasonable be identified as the den; see Figure 4(b).

Second behavior (*HE*1). As far as the second behavior is concerned, the speed increases and the circular distribution has a well-defined mode around zero, thus indicating that the wolf tends to move in a straight line; see Figure 4. This behavior is relevant in the first time window while it almost disappears in the second, as can be seen from the associated regressive coefficients. This behavior represents the nomadic phase of wolf movement patterns during winter, when the main activities are hunting and patrolling the territory (Mech and Boitani (2003)). Moreover, F24 established her home range in March 2010, and these high speeds may also represent the need to control and mark the territory, as newly formed pairs are the ones with the highest marking rates in wolf populations (Rothman and Mech (1979)).

Third behavior (*HE*2). As far as the third behavior is concerned, the predictive distribution of the step-length is similar to that of the second behavior (Fig-

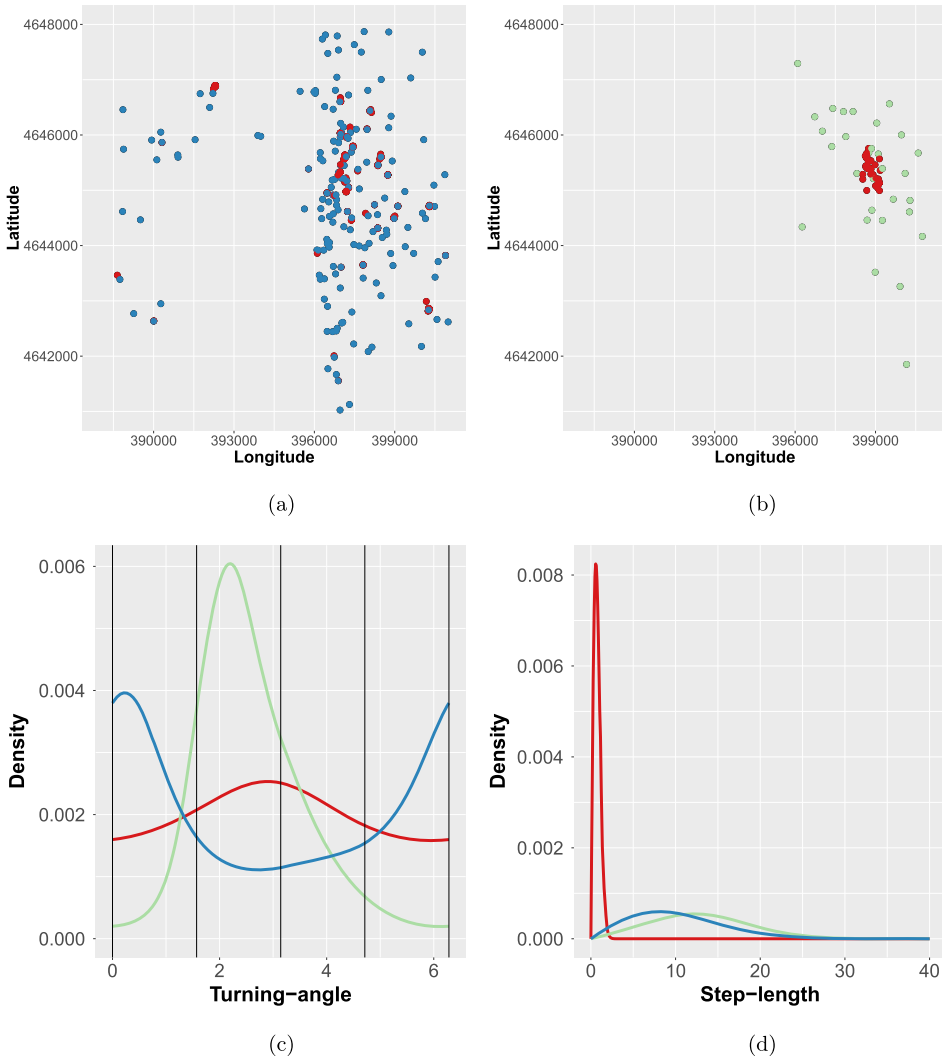


FIG. 4. *The observed spatial locations in the first* (*a*) *and second* (*b*) *time windows*, *predictive distributions of the turning-angle* (*c*) *and the step-length* (*d*) *for the three different behaviors*. *The behaviors are encoded with the same colors as in Figure* 3.

ure 4(d)). The turning-angle has a mode at approximately 2 and limited variability, thus indicating that the animal moves in an anticlockwise direction. This behavior is almost absent in the first temporal window, while it represents the main movement type in the second one. This pattern seems to correctly estimate the star-shaped movements of wolves in the presence of cubs at dens (Mech and Boitani (2003)). This is in line with the tendency of breeding females to restrict their movements to a smaller portion of the territory during the period of reproduction,

compared to other times of the year (Jedrzejewski et al. (2001)). The counterclockwise tendency may be related to the necessity of exploiting different portions of the home range to locate vulnerable prey. Since wolves have a spatial map of resources within their territory (Peters (1979)), varying their hunting routes to surprise prey could improve their hunting success (Jedrzejewski et al. (2001)), and this could result in a rotational use of the home range (Demma and Mech (2011)).

It is possible to see from the off-diagonal elements of $\mathbf{\Sigma}$ (Table 3) that there could be dependence; it should be recalled that independence between the elements of the compositional vectors requires a diagonal $\mathbf{\Sigma}$. To better analyze the dependence structure we look at the temporal evolution of the "correlation". There is no unique or generally accepted way of evaluating the temporal correlation of compositional data (see, e.g., Filzmoser and Hron (2009) or Long and Wang (2013)), but since we are here mainly interested in the temporal evolution of the dependence, the log-ratio covariance, equation (12), is divided by its value when $t = t'$ (at lag 0), thus showing how dependence changes over time. These log-ratio correlations, indicated with ν, are shown in Figure 5.

It is interesting to note that $\nu_{11,2}$ and $\nu_{11,3}$, being the correlation functions of the second and third behaviors with respect to the first one, are indistinguishable, thus highlighting that the difference is mostly due to the direction of movement (the turning-angle). In Figure 5(b), which shows the cross-correlations, only $\nu_{23,1}$ has 95% CIs that does not contain zero. Moreover, dependence almost disappears for a lag of 12 hours (all the correlation values are close to zero) and it is numerically zero after 24 hours. From a biological point of view, this means that the probability of following a particular behavior at time t is influenced to a great extent by the probabilities and behaviors of the 12 previous hours, is slightly influenced by the 12 to 24 previous hours, but what the animal did the day before shows no influence.

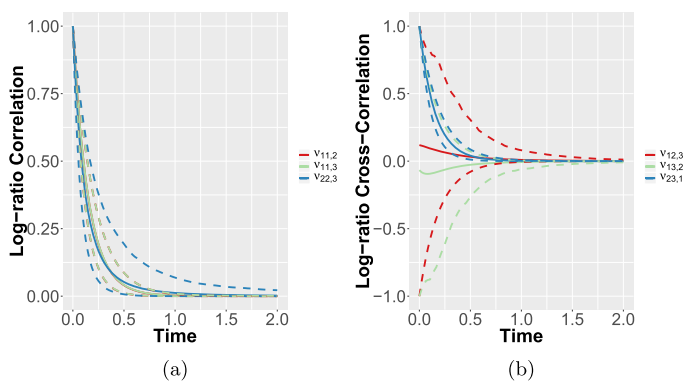


FIG. 5. *Plots of the log-ratios as a function of the temporal distance (expressed in days). The solid lines represent the posterior means, while the dotted lines are the limits of the 95% CIs.*

5.3. *Time window description.*

First time window. In the first window, the slow-movement behavior (R) has a high probability of occurring during daylight hours, whereas the HR1 seems to be more likely during the night (Figure 3(a)). This complementary pattern is in line with the circadian activity of wolves in human-modified environments, where they are mainly nocturnal to avoid disturbances from human activities during the day (Ciucci et al. (1997), Theuerkauf (2009)).

Second time window. In the second window, the slow-movement regime has a probability close to one during the first days, because F24 likely entered the den and gave birth during that time. According to previous research on wolf reproducing behavior, breeding females are stationary on the day of birth, and with limited movements during the following period (Alfredeén (2006)). During the days after reproduction occurred, the slow-movement regime is often concentrated around dusk and during the first hours of the night. Instead, the star-shaped moving regime is concentrated during daylight hours, and shows two or more peaks at different times of the day (Figure 3(b)). This result can be interpreted as a reduction in the nocturnal activity of this wolf due to the presence of cubs, which is also accompanied by a relative increase in diurnality. During the reproduction period, breeding females spend most of the time at their den and rendezvous sites (Ayres, Gardner and Foster (1991), Harrington and Mech (1982)). Since other wolves from the pack usually ensure the feeding of breeding females during this time (Mech, Wolf and Packard (1999)), females do not have to maintain an activity pattern based on hunting which, in our study area, may be nocturnal due to the presence of humans. This situation may have led F24 to mainly leave the den during the day, when sunlight can help keep the unattended cubs warm, and other large carnivores (such as Apennine brown bears) are less active (Vila (1995)).

It should be noted that, since the temporal evolution of $\boldsymbol{\pi}$ is modeled nonparametrically, we found the daily-temporal patterns in Figure 3 without introducing any constraint into our model or introducing factors that can be used to model it directly.

6. Final remark. Motivated by our dataset, in which a female wolf is observed over two time windows, we have proposed a novel approach to analyze trajectory tracking data. This approach is aimed at defining the posterior distribution of the clustering probabilities, where the clusters are representative of different behaviors that the animal exhibits, and at describing the trajectory conditionally on the particular behavior by characterizing the associated step-length and turning-angle.

Our model is based on a Bayesian nonparametric approach, where the time varying probabilities $\boldsymbol{\pi}(t)$ are modeled through a LogitGP. Our model is invariant with respect to the arbitrary choices of the reference element and the reordering of

the labels. We have proposed fitting the model under a Bayesian framework and the number of latent behaviors through ICL, which is one of the most frequently used information criterion.

We fitted the model on the wolf data. The results we obtained are easy to interpret and give insight into wolf behavior, both in terms of movement metrics (in particular, step-length and turning-angle) and the time evolution.

The advantage of a nonparametric approach is that even though some features of the data are not modeled explicitly, posterior inference may reveal them afterwards. An example is the attractive effect of the unknown location of the den. The distribution of the turning-angle in HE2 has a peak close to π, which may be in fact interpretable as the effect of such an attraction (Parton and Blackwell (2017)). A second example is the seasonality of the components of $\boldsymbol{\pi}(t)$ which is discussed in detail in Section 5.

The model we have proposed was motivated by a trajectory tracking dataset, but it could be employed in different contexts, in particular in the case of environmental sciences, where spatial information can be incorporated in the probability dependence structure, so that the response variable behaves in a similar way in locations that are close in space. This aspect will be investigated in the future.

Another issue that deserves further attention is the estimation of the unknown number of clusters. For the moment, we decided to fit the model with different numbers of components, and then use a model-fit measure to select the *best* one. A fully Bayesian analysis, where a prior distribution is defined for the number of components, would also be very interesting.

Future lines of research include the extension of the present model to multi-animal trajectories. The generalization of the model proposed in this work would be an extension of the hidden Markov models proposed in Langrock et al. (2014), with relation to the continuous-time approach of Niu, Blackwell and Skarin (2016).

APPENDIX A: MCMC IMPLEMENTATION

The MCMC implementation is straightforward. Given a value of the entire multivariate GP, its parameters can be simulated as in the usual GP framework. In details, we update all the parameters at the same time using a Metropolis step, according to algorithm 4 of Andrieu and Thoms (2008). Before applying the algorithm, it is necessary to transform the parameters so that they belong to $\mathbb{R}$. Then, we take the logarithm of the decay parameters and in order to eliminate the constraints over the parameters of the nonnegative definite matrix $\boldsymbol{\Sigma}$, we re-express it using the Bartlett decomposition (Anderson (2003)), which is based on random variables that are normally and chi-squared distributed; the latter is then transformed using the logarithm. Given the probabilities $\boldsymbol{\pi}$ and the data $\mathbf{y}$, the parameters and the latent variable $\mathbf{z}$ are simulated as in a mixture model using Gibbs steps. In order to simulate the GP elements, we use the novel approach of Polson, Scott and

Windle (2013) and its extension, as proposed in Linderman, Johnson and Adams (2015).

When the missing $\mathbf{y}$ are simulated, we have to ensure that their values are "coherent", meaning that $\mathbf{y}$ have to define a trajectory $\mathbf{s}$ that goes through the observed locations. Equation (1) shows that having a realization of $\mathbf{s}$ we can easily compute $\mathbf{y}$. However, it is easier to simulate $\mathbf{s}$ since the animal trajectory follows an ARMA-type model. Therefore, the missing observations are obtained by first simulating the missing $\mathbf{s}$ and then deriving the corresponding $\mathbf{y}$; in this way, the uncertainty implicit in the presence of missing data is automatically incorporated in the inferential method.

All codes are available from the first author, upon request.

APPENDIX B: SIMULATED EXAMPLES

In this section, we show that the MCMC is able to estimate the parameters in a satisfactory way and, moreover, that ICL retrieves the right number of latent classes.

We simulate data with $K = 3$, and three different number of observations $T = 250, 500, 1000$ over a time window of length 20. The time-lag between observations is assumed constant and therefore $t_i - t_{t-1} = 20/T$; in other words, the intervals between the observations decrease as T increases. The parameters are $\boldsymbol{\xi}_1 = (0, 0)'$, $\boldsymbol{\xi}_2 = (3, 0)'$, $\boldsymbol{\xi}_3 = (0, -3)'$ and

$$\boldsymbol{\Omega}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \qquad \boldsymbol{\Omega}_2 = \begin{pmatrix} 1 & 1.272 \\ 1.272 & 2 \end{pmatrix}, \qquad \boldsymbol{\Omega}_3 = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}.$$

We also assume

$$\boldsymbol{\Sigma} = \begin{pmatrix} 5 & -2 & 0 \\ -2 & 5 & 3 \\ 0 & 3 & 5 \end{pmatrix},$$

and use exponential correlation functions with decay parameters equal to 1, 0.8 and 1.5, respectively. We assume $\mu_1(t) = \beta_{10} + \beta_{11} t$ and $\mu_1(t) = \beta_{20} + \beta_{21} t$ with $\beta_{10} = 0$, $\beta_{11} = -5$, $\beta_{20} = 3$, $\beta_{21} = -7$ for the mean function of the GPs.

We simulate 100 datasets for each T, and we perform inference by fixing K to values between 2 and 6. To ensure that three separate clusters can be identified, datasets with a cluster composed of less than 5% of the total sample size are discarded. Model choice is performed using the ICL.

We assume $\boldsymbol{\xi}_k \sim N_2(\mathbf{0}_2, 100\mathbf{I}_2)$ and $\boldsymbol{\Omega}_k \sim \mathrm{IW}(3, \mathbf{I}_2)$ as prior distributions for the likelihood parameters, $U(0.3, 6)$ for the temporal decays, the regression coefficients are normally distributed with a mean of 0 and a variance of 100 while $\boldsymbol{\Sigma} \sim \mathrm{IW}(K+1, \mathbf{I}_K)$. The MCMC is implemented with 1,000,000 iterations, burnin 70,000 and thin 6, which results in 5000 posterior samples.

TABLE 4
Simulated Example—Posterior means and 95% CIs under $K = 3$

	$k = 1$	$k = 2$	$k = 3$
$[\boldsymbol{\xi}_k]_1$	0.043	2.953	−0.234
(CI)	(−0.165 0.277)	(2.745 3.236)	(−0.441 −0.112)
$[\boldsymbol{\xi}_k]_2$	−0.098	−0.124	−2.897
(CI)	(−0.467 0.301)	(−0.364 0.117)	(−3.001 −2.791)
$[\boldsymbol{\Omega}_k]_{11}$	1.446	1.271	2.076
(CI)	(1.074 1.886)	(1.015 1.625)	(1.764 2.427)
$[\boldsymbol{\Omega}_k]_{12}$	−0.004	1.464	−0.539
(CI)	(−0.373 0.356)	(1.151 1.841)	(−0.706 −0.389)
$[\boldsymbol{\Omega}_k]_{22}$	3.338	2.351	0.635
(CI)	(2.518 4.182)	(1.887 2.916)	(0.519 0.771)
β_{k0}	0.580	2.206	
(CI)	(−1.955 3.169)	(0.764 3.721)	()
β_{k1}	−2.508	−5.09	
(CI)	(−6.743 1.463)	(−7.724 −2.649)	()
$\boldsymbol{\psi}_k$	0.987	2.247	2.89
(CI)	(0.563 1.639)	(0.422 5.361)	(0.652 5.208)
$[\boldsymbol{\Sigma}^*]_{kk}$	3.889	0.78	1.979
(CI)	(0.358 8.784)	(0.143 3.579)	(0.549 5.001)
	$[\boldsymbol{\Sigma}]_{12}$	$[\boldsymbol{\Sigma}]_{13}$	$[\boldsymbol{\Sigma}]_{23}$
	−0.118	0.557	0.338
(CI)	(−1.072 1.479)	(−1.292 3.684)	(−0.891 3.150)

The ICL selects $K = 3$ 95% of the times with $T = 250$, 99% with $T = 500$ and 100% if $T = 1000$. In order to obtain a better insight into this result, we randomly selected one dataset with $T = 1000$, and we show the parameter estimates in Table 4, trajectory and $\mathbf{y}$ are shown in Figure 6, while the posterior estimates of the compositional vector time series, with the associated 95% CIs, are depicted, in Figure 7, for the first behavior. The "true" compositional vector time series. From these it is possible to see that most of the true values of the parameters are inside the associated 95% CIs, as well as the true compositional vector time series.

Acknowledgments. The GPS data were collected within the "Large Carnivores" project, which took place from 2006 to 2011 in the Abruzzo, Lazio and Molise National Park. The project was supervised by the Department of Biology and Biotechnology "Charles Darwin" of the University of Rome "La Sapienza" and funded by a private American donor through the "Wildlife Conservation Society".

The authors would like to thank Tilman Gneiting, the Editor, the Associate Editor and the referees for their valuable suggestions.

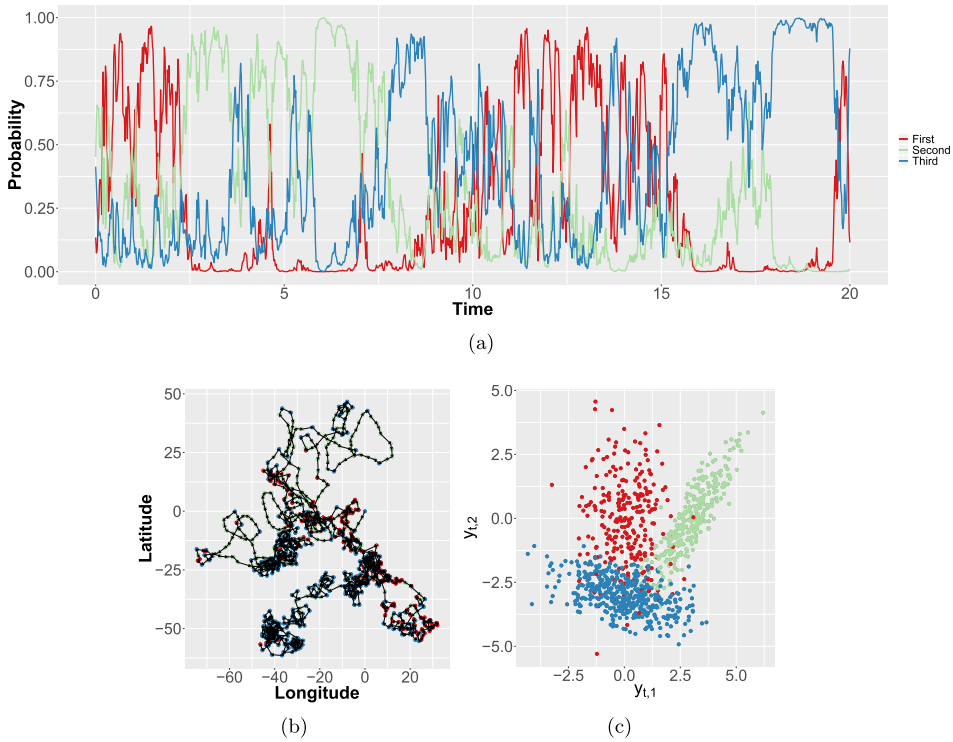


FIG. 6. *Simulated example—Probability vector time series* (*a*), *trajectory* (*b*), *data* **y** (*c*). *The colors represent the different behaviors.*

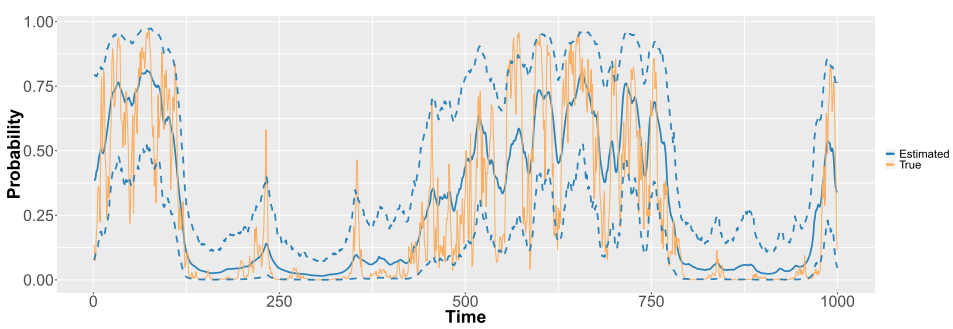


FIG. 7. *Simulated example—Posterior estimates of the probability vector time series of the first behavior. The solid lines represent the posterior means while the dotted lines are the limits of the 95% CIs.*

REFERENCES

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0865647

ALFREDEÉN, A. C. (2006). Denning behavior and movement pattern during summer of wolves, Canis lupus. Grimsö Institute of Wildlife Biology.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

ANDERSON, C. R. and LINDZEY, F. G. (2003). Estimating cougar predation rates from GPS location clusters. *J. Wildl. Manag*. **67** 307–316.

ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput*. **18** 343–373. MR2461882

ANTONACCI, M., BELLOTTI, R., CAFAGNA, F., DE PALMA, M., DIACONO, D., DONVITO, G., ITALIANO, A., GERVASONI, R., MAGGI, G. et al. (2017). The ReCaS Project: The Bari Infrastructure. In *High Performance Scientific Computing Using Distributed Infrastructures* 17–33 Chapter 3.

AYRES, L. A., GARDNER, C. L. and FOSTER, J. W. (1991). Den site activity patterns of gray wolves, Canis lupus, in Southcentral Alaska. *Can. Field-Nat*. **105** 497–504.

BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated classification likelihood. *IEEE Trans. Pattern Anal. Mach. Intell*. **22** 719–725. DOI:10.1109/34.865189.

BLACKWELL, P. G. (1997). Random diffusion models for animal movement. *Ecol. Model*. **100** 87–102.

BLACKWELL, P. G. (2003). Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika* **90** 613–627. MR2006839

BLACKWELL, P. G. (2018). Integrated continuous-time hidden Markov models. arXiv preprint arXiv:1807.11907.

BLACKWELL, P. G., NIU, M., LAMBERT, M. S. and LAPOINT, S. D. (2016). Exact Bayesian inference for animal movement in continuous time. *Methods Ecol. Evol*. **7** 184–195.

BRILLINGER, D. R. (2010). Modeling spatial trajectories. In *Handbook of Spatial Statistics*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 463–475. CRC Press, Boca Raton, FL. MR2730961

BROST, B. M., HOOTEN, M. B., HANKS, E. M. and SMALL, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. *Ecology* **96** 2590–2597.

BROWNLEE, J. (1911). The mathematical theory of random migration and epidemic distribution. *Proc. Roy. Soc. Edinburgh* **31** 262–289. DOI:10.1017/S0370164600025116.

BRUNSDON, T. M. and SMITH, T. M. F. (1998). The time series analysis of compositional data. *J. Off. Stat*. **14** 237–253.

CAGNACCI, F., BOITANI, L., POWELL, R. A. and BOYCE, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *Philos. Trans. R. Soc. Lond. B, Biol. Sci*. **365** 2157–2162. DOI:10.1098/rstb.2010.0107.

CALABRESE, J. M., FLEMING, C. H. and GURARIE, E. (2016). ctmm: An R package for analyzing animal relocation data as a continuous-time stochastic process. *Methods Ecol. Evol*. **7** 1124–1132.

CIUCCI, P., BOITANI, L., FRANCISCI, F. and ANDREOLI, G. (1997). Home range, activity and movements of a wolf pack in central Italy. *J. Zool*. **243** 803–819. DOI:10.1111/j.1469-7998.1997.tb01977.x.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. Revised reprint of the 1991 edition. MR1239641

DEMMA, D. J. and MECH, L. D. (2011). Accuracy of estimating wolf summer territories by daytime locations. *Am. Midl. Nat*. **165** 436–445. DOI:10.1674/0003-0031-165.2.436.

FILZMOSER, P. and HRON, K. (2009). Correlation analysis for compositional data. *Math. Geosci.* **41** 905–919. MR2557238

FLEMING, C. H. and CALABRESE, J. M. (2019). ctmm: Continuous-time movement modeling. R package version 0.5.3.

FLEMING, C. H., CALABRESE, J. M., MUELLER, T., OLSON, K. A., LEIMGRUBER, P. and FAGAN, W. F. (2014a). From fine-scale foraging to home ranges: A semivariance approach to identifying movement modes across spatiotemporal scales. *Amer. Nat.* **183** E154–E167.

FLEMING, C. H., CALABRESE, J. M., MUELLER, T., OLSON, K. A., LEIMGRUBER, P. and FAGAN, W. F. (2014b). Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods Ecol. Evol.* **5** 462–472.

FRAIR, J. L., FIEBERG, J., HEBBLEWHITE, M., CAGNACCI, F., DECESARE, N. J. and PEDROTTI, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **365** 2187–2200. DOI:10.1098/rstb.2010.0084.

FRANKE, A., CAELLI, T., KUZYK, G. and HUDSON, R. J. (2006). Prediction of wolf (Canis lupus) kill-sites using hidden Markov models. *Ecol. Model.* **197** 237–246.

GELFAND, A. E., DIGGLE, P. J., FUENTES, M. and GUTTORP, P., eds. (2010). *Handbook of Spatial Statistics. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR2761512

GURARIE, E., FLEMING, C. H., FAGAN, W. F., LAIDRE, K. L., HERNÁNDEZ-PLIEGO, J. and OVASKAINEN, O. (2017). Correlated velocity models as a fundamental unit of animal movement: Synthesis and applications. *Movement Ecology* **5** 13.

HARRINGTON, F. H. and MECH, L. D. (1982). An analysis of howling response parameters useful for wolf pack censusing. *J. Wildl. Manag.* **46** 686–693.

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for Analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425. MR3996451

HEBBLEWHITE, M. and MERRILL, E. (2008). Modelling wildlife-human relationships for social species with mixed-effects resource selection models. *J. Appl. Ecol.* **45** 834–844.

HOOTEN, M. B. and JOHNSON, D. S. (2017). Basis function models for animal movement. *J. Amer. Statist. Assoc.* **112** 578–589. MR3671753

HOOTEN, M. B., JOHNSON, D. S., MCCLINTOCK, B. T. and MORALES, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, FL.

HOOTEN, M. B., SCHARF, H. R., HEFLEY, T. J., PEARSE, A. T. and WEEGMAN, M. D. (2018). Animal movement models for migratory individuals and groups. *Methods Ecol. Evol.* **9** 1692–1705.

HOUSTON, A. I. and MCNAMARA, J. M. (1999). *Models of Adaptive Behaviour: An Approach Based on State*. Cambridge Univ. Press, Cambridge.

JEDRZEJEWSKI, W., SCHMIDT, K., THEUERKAUF, J., JEDRZEJEWSKA, B. and OKARMA, H. (2001). Daily movements and territory use by radio-collared wolves (Canis lupus) in Bialowieza Primeval Forest in Poland. *Can. J. Zool.* **79** 1993–2004. DOI:10.1139/z01-147.

JOHNSON, D. S., HOOTEN, M. B. and KUHN, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *J. Anim. Ecol.* **82** 1155–1164.

JOHNSON, D. S., LONDON, J. M., LEA, M.-A. and DURBAN, J. W. (2008a). Continuous-time correlated random walk model for animal telemetry data. *Ecology* **89** 1208–1215.

JOHNSON, D. S., THOMAS, D. L., VER HOEF, J. M. and CHRIST, A. (2008b). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics* **64** 968–976. MR2571498

JONA LASINIO, G., MASTRANTONIO, G. and POLLICE, A. (2013). Discussing the "big *n* problem". *Stat. Methods Appl.* **22** 97–112. MR3038184

JONSEN, I. D., FLEMMING, J. M. and MYERS, R. A. (2005). Robust state space modeling of animal movement data. *Ecology* **86** 2874–2880.

LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. and MORALES, J. M. (2012). Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions. *Ecology* **93** 2336–2342. DOI:10.1890/11-2241.1.

LANGROCK, R., HOPCRAFT, J. G. C., BLACKWELL, P. G., GOODALL, V., KING, R., NIU, M., PATTERSON, T. A., PEDERSEN, M. W., SKARIN, A. et al. (2014). Modelling group dynamic animal movement. *Methods Ecol. Evol.* **5** 190–199.

LARK, R. M. and BISHOP, T. F. A. (2007). Cokriging particle size fractions of the soil. *Eur. J. Soil Sci.* **58** 763–774.

LINDERMAN, S. W., JOHNSON, M. J. and ADAMS, R. P. (2015). Dependent multinomial models made easy: Stick breaking with the Pólya–gamma augmentation. In *Proceedings of the* 28*th International Conference on Neural Information Processing Systems* 3456–3464.

LONG, W. and WANG, Q. (2013). Two methods of correlation coefficient on compositional data. *Proc. Comput. Sci.* **18** 1757–1763. 2013 International Conference on Computational Science.

MANCINELLI, S., BOITANI, L. and CIUCCI, P. (2018). Determinants of home range size and space use patterns in a protected wolf (Canis lupus) population in the central Apennines, Italy. *Can. J. Zool.* **96** 828–838.

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. Revised reprint of *Statistics of Directional Data* by Mardia [MR0336854 (49 #1627)]. MR1828667

MARTINS, A. B. T., BONAT, W. H. and RIBEIRO, P. J. JR. (2016). Likelihood analysis for a class of spatial geostatistical compositional models. *Spat. Stat.* **17** 121–130. MR3532093

MARUOTTI, A., PUNZO, A., MASTRANTONIO, G. and LAGONA, F. (2015). A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. *Stoch. Environ. Res. Risk Assess.* **30** 1725–1740.

MASTRANTONIO, G., JONA LASINIO, G. and GELFAND, A. E. (2016). Spatio-temporal circular models with non-separable covariance structure. *TEST* **25** 331–350. MR3493522

MASTRANTONIO, G., MARUOTTI, A. and JONA-LASINIO, G. (2015). Bayesian hidden Markov modelling using circular-linear general projected normal distribution. *Environmetrics* **26** 145–158. MR3324908

MCCLINTOCK, B. T., KING, R., THOMAS, L., MATTHIOPOULOS, J., MCCONNELL, B. J. and MORALES, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecol. Monogr.* **82** 335–349. DOI:10.1890/11-0326.1.

MECH, D. and BOITANI, L. (2003). *Wolves*: *Behavior, Ecology and Conservation*. Univ. Chicago Press, Chicago, IL.

MECH, L. D., WOLF, P. C. and PACKARD, J. M. (1999). Regurgitative food transfer among wild wolves. *Can. J. Zool.* **77** 1192–1195.

MERRILL, S. B. and DAVID MECH, L. (2000). Details of extensive movements by Minnesota wolves (Canis lupus). *Am. Midl. Nat.* **144** 428–433.

MORALES, J. M., HAYDON, D. T., FRAIR, J., HOLSINGER, K. E. and FRYXELL, J. M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology* **85** 2436–2445.

MULLER, P. (2017). *Bayesian Nonparametric Data Analysis*. CreateSpace Independent Publishing Platform.

NIU, M., BLACKWELL, P. G. and SKARIN, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics* **72** 315–324. MR3515758

PACI, L. and FINAZZI, F. (2018). Dynamic model-based clustering for spatio-temporal data. *Stat. Comput.* **28** 359–374. MR3747568

PARTON, A. and BLACKWELL, P. G. (2017). Bayesian inference for multistate 'step and turn' animal movement in continuous time. *J. Agric. Biol. Environ. Stat.* **22** 373–392. MR3692470

PATTERSON, T. A., BASSON, M., BRAVINGTON, M. V. and GUNN, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *J. Anim. Ecol.* **78** 1113–1123.

PATTERSON, T. A., PARTON, A., LANGROCK, R., BLACKWELL, P. G., THOMAS, L. and KING, R. (2017). Statistical modelling of individual animal movement: An overview of key methods and a discussion of practical challenges. *AStA Adv. Stat. Anal.* **101** 399–438. MR3712406

PAWLOWSKY, V. and BURGER, H. (1992). Spatial structure analysis of regionalized compositions. *Math. Geol.* **24** 675–691.

PETERS, R. P. (1979). Mental maps in wolf territoriality. In *The Behavior and Ecology of Wolves* (E. K. Klinghammer, ed.) 119–152. Garland STPM Press, New York.

POHLE, J., LANGROCK, R., VAN BEEST, F. M. and SCHMIDT, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement. *J. Agric. Biol. Environ. Stat.* **22** 270–293. MR3692465

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539–550. MR2672482

ROTHMAN, R. J. and MECH, L. D. (1979). Scent-marking in lone wolves and newly formed pairs. *Anim. Behav.* **27** 750–760.

SAND, H., ZIMMERMANN, B., WABAKKEN, P., ANDREN, H. and PEDERSEN, H. C. (2010). Using GPS technology and GIS cluster analyses to estimate kill rates in wolf-ungulate ecosystems. *Wildl. Soc. Bull.* **33** 914–925.

SCHARF, H. R., HOOTEN, M. B., JOHNSON, D. S. and DURBAN, J. W. (2018). Process convolution approaches for modeling interacting trajectories. *Environmetrics* **29** e2487, 18. MR3799912

THEUERKAUF, J. (2009). What drives wolves: Fear or hunger? Humans, diet, climate and wolf activity patterns. *Ethology* **115** 649–657.

TJELMELAND, H. and LUND, K. V. (2003). Bayesian modelling of spatial compositional data. *J. Appl. Stat.* **30** 87–100. MR1957363

VILA, C. (1995). Observations on the daily activity patterns in the Iberian wolf. *Ecology and Conservation of Wolves in a Changing World* **35** 335–340.

WANG, F. and GELFAND, A. E. (2013). Directional data analysis under the general projected normal distribution. *Stat. Methodol.* **10** 113–127. MR2974815

ZUCCHINI, W. and MACDONALD, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R. Monographs on Statistics and Applied Probability* **110**. CRC Press, Boca Raton, FL. MR2523850

G. MASTRANTONIO
DIPARTIMENTO DI SCIENZE MATEMATICHE
POLITECNICO DI TORINO
CORSO DUCA DEGLI ABRUZZI 24
10129 TORINO
ITALY
E-MAIL: gianluca.mastrantonio@polito.it

C. GRAZIAN
DIPARTIMENTO DI ECONOMIA
UNIVERSITÀ DEGLI STUDI
"GABRIELE D'ANNUNZIO"
VIALE PINDARO 42
65127 PESCARA
ITALY
AND
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW SOUTH WALES
THE RED CENTRE, CENTRE WING, KENSINGTON
NSW 2052 SYDNEY
AUSTRALIA

S. MANCINELLI
DIPARTIMENTO DI BIOLOGIA E BIOTECNOLOGIE
UNIVERSITÀ DI ROMA "LA SAPIENZA"
PIAZZALE ALDO MORO 5
00185 ROMA
ITALY

E. BIBBONA
DIPARTIMENTO DI SCIENZE MATEMATICHE
POLITECNICO DI TORINO
CORSO DUCA DEGLI ABRUZZI 24
10129 TORINO
ITALY