POLITECNICO DI TORINO
Repository ISTITUZIONALE

The modelling of movement of multiple animals that share behavioural features

Original
The modelling of movement of multiple animals that share behavioural features / Mastrantonio, G.. - In: JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS. - ISSN 0035-9254. - (2022). [10.1111/rssc.12561]

Availability:
This version is available at: 11583/2961714 since: 2022-04-20T16:02:08Z

Publisher:
Wiley

Published
DOI:10.1111/rssc.12561

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 November 2026

Received: 30 December 2020 | Accepted: 24 January 2022

DOI: 10.1111/rssc.12561

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY
Series C Applied Statistics

The modelling of movement of multiple animals that share behavioural features

Gianluca Mastrantonio

Department of Mathematical Sciences, Politecnico di Torino, Turin, Italy

Correspondence
Gianluca Mastrantonio, Department of Mathematical Sciences, Politecnico di Torino, Turin, Italy.
Email: gianluca.mastrantonio@polito.it

Abstract

In this work, we propose a model that can be used to infer the behaviour of multiple animals. Our proposal is defined as a set of hidden Markov models that are based on the sticky hierarchical Dirichlet process, with a shared base-measure, and a step and turn with an attractive point (STAP) emission distribution. The latent classifications are representative of the behaviour assumed by the animals, which is described by the STAP parameters. Given the latent classifications, the animals are independent. As a result of the way we formalize the distribution over the STAP parameters, the animals may share, in different behaviours, the set or a subset of the parameters, thereby allowing us to investigate the similarities between them. The hidden Markov models, based on the Dirichlet process, allow us to estimate the number of latent behaviours for each animal, as a model parameter. This proposal is motivated by a real data problem, where the global positioning system (GPS) coordinates of six Maremma Sheepdogs have been observed. Among the other results, we show that four dogs share most of the behaviour characteristics, while two have specific behaviours.

KEYWORDS

attractive-point, directional-persistence, Maremma Sheepdogs, STAP density

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

1 | INTRODUCTION

Movement data are often based on a time-series of two-dimensional spatial coordinates recorded using a global positioning system (GPS) device attached to an animal. Since the first paper by Dunn and Gipson (1977), the statistical models used to analyse such data have become increasingly popular, and are used to understand different aspects of the movement of animals, ranging from habitat selection (Hebblewhite & Merrill, 2008) to behaviour analysis (Anderson & Lindzey, 2003; Maruotti et al., 2016; Mastrantonio, 2018; Merrill & David Mech, 2000); for a detailed review, the reader may refer to Hooten et al. (2017).

Three major categories of movement-description models can be identified in the behaviour modelling framework: biased random walks (BRWs), correlated random walks (CRWs), and bias and correlated random walks (BCRWs). In a BRW, the animal movement is attracted (or biased) toward a point in space, which is called *center-of-attraction* (see, e.g., Blackwell, 1997; Dunn & Gipson, 1977). The center-of-attraction can be interpreted as a proxy of the home-range (Christ et al., 2008) or it can describe a movement toward a patch of space (McClintock et al., 2012). In a CRW, the movement direction, at any given time, depends on the previous direction. This characteristic is called *directional persistence* (Jonsen et al., 2005) and it is useful to describe a constant change in direction between consecutive observations. If both directional persistence and attractors are used to describe a movement, the model is a BCRW (Codling et al., 2008; Fortin et al., 2005; McClintock et al., 2012). A movement-description model is generally used as an emission distribution of a mixture-type model, where a latent cluster-membership variable is used to identify the behaviour assumed by an animal. If the observed time-window is wide enough, the use of a mixture-type model is justified by the assumption that an animal exhibits more than one behaviour during the day (Patterson et al., 2008), e.g., sleeping and hunting. The switching between behaviours is often temporally structured and, if formulated in a discrete-time framework, the model is usually the hidden Markov model (HMM) (Langrock et al., 2012; Michelot et al., 2016).

The literature on the modelling of multiple animals is not as extensive as that on single individuals, even though coordinates of different animals are often recorded. Nonetheless, interest in this topic is increasing (see, e.g., Westley et al., 2018) since, as shown by (Jonsen, 2016), the joint modelling of multiple animals often increases the precision of the estimates. By adopting the classification given by Scharf and Buderman (2020), it is possible to model multiple animals using two approaches. In the first approach, called *indirect*, the parameters that govern the behaviours are seen as random effects across animals, that is, they come from a common distribution, whose parameters must be estimated, and the animals are conditionally independent (see, e.g., Buderman et al., 2018; McClintock et al., 2013; Michelot et al., 2017). However, in the *direct* approach, the dependence between animals is described by an unobserved graph or social network, see, e.g., Calabrese et al. (2018), Hooten et al. (2018), Milner et al. (2021) and Niu et al. (2020).

We here propose a Bayesian model which can be used to describe multiple animals that share certain movement characteristics, observed over different time-windows. The model is based on the hierarchical Dirichlet process (HDP) (Teh et al., 2006) and it is a generalization of the sticky hierarchical Dirichlet process HMM (sHDP-HMM) of Fox et al. (2011). In the model, given the latent classification and likelihood parameters, the animals are independent and the behaviours are described by the five parameters of the step and turn with an attractive point (STAP) distribution, which is a BCRW emission-distribution that has recently been proposed by Mastrantonio (2020). The main contributions of the present proposal are the possibility of estimating the number of latent behaviours of each animal as model parameters, and of

introducing the sharing of parameters between behaviours and animals in HMMs based on Dirichlet processes (DPs). The former is a by-product of the DP modelling, which allows us to avoid the use of information criteria to select the number of behaviours, which has been shown to be problematic in this context (Pohle et al., 2017), or a trans-dimensional Markov chain Monte Carlo (MCMC) algorithm, such as the reversible jump MCMC (RJMCMC), which presents challenges in its implementation (Hastie & Green, 2012). The sharing of parameters is introduced in the lower level of the model hierarchy, where the distribution over the STAP parameters is defined by combining five different DPs. This distribution is discrete, with a countable number of atoms being defined so that they can share some of their multivariate components. This approach is similar to the one proposed by Mastrantonio et al. (2021), which was used to model climate data in a change-point framework. Therefore, the behaviours within or between animals can have the same value of an STAP parameter, and this allows us to investigate similarities and differences between the analysed animals. Other approaches also exist that have the sharing of parameters as one of their characteristics (see, e.g., Jonsen (2016) or Milner et al. (2021)), but they require one to select, a priori, what parameters are allowed to change and the number of values that a parameter can assume. However, in our proposal, everything is done during the model fitting and driven by the information within the data.

Our proposal has been used to model the trajectories of 6 Maremma Sheepdogs, observed in Australia with recorded coordinates every 30 min. These dogs are used all over Europe and Asia to protect livestock from possible predators and, in recent years, also in Australia (see, e.g., van Bommel & Johnson, 2016; Gehring et al., 2017). Maremma Sheepdogs are able to work in synergy with the shepherd to keep the stock together but this is not always possible when the property is too large. For this reason, the dogs are often left alone and are rarely visited by the shepherd. The owner has no supervision over the dogs and it is therefore interesting to analyse and understand their behaviour. The used dataset was taken from the movebank repository (`www.movebank.org`) and is described in detail in van Bommel and Johnson (2014a) and van Bommel and Johnson (2014b). With our model, we have identified many similarities and some specific features between dogs, that are easy to interpret and which give a better insight into the behaviour of the dogs. Two competitive models have also been estimated on the same data and the results are compared with our proposal.

The paper is organized as follows. We introduce the STAP density in Section 2, and the hierarchical formalization of our proposal in Section 3, while Section 4 contains the results of the real data application. The paper ends with some conclusive remarks in Section 5. The Web-based supporting materials, available on the web page of the journal, contain details of the MCMC algorithm and on the results of the competitive models.

2 | THE STAP DISTRIBUTION

With the aim of better understanding the results of the real data application (Section 4) and the formalization of our proposal, we briefly describe the STAP distribution, which was introduced in Mastrantonio (2020), and its parameters; for a more detailed description the reader may refer to Mastrantonio (2020).

We assume we have a time-series of two-dimensional spatial locations $\mathbf{s} = (\mathbf{s}_{t_1}, \dots, \mathbf{s}_{t_T})'$ that represent an animal's path, where $\mathbf{s}_{t_i} = (s_{t_i,1}, s_{t_i,2}) \in \mathcal{D} \subset \mathbb{R}^2$, and t_i is a temporal index. The coordinates are recorded without any measurement error and the time difference between consecutive points is constant. In order to formalize the STAP, we introduce the bearing angle

$$\phi_{t_i} = \text{atan}^*(s_{t_{i+1},2} - s_{t_i,2}, s_{t_{i+1},1} - s_{t_i,1}) \in [-\pi, \pi),$$

and the rotation matrix

$$\mathbf{R}(x) = \begin{pmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{pmatrix},$$

where $\text{atan}^*(\cdot)$ is the two-argument tangent function (Jammalamadaka & Kozubowski, 2004). The bearing-angle measures the direction of the movement between time t_i and t_{i+1}, and the rotation matrix can be used to perform a rotation in a two-dimensional space, so that if it multiplies a two-dimensional vector, the vector is rotated anti-clockwise by an angle x. The conditional distribution of $\mathbf{s}_{t_{i+1}}$ is assumed to be second-order Markovian, with the following specification:

$$\begin{aligned} \mathbf{s}_{t_{i+1}} | \mathbf{s}_{t_i}, \mathbf{s}_{t_{i-1}} &\sim N(\mathbf{s}_{t_i} + \mathbf{M}_{t_i}, \mathbf{V}_{t_i}), i \in \{1, \dots, T-1\}, \\ \mathbf{M}_{t_i} &= (1-\rho)\tau \left(\boldsymbol{\mu} - \mathbf{s}_{t_i}\right) + \rho \mathbf{R}(\phi_{t_{i-1}})\boldsymbol{\eta}, \\ \mathbf{V}_{t_i} &= \mathbf{R}(\rho\phi_{t_{i-1}})\boldsymbol{\Sigma}\mathbf{R}'(\rho\phi_{t_{i-1}}), \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}, \boldsymbol{\eta} \in \mathbb{R}^2$, $\tau \in (0, 1)$, $\rho \in [0, 1]$, and $\boldsymbol{\Sigma}$ is a two-dimensional covariance matrix. The location $\mathbf{s}_{t_1}$ is fixed, and $\mathbf{s}_{t_0}$ is another parameters that is needed to compute ϕ_{t_0} in the conditional distribution of $\mathbf{s}_{t_2}$. If the path follows Equation (1), we write $\mathbf{s}_{t_{i+1}} | \mathbf{s}_{t_i}, \mathbf{s}_{t_{i-1}} \sim \text{STAP}(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Sigma}, \tau, \rho)$.

The movement described by the STAP can have directional-persistence and attraction to a point in space, therefore, the STAP is a BCRW. To better understand these two properties and how they are formalized in Equation (1), we introduce the vector $\vec{\mathbf{F}}_{t_i}$, which is a vector with initial and terminal points equal to $\mathbf{s}_{t_i}$ and $\mathbf{s}_{t_i} + \mathbf{M}_{t_i}$ respectively. This vector represents the expected movement between time t_i and t_{i+1}, since its initial point is the previously observed location $\mathbf{s}_{t_i}$ and the terminal one is equal to $\mathbb{E}(\mathbf{s}_{t_{i+1}} | \mathbf{s}_{t_i})$. If $\rho = 0$, the STAP reduces to a two-dimensional AR(1) (a BRW), and $\vec{\mathbf{F}}_{t_i}$ points to the spatial location $\boldsymbol{\mu}$, which is therefore the attractor. The length of $\vec{\mathbf{F}}_{t_i}$ is $\tau \|(\boldsymbol{\mu} - \mathbf{s}_{t_i})\|$, which shows that τ measures how much of the total distance between the last observation ($\mathbf{s}_{t_i}$) and the attractor ($\boldsymbol{\mu}$) is expected to be covered or, in other words, how strong the attraction to $\boldsymbol{\mu}$ is. If $\rho = 1$, the STAP reduces to a CRW, based on a normal density. In this case, the direction of $\vec{\mathbf{F}}_{t_i}$ is the same as the direction of $\mathbf{R}(\phi_{t_{i-1}})\boldsymbol{\eta}$, which depends on the previous bearing angle $\phi_{t_{i-1}}$, and thus induces a directional-correlation between consecutive points. If $\rho \in (0, 1)$, $\vec{\mathbf{F}}_{t_i}$ is a weighted mean between its value in a BRW and a CRW, with weights given by $(1 - \rho)$ and ρ respectively. The covariance matrix of the conditional distribution of $\mathbf{s}_{t_{i+1}}$ is fixed in a BRW ($\text{Cov}(\mathbf{s}_{t_{i+1}} | \mathbf{s}_{t_i}) = \boldsymbol{\Sigma}$), while it rotates with the bearing-angle for any $\rho > 0$ ($\text{Cov}(\mathbf{s}_{t_{i+1}} | \mathbf{s}_{t_i}) = \mathbf{R}(\rho\phi_{t_{i-1}})\boldsymbol{\Sigma}\mathbf{R}'(\rho\phi_{t_{i-1}})$): for more details, the reader may refer to Mastrantonio (2020).

We show examples of STAP densities, and the associated BRW ($\rho = 0$) and CRW ($\rho = 1$) in Figure 1, to better understand the differences between the BRW, CRW and BCRW. The dashed arrow in the figure is the movement between times t_{i-1} and t_i, the solid arrow is $\vec{\mathbf{F}}_{t_i}$, and the ellipse is a contour of the conditional distribution of $\mathbf{s}_{t_{i+1}}$ with a constant density containing 95% of the total probability mass. From 1 (a), we can see that the direction of $\vec{\mathbf{F}}_{t_i}$ and the ellipse change according to $\phi_{t_{i-1}}$ in a CRW. However, $\vec{\mathbf{F}}_{t_i}$ and the ellipse are independent from the previous direction in a BRW (Figure 1b), and $\vec{\mathbf{F}}_{t_i}$ points to the spatial attractor $\boldsymbol{\mu} = (0, 0)'$. When $\rho \in (0, 1)$, the

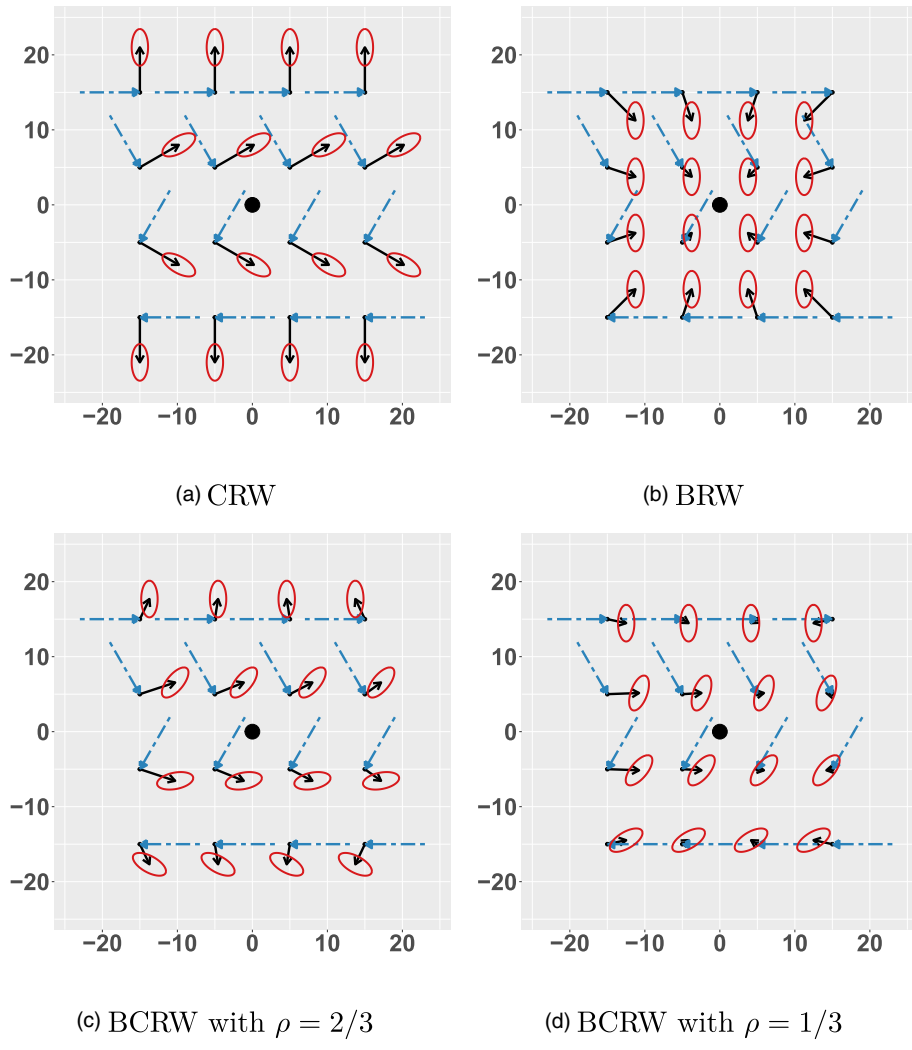


FIGURE 1 Graphical representation of the conditional distribution of $\mathbf{s}_{i+1}$ ((a) CRW, (b) BRW, (c), (d) BCRW), for different possible values of $\mathbf{s}_i$ and the previous directions. The dashed arrow represents the movement between $\mathbf{s}_{i-1}$ and $\mathbf{s}_i$. The solid arrow is $\vec{\mathbf{F}}_i$, while the ellipse is the area containing 95% of the probability mass of the conditional distribution of $\mathbf{s}_{i+1}$. $\boldsymbol{\mu} = (0, 0)'$, $\boldsymbol{\eta} = (0, 6)'$, $\tau = 0.25$, $\boldsymbol{\Sigma} = \begin{pmatrix} 0.2 & 0 \\ 0 & 1 \end{pmatrix}$ in all figures, and the central dot is the location $\boldsymbol{\mu}$, which is the attractor in the biased random walks and the bias and correlated random walks

ellipse and $\vec{\mathbf{F}}_{t_i}$ are dependent on both the previous direction and the spatial attractor, see Figures 1c and d.

3 | THE PROPOSED MODEL

In this section, we introduce the components of the model and how they are used to introduce the characteristics of our proposal. We extend the notation of the previous section to describe the path of m animals, and to allow changes in the behaviour to be considered.

We indicate the path of the j th animal with $\mathbf{s}_j = (\mathbf{s}_{j,t_{j,1}}, \mathbf{s}_{j,t_{j,2}}, \ldots, \mathbf{s}_{j,t_{j,T_j}})$, where $j = 1, \ldots, m$, and $\mathcal{T}_j \equiv (t_{j,1}, t_{j,2}, \ldots, t_{j,T_j})$ is the set of temporal points, equally spaced in time, where the position of the j th dog is recorded. The sets $\mathcal{T}_j$ and $\mathcal{T}_{j'}$ can contain different time-points, but the time difference must be constant across animals, that is, $t_{j,i+1} - t_{j,i} = c$ for all $j = 1, \ldots, m$ and $i = 1, \ldots T_j - 1$. We introduce a discrete random variable $z_{j,t_{j,i}} \in \mathbb{N}$ to represent the animal behaviour at time $t_{j,i}$, where $z_{j,t_{j,i}} = k$ indicates that animal j follows behaviour k at time $t_{j,i}$. Given the behaviour assumed by each animal, the paths are independent and

$$\mathbf{s}_{j,t_{j,i+1}} | \mathbf{s}_{j,t_{j,i}}, \mathbf{s}_{j,t_{j,i-1}}, z_{j,t_{j,i}} \sim \text{STAP}(\boldsymbol{\theta}_{z_{j,t_{j,i}}})$$

where $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\eta}_k, \boldsymbol{\Sigma}_k, \tau_k, \rho_k)$. In other words, if the j th animal is following the k th behaviour at time $t_{j,i}$ (i.e. $z_{j,t_{j,i}} = k$), the path is described by the set $\boldsymbol{\theta}_k$ of STAP parameters. It should be noted that the k th behaviours are represented by the same set of parameters $\boldsymbol{\theta}_k$ for all animals.

Let $\mathbf{s} = \{\mathbf{s}_j\}_{j=1}^m$, $\mathbf{z}_j = \{z_{j,t}\}_{t \in \mathcal{T}_j}$, $\mathbf{z} = \{\mathbf{z}_j\}_{j=1}^m$, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k\}_{k \in \mathbb{N}}$, then the model we propose is

$$f(\mathbf{s}|\boldsymbol{\theta}, \mathbf{z}) = \prod_{j=1}^{m} \prod_{i=1}^{T_j - 1} f\left(\mathbf{s}_{j,t_{j,i+1}} | \mathbf{s}_{j,t_{j,i}}, \mathbf{s}_{j,t_{j,i-1}}, \boldsymbol{\theta}_{z_{j,t_{j,i}}}\right), \tag{2}$$

$$\mathbf{s}_{j,t_{j,i+1}} | \mathbf{s}_{j,t_{j,i}}, \mathbf{s}_{j,t_{j,i-1}}, z_{j,t_{j,i}}, \boldsymbol{\theta}_{z_{j,t_{j,i}}} \sim \text{STAP}\left(\boldsymbol{\theta}_{z_{j,t_{j,i}}}\right), \quad \mathbf{s}_{j,t_{j,0}} \sim \text{Unif}(\mathcal{D}), \tag{3}$$

$$\begin{aligned} z_{j,t_{j,i}} | z_{j,t_{j,i-1}}, \boldsymbol{\pi}_{j,z_{j,t_{j,i-1}}} &\sim \text{Multinomial}\left(1, \boldsymbol{\pi}_{j,z_{j,t_{j,i-1}}}\right), \\ z_{j,t_{j,0}} &\sim \text{Geom}(\epsilon), \end{aligned} \tag{4}$$

$$\boldsymbol{\pi}_{j,l} | \alpha, \nu, \boldsymbol{\beta} \sim \text{DP}\left(\alpha + \nu, \frac{\alpha \boldsymbol{\beta} + \nu \delta_l}{\alpha + \nu}\right), \tag{5}$$

$$\{\beta_k\}_{k \in \mathbb{N}} = C_1(\boldsymbol{\beta}^*_{\boldsymbol{\mu}}, \boldsymbol{\beta}^*_{\boldsymbol{\eta}}, \boldsymbol{\beta}^*_{\boldsymbol{\Sigma}}, \boldsymbol{\beta}^*_{\tau}, \boldsymbol{\beta}^*_{\rho}), \tag{6}$$

$$\{\boldsymbol{\theta}_k\}_{k \in \mathbb{N}} = C_2(\boldsymbol{\mu}^*, \boldsymbol{\eta}^*, \boldsymbol{\Sigma}^*, \boldsymbol{\tau}^*, \boldsymbol{\rho}^*), \tag{7}$$

$$\begin{aligned} \boldsymbol{\beta}^*_{\boldsymbol{\mu}} | \gamma_{\boldsymbol{\mu}} \sim \text{Gem}(\gamma_{\boldsymbol{\mu}}), \boldsymbol{\beta}^*_{\boldsymbol{\eta}} | \gamma_{\boldsymbol{\eta}} \sim \text{Gem}(\gamma_{\boldsymbol{\eta}}), \boldsymbol{\beta}^*_{\boldsymbol{\Sigma}} | \gamma_{\boldsymbol{\Sigma}} \sim \text{Gem}(\gamma_{\boldsymbol{\Sigma}}), \\ \boldsymbol{\beta}^*_{\tau} | \gamma_{\tau} \sim \text{Gem}(\gamma_{\tau}), \boldsymbol{\beta}^*_{\rho} | \gamma_{\rho} \sim \text{Gem}(\gamma_{\rho}), \end{aligned} \tag{8}$$

$$\begin{aligned} \boldsymbol{\mu}^*_p | H_{\boldsymbol{\mu}} \sim H_{\boldsymbol{\mu}}, \boldsymbol{\eta}^*_p | H_{\boldsymbol{\eta}} \sim H_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}^*_p | H_{\boldsymbol{\Sigma}} \sim H_{\boldsymbol{\Sigma}}, \\ \tau^*_p | H_{\tau} \sim H_{\tau}, \rho^*_p | H_{\rho} \sim H_{\rho}, \end{aligned} \tag{9}$$

where $p \in \mathbb{N}$, $l \in \mathbb{N}$, and $i = 1, \ldots, T_j - 1$. A full description of the components of the model and how they are used to introduce the main novelties of our model is given below.

The DPs. In order to simplify the description of the lower levels of the model hierarchy, we use χ as a variable that can be $\boldsymbol{\mu}$, $\boldsymbol{\eta}$, $\boldsymbol{\Sigma}$, ν, or ρ, and it is used when whatever is described can be applied to any of the five parameters. In Equation (9) the values of the STAP parameter χ_p are sampled from the distribution H_χ, and the infinite-dimensional vector of probabilities $\boldsymbol{\beta}^*_\chi = \{\beta^*_{\chi,p}\}_{p \in \mathbb{N}}$, associated with parameter χ, is Gem distributed (Gnedin et al., 2001) with *scaling parameter*γ_χ. The scaling parameter can easily be interpreted with the stick-breaking representation of the GEM distribution, defined as follows:

$$\begin{aligned} \beta^*_{\chi,1} &\sim \mathrm{B}(1,\gamma_\chi), \\ \frac{\beta^*_{\chi,p}}{1-\sum_{h=1}^{p-1}\beta^*_{\chi,h}} &\sim \mathrm{B}(1,\gamma_\chi), \quad p \neq 1. \end{aligned} \tag{10}$$

From Equation (10), we see that the smaller γ_χ is, and the smaller is the number of elements of $\boldsymbol{\beta}_\chi$ that contains most of the probability mass with $\lim_{\gamma_\chi \to 0} \beta^*_{\chi,1} = 1$ and $\lim_{\gamma_\chi \to 0} \beta^*_{\chi,p} = 0$ for all $p \neq 1$. The vectors $\boldsymbol{\beta}^*_\chi$ and $\boldsymbol{\chi}^* = \{\chi^*_p\}_{p\in\mathbb{N}}$ can be used to define the discrete distribution

$$G_\chi = \sum_{p\in\mathbb{N}} \beta^*_{\chi,p}\delta_{\chi^*_p}, \tag{11}$$

where $\delta_\cdot$ is the Dirac delta function. Equation (11) is a draw from a $\mathrm{DP}(\gamma_\chi, H_\chi)$, and thus we can equivalently describe $\boldsymbol{\beta}^*_\chi$ and χ_p as the components of a sample from $\mathrm{DP}(\gamma_\chi, H_\chi)$, or as Equations (8) and (9). The vectors $\boldsymbol{\chi}^*$ and $\boldsymbol{\beta}^*_\chi$ contain, respectively, the values that the parameters can assume (χ^*_p) and the 'base' probability ($\beta^*_{\chi_p}$) that a particular value of the parameter is selected in a behaviour (see Equation (15) below).

The functions $C_1(\cdot)$ ***and*** $C_2(\cdot)$. The function $C_1(\cdot)$ and $C_2(\cdot)$ (Equations (6) and (7)) introduce the sharing of parameters between behaviours, which is one of the novelties of our proposal. The role of function $C_2(\cdot)$ is to produce the set of STAP parameters $\{\boldsymbol{\theta}_k\}_{k\in\mathbb{N}}$, where we remind the reader that $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\eta}_k, \boldsymbol{\Sigma}_k, \tau_k, \rho_k)$. The set $\{\boldsymbol{\theta}_k\}_{k\in\mathbb{N}}$ is comprised of all the possible combinations of the 5 STAP parameters, without repetitions. This means that $\boldsymbol{\theta}_k \neq \boldsymbol{\theta}_{k'}$, if $k \neq k'$, but we can have a subset of elements that has the same value, for example, $\tau_k \equiv \tau_{k'}$. Hence, since each behaviour selects its STAP parameters in $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k\}_{k\in\mathbb{N}}$, different behaviours can share parameters, even though they are described by a different $\boldsymbol{\theta}_k$.

Function $C_1(\cdot)$ is closely related to $C_2(\cdot)$ since, if we introduce the new variables $\lambda_{\boldsymbol{\mu},k}$, $\lambda_{\boldsymbol{\eta},k}$, $\lambda_{\boldsymbol{\Sigma},k}$, $\lambda_{\tau,k}$ and $\lambda_{\rho,k}$ that represent what parameter is in $\boldsymbol{\theta}_k$, that is,

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}^*_{\lambda_{\boldsymbol{\mu},k}}, \boldsymbol{\eta}_k = \boldsymbol{\eta}^*_{\lambda_{\boldsymbol{\eta},k}}, \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}^*_{\lambda_{\boldsymbol{\Sigma},k}}, \boldsymbol{\tau}_k = \tau^*_{\lambda_{\tau,k}}, \boldsymbol{\rho}_k = \rho^*_{\lambda_{\rho,k}}, \tag{12}$$

we can associate a value β_k to $\boldsymbol{\theta}_k$ which is computed as

$$\beta_k = \beta^*_{\boldsymbol{\mu},\lambda_{\boldsymbol{\mu},k}}\beta^*_{\boldsymbol{\eta},\lambda_{\boldsymbol{\eta},k}}\beta^*_{\boldsymbol{\Sigma},\lambda_{\boldsymbol{\Sigma},k}}\beta^*_{\tau,\lambda_{\tau,k}}\beta^*_{\rho,\lambda_{\rho,k}}.$$

The set $\{\beta_k\}_{k\in\mathbb{N}}$ is the output of $C_1(\cdot)$ and it is a probability vector, since $\sum_{k\in\mathbb{N}}\beta_k = 1$ and $\beta_k \in (0,1)$. We can define the discrete distribution

$$G_0 = \sum_{k\in\mathbb{N}} \beta_k \delta_{\boldsymbol{\theta}_k}, \tag{13}$$

where, similarly to Equation (11), its atoms contain all the possible values that $\boldsymbol{\theta}_k$ can assume and β_k is connected to the expected value of the probability of selection $\boldsymbol{\theta}_k$ as the vector of parameter in a behaviour (see Equation (14) below). This way to define the distribution G_0 is closely related to the shared base-distribution of the change-point model of Mastrantonio et al. (2021).

Behaviour switching. Let $\boldsymbol{\Pi}_j$ be the matrix that has $\boldsymbol{\pi}_{j,l} = \{\pi_{j,l,k}\}_{k\in\mathbb{N}}$ as lth row. Matrix $\boldsymbol{\Pi}_j$ rules the switching between the behaviours of animal j (Equation 4) and if the jth animal is following

behaviour l at time $t_{j,i-1}$, the probability of switching to behaviour k is given by the element of $\mathbf{\Pi}_j$ in row l and column k. Hence, the time evolution of $z_{j,t_{j,i}}$ is modeled by a discrete first-order Markov process, which defines an HMM with transition matrix $\mathbf{\Pi}_j$ and initial state $z_{j,t_{j,0}}$. The initial state is drawn from a Geometric distribution with parameter ϵ, which is defined as the number of Bernoulli trials needed to have one success. The row $\boldsymbol{\pi}_{j,l}$ is DP distributed (see Equation (5)) and the expected value of $\pi_{j,l,k}$ is equal to

$$\mathrm{E}(\pi_{j,l,k}|\alpha,\nu,\boldsymbol{\beta}) = \frac{\alpha\beta_k + \nu\delta(l,k)}{\alpha+\nu} \tag{14}$$

(see Fox et al., 2011), where $\delta(l, k)$ is equal to 1 if $l = k$, and 0 otherwise. From Equation (14), we can see that the kth element of $\boldsymbol{\beta}$ is associated with the expected value of $\pi_{j,l,k}$, for all the animals ($j = 1, \ldots, m$) and $l \in \mathbb{N}$. Hence, a larger β_k increases the probability of switching from any behaviour l to the kth, described by $\boldsymbol{\theta}_k$. However, Equation (14) can also be stated as

$$\mathrm{E}(\pi_{j,l,k}|\alpha,\nu,\boldsymbol{\beta}) = \frac{\alpha\beta^*_{\boldsymbol{\mu},\lambda_{\boldsymbol{\mu},k}}\beta^*_{\boldsymbol{\eta},\lambda_{\boldsymbol{\eta},k}}\beta^*_{\boldsymbol{\Sigma},\lambda_{\boldsymbol{\Sigma},k}}\beta^*_{\tau,\lambda_{\tau,k}}\beta^*_{\rho,\lambda_{\rho,k}} + \nu\delta(l,k)}{\alpha+\nu}, \tag{15}$$

which highlights how the value $\beta^*_{\chi,p}$ is connected to all $\pi_{j,l,k}$, with $l, k \in \mathbb{N}$ and $j = 1, \ldots, m$, so that $\lambda_{\chi,k} = p$. Therefore, a larger $\beta^*_{\chi,p}$ increases the expected values of all these probabilities, and for this reason we call β^*_χ the 'base probability' as χ^*_p. The variable α is the scaling parameter of the DP of Equation (5), which has the same interpretation of γ_χ, while ν is a weight that is added to the self transitions $\pi_{j,l,l}$ to increase their expected value, see Equation (14), which in turn is used to reduce the tendency of the HDP-HMM to create redundant behaviours, that is, behaviours with similar parameter vectors. For a more detailed description of parameters α and ν, the reader may refer to Fox et al. (2011).

It should be noted that, in most applications, see, e.g., McClintock et al. (2012) and Leos-Barajas et al. (2017), $z_{j,t_{j,i}} \in \{1, 2, \ldots, K^*\}$ is assumed, where K^* indicates the maximum number of behaviours, while we have $z_{j,t_{j,i}} \in \mathbb{N}$ in this work, since we define the HMM using DPs. Thus, the model assumes an infinite and countable number of possible behaviours for each animal, but, since we have a finite number of observed time-points, only a finite number K_j of them can be 'occupied'; these are generally called 'non-empty states' (Frühwirth-Schnatter & Malsiner-Walli, 2019), or, in this context, 'non-empty behaviours'. The random variable K_j is used to estimate the number of latent behaviours of the jth animal. Parameters α, ν and γ (through $\boldsymbol{\beta}$) determine the number of non-empty states K_j, since they are responsible with the total mass associated with each of the $\mathbf{\Pi}_j$ elements.

The emission-distribution. The model specification is concluded with the emission distribution, which is given by Equations (2) and (3). It should be noted that, given the latent behaviours, we consider the animals independent but, since they share the same set of atoms $\{\boldsymbol{\theta}_k\}_{k\in\mathbb{N}}$, the behaviours of the different animals can be described by the same STAP distribution. From Equation (12), we know that $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k'}$ can have common components and therefore, when an animal changes behaviour, it is not necessary for all the parameters to change and, more importantly, we can also identify the features that two animals share for different behaviours, for example, two behaviours can have the same attractive-point ($\boldsymbol{\mu}^*_p$), even though the strength of attraction (τ^*_p) is different. This feature is one of the main novelties of our proposal, and, although other approaches have a similar characteristic (see, e.g., Jonsen (2016)

or Milner et al. (2021)), they do not allow the number of latent behaviours to be estimated, which is the other main novelty of our proposal, the possibility of evaluating, during model fitting, what parameters are allowed to change, or the number of values that a parameter can assume. The sharing of a subset of the parameters between states is new in the context of HMMs based on DPs.

Connection to the sHDP-HMM. To conclude this section, we would like to show that our proposal can be considered as a generalization of the sHDP-HMM. The model of Fox et al. (2011) is defined for a single time-series, and G_0 is a draw from a DP. It is easy to see that, if we consider only one animal and use only one multivariate parameter in Equation (9), for example, $\boldsymbol{\theta}_p^* = (\boldsymbol{\mu}_p^*, \boldsymbol{\eta}_p^*, \boldsymbol{\Sigma}_p^*, \tau_p^*, \rho_p^*)$, with the associated vector of probability $\boldsymbol{\beta}_\theta^*$, the distribution G_0 is a draw from a DP, since $\beta_k = \beta_{\theta,p}^*$ and $\boldsymbol{\theta}_k = \boldsymbol{\theta}_p^*$. Hence, the model reduces to a sHDP-HMM.

4 | REAL DATA APPLICATION

We have the recorded coordinates of 6 dogs, taken every 30 min at the Heatherlie property in Australia, between 2012-11-10 15:30 and 2012-08-02 15:30. The data[1] consist of 4801 observations for each dog, with less than 1% of missing points. To facilitate the specification of the prior the coordinates are centered using the bivariate sample mean and scaled with a common standard deviation, computed using both the X and Y coordinates, to maintain the relative scale between the two coordinates; the recorded locations are shown in Figure 2. The dogs are called Woody, Sherlock, Alvin, Rosie, Bear, and Lucy. Rosie and Lucy are female, while the other four are male, and Woody, Sherlock, Bear and Lucy form a cohesive group, which is responsible for livestock protection, while Rosie, due to her advanced age, is solitary, and Alvin suffers from social exclusion, which restricted his movement (van Bommel & Johnson, 2016).

Maremma Sheepdogs originate from Europe, and have been used for centuries to protect livestock from potential predators (Gehring et al., 2017). They are trained to live with the livestock from birth and, as a result, they develop a strong bond with them and an instinct to protect them. They can be fence-trained, but are generally allowed to move freely. The use of livestock guardian dogs is relatively new outside Europe, especially in Australia, and, due to their effectiveness, interest in their use is increasing (van Bommel & Invasive Animals Cooperative Research Centre, 2010; van Bommel & Johnson, 2016). Since the extension of properties in Australia can be as much as several thousand hectares, it is hard for the owner to supervise the dogs (van Bommel & Johnson, 2012) and to have information about their behaviour (van Bommel & Invasive Animals Cooperative Research Centre, 2010).

4.1 | Comparison of the model and implementation details

We compare the predictive performances between our proposal (M1) and two competitive models (M2 and M3), using the integrated completed likelihood (ICL) (Biernacki et al., 2000) and the

[1]The dataset is freely available from the movebank repository https://www.datarepository.movebank.org/handle/10255/move.395

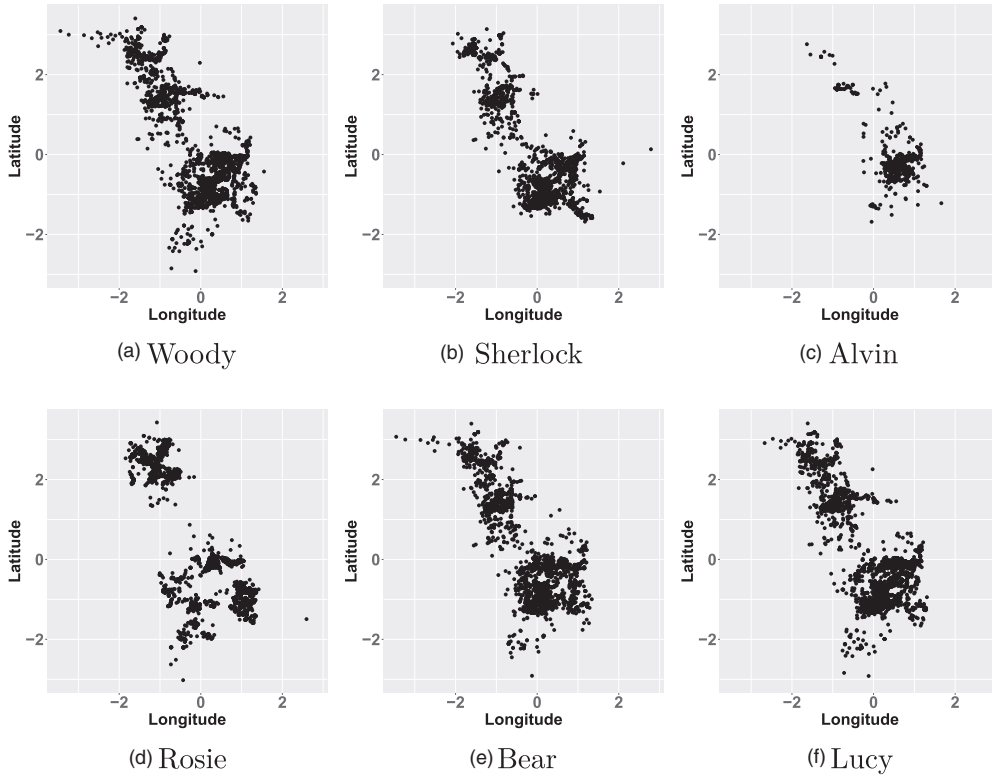


FIGURE 2 The observed spatial locations of the six dogs

deviance information criteria (DIC) DIC_5 and DIC_7 (Celeux et al., 2006). In the first competitive model (M2), we assume that only the entire vector of STAP parameters can be shared between animals, which means that each time-series is an sHDP-HMM with a share-based distribution. In terms of model formalization, we assume

$$\{\beta_k\}_{k\in\mathbb{N}}|\gamma \sim \text{Gem}(\gamma),$$
$$\boldsymbol{\theta}_k|H_{\boldsymbol{\mu}}, H_{\boldsymbol{\eta}}, H_{\boldsymbol{\Sigma}}, H_\tau, H_\rho \sim H_{\boldsymbol{\mu}} \times H_{\boldsymbol{\eta}} \times H_{\boldsymbol{\Sigma}} \times H_\tau \times H_\rho,$$

for the set of atoms and weights of G_0 (Equation 13). In the second competitor (M3), each animal follows the model of Mastrantonio (2020), which means they are completely independent, and there is no sharing of parameters. Hence, we substitute G_0 with the animal-specific distribution

$$G_{0,j} = \sum_{k\in\mathbb{N}} \beta_{j,k}\delta_{\tilde{\boldsymbol{\theta}}_{j,k}},$$

where

$$\{\beta_{j,k}\}_{k\in\mathbb{N}}|\gamma_j \sim \text{Gem}(\gamma_j),$$
$$\tilde{\boldsymbol{\theta}}_{j,k}|H_{\boldsymbol{\mu}}, H_{\boldsymbol{\eta}}, H_{\boldsymbol{\Sigma}}, H_\tau, H_\rho \sim H_{\boldsymbol{\mu}} \times H_{\boldsymbol{\eta}} \times H_{\boldsymbol{\Sigma}} \times H_\tau \times H_\rho,$$
$$\tilde{\boldsymbol{\theta}}_{j,k} = (\boldsymbol{\mu}_{j,k}, \boldsymbol{\eta}_{j,k}, \boldsymbol{\Sigma}_{j,k}, \tau_{j,k}, \rho_{j,k}),$$

TABLE 1 Information criteria for the proposed model (M1), sHDP-HHMMs with a common G_0 (M2), sHDP-HHMMs with animal-specific $G_{0,j}$ (M3). The model selected by each index is indicated in bold

	M1	M2	M3
ICL	**209593**	201880	192145
DIC5	**−457502**	−441264	−407823
DIC7	**−417544**	−400786	−376037

and then

$$\mathbf{s}_{j,t_{j,i+1}} | \mathbf{s}_{j,t_{j,i}}, \mathbf{s}_{j,t_{j,i-1}}, z_{j,t_{j,i}}, \tilde{\boldsymbol{\theta}}_{j,z_{j,t_{j,i}}} \sim \text{STAP}\left(\tilde{\boldsymbol{\theta}}_{j,z_{j,t_{j,i}}}\right).$$

By changing the way distribution G_0 is defined, we aim to show that the main feature of our proposal, that is, the sharing of sets and subsets of parameters, improves the model fitting and leads to a better description of the data.

The models are implemented assuming $H_{\boldsymbol{\mu}} \equiv N(\mathbf{0}, 20\mathbf{I})$, $H_{\boldsymbol{\eta}} \equiv N(\mathbf{0}, 20\mathbf{I})$, $H_{\tau} \equiv U(0,1)$, and $H_{\boldsymbol{\Sigma}} \equiv IW(3, \mathbf{I})$. The distribution H_{ρ} is assumed to be a mixture of a $U(0,1)$ and two bulks of probability on 0 and 1, with the 3 mixture weights equal to 1/3. This allows ρ_k (in M1 and M2) and $\rho_{j,k}$ (in M3) to be, a posteriori, equal to 0 or 1 with a greater probability than 0, which allows us to detect a pure CRW or BRW behaviour. We assume $\alpha + \nu, \gamma_{\boldsymbol{\mu}}, \gamma_{\boldsymbol{\eta}}, \gamma_{\boldsymbol{\Sigma}}, \gamma_{\rho}, \gamma_{\tau}, \gamma, \gamma_1, \ldots, \gamma_m \sim G(0.01, 0.01)$ and $\nu/(\alpha + \nu) \sim B(1, 1)$, which allows us to easily sample from their full conditionals, see Fox et al. (2011) and Section A of the Web-based Supporting Materials. All the distributions are chosen to be weakly informative. The domain $\mathcal{D}$ is a square $[-5, 5] \times [-5, 5]$ and the parameter ϵ of the Geometric distribution is equal to 0.00001, which means that the distribution over the initial state, $z_{j,t_{j,0}}$, is approximately uniform over the positive integers. Posterior estimates are obtained with 15,000 iterations, burnin 7500, thin 3, and thus 2500 samples are available for posterior inference. Convergence has been checked by means of a visual inspection of the posterior chains and using the $\hat{R}$ statistics (Gelman et al., 2013) Details on the MCMC algorithm, implemented in Julia 1.3 (Bezanson et al., 2017), can be found in the Web-based Supporting Materials, Section A, and the codes used to replicate the results, tables, and figures are available at https://github.com/GianlucaMastrantonio/multiple_animals_movement_model.

In Table 1, we can see that the three indices indicate that our model is the one with the best fit,[2] model M2 is the second, while M3 is always the last. Therefore, the joint modelling of the six dogs improves the performances of the model (since M2 is always preferable to M3), but the sharing of a subset of parameters also leads to a better description of the data (since M1 is better than M2). We provide a description of the results obtained with M2 and M3 in the Web-based Supporting Materials, Section B.

4.2 | Description and interpretation of the output

Using the algorithm proposed by Wade and Ghahramani (2018), we find a representative behaviour $\hat{z}_{j,t_{j,i}}$ associated with each animal and time, that we indicate as *MAP behaviour*. We indicate the kth behaviour of the jth dog based on $\hat{z}_{j,t_{j,i}}$ as B_{jk}, and let $n_{j,k}$ be the number of times we have $\hat{z}_{j,t_{j,i}} = k$, without any loss of generality, we assume $n_{j,1} > n_{j,2} > \cdots$, that is, the

[2]It should be noted that a higher ICL and a lower DIC indicate a better fit.

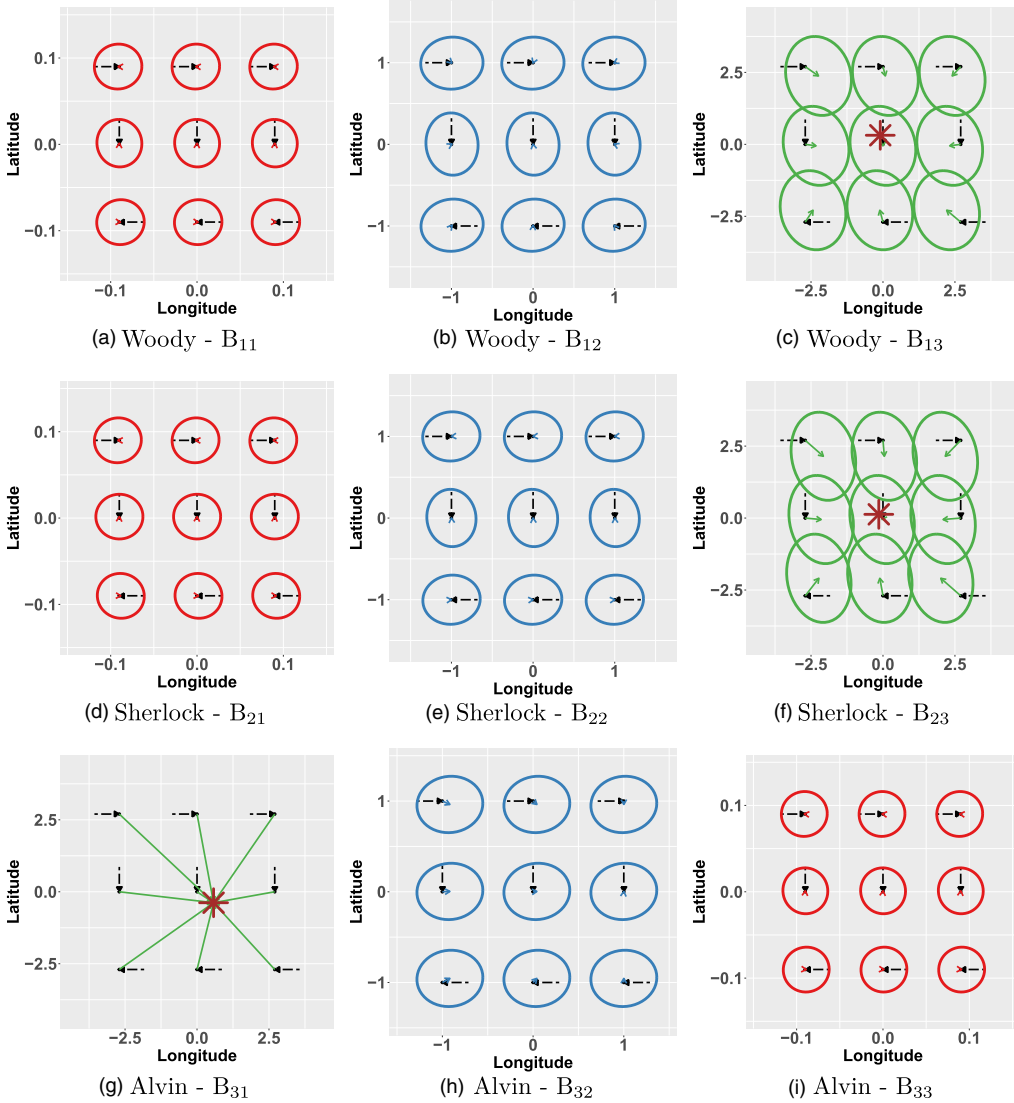


FIGURE 3 Graphical representation of the conditional distribution of $\mathbf{s}_{j,t_{j,i+1}}$ for the first three dogs ((a)(b)(c) Woody, (d)(e)(f) Sherlock, (g)(h)(i) Alvin), for different possible values of $\mathbf{s}_{j,t_{j,i}}$ and previous directions. The images has been obtained by using the posterior values that maximize the data likelihood of each animal, given the representative clusterization $\hat{z}_{j,t_{j,i}}$. The dashed arrow represents the movement between $\mathbf{s}_{j,t_{j,i-1}}$ and $\mathbf{s}_{j,t_{j,i}}$. The solid arrow is $\vec{\mathbf{F}}_{j,t_{j,i}}$, while the ellipse is an area containing 95% of the probability mass of the conditional distribution of $\mathbf{s}_{j,t_{j,i+1}}$. The asterisk represents the attractor, and it is only shown for behaviours that have posterior values of $\rho_{j,k} < 0.9$and $\tau_{j,k} > 0.1$

behaviours are ordered with respect to the number of times they are observed. It should be noted that B_{jk} is not the same as $B_{j'k}$, if $j \neq j'$, and therefore, to avoid confusion, we indicate the vector of STAP parameters for B_{jk} with $\boldsymbol{\theta}_{j,k} = (\boldsymbol{\mu}_{j,k}, \boldsymbol{\eta}_{j,k}, \boldsymbol{\Sigma}_{j,k}, \tau_{j,k}, \rho_{j,k})$. For easiness of interpretation, we only discuss behaviours that have been observed, on average, at least once a day ($n_{j,k} > 100$), thus obtaining then 3 behaviours for each dogs, with the exception of Rosie (dog 4) that has 2 behaviours; see Tables B.1–B.6 in the Web-based Supporting Materials, where the posterior

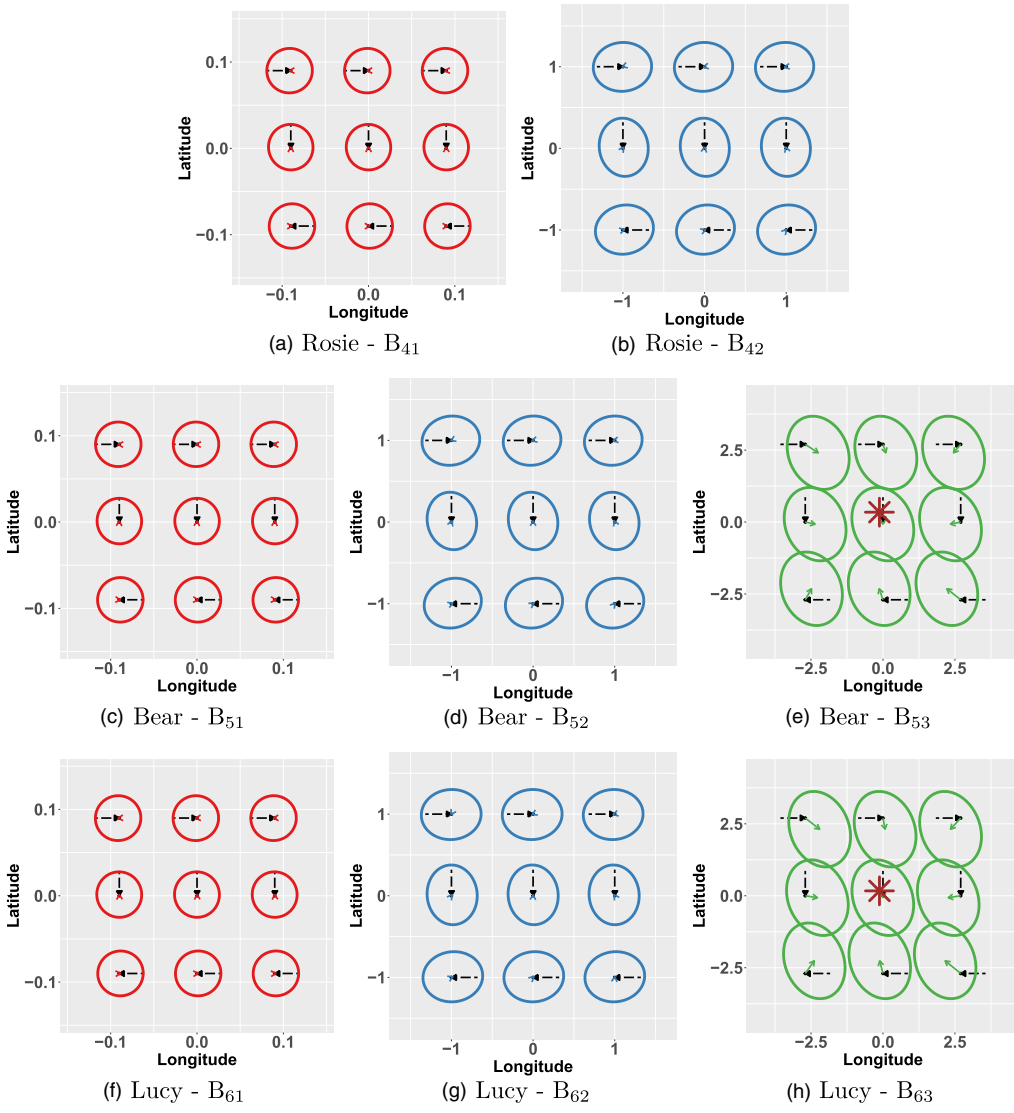


FIGURE 4 Graphical representation of the conditional distribution of $\mathbf{s}_{j,t_{j,i+1}}$ for the last three dogs ((a)(b) Rosie, (c)(d)(e) Bear, (f)(g)(h) Lucy), for different possible values of $\mathbf{s}_{j,t_{j,i}}$ and previous directions. The images has been obtained by using the posterior values that maximize the data likelihood of each animal, given the representative clusterization $\hat{z}_{j,t_{j,i}}$. The dashed arrow represents the movement between $\mathbf{s}_{j,t_{j,i-1}}$ and $\mathbf{s}_{j,t_{j,i}}$. The solid arrow is $\vec{\mathbf{F}}_{j,t_{j,i}}$, while the ellipse is an area containing 95% of the probability mass of the conditional distribution of $\mathbf{s}_{j,t_{j,i+1}}$. The asterisk represents the attractor, and it is only shown for behaviours that have posterior values of $\rho_{j,k} < 0.9$ and $\tau_{j,k} > 0.1$

means (^) and credible intervals (CIs) for the STAP parameters, $n_{j,k}$, and the transition probabilities for all the dogs and behaviours are shown. Using similar pictures to the ones used in Figure 1, we show a graphical description of the behaviours found by the model in Figures 3 and 4; the behaviours are represented on different spatial scales. From the model output, we computed the posterior mean of the variable $\delta(\chi_{j,k}, \chi_{j',k'})$, which is the posterior probability that a STAP parameter assumes the same value in B_{jk} and $B_{j'k'}$. These probabilities are depicted in

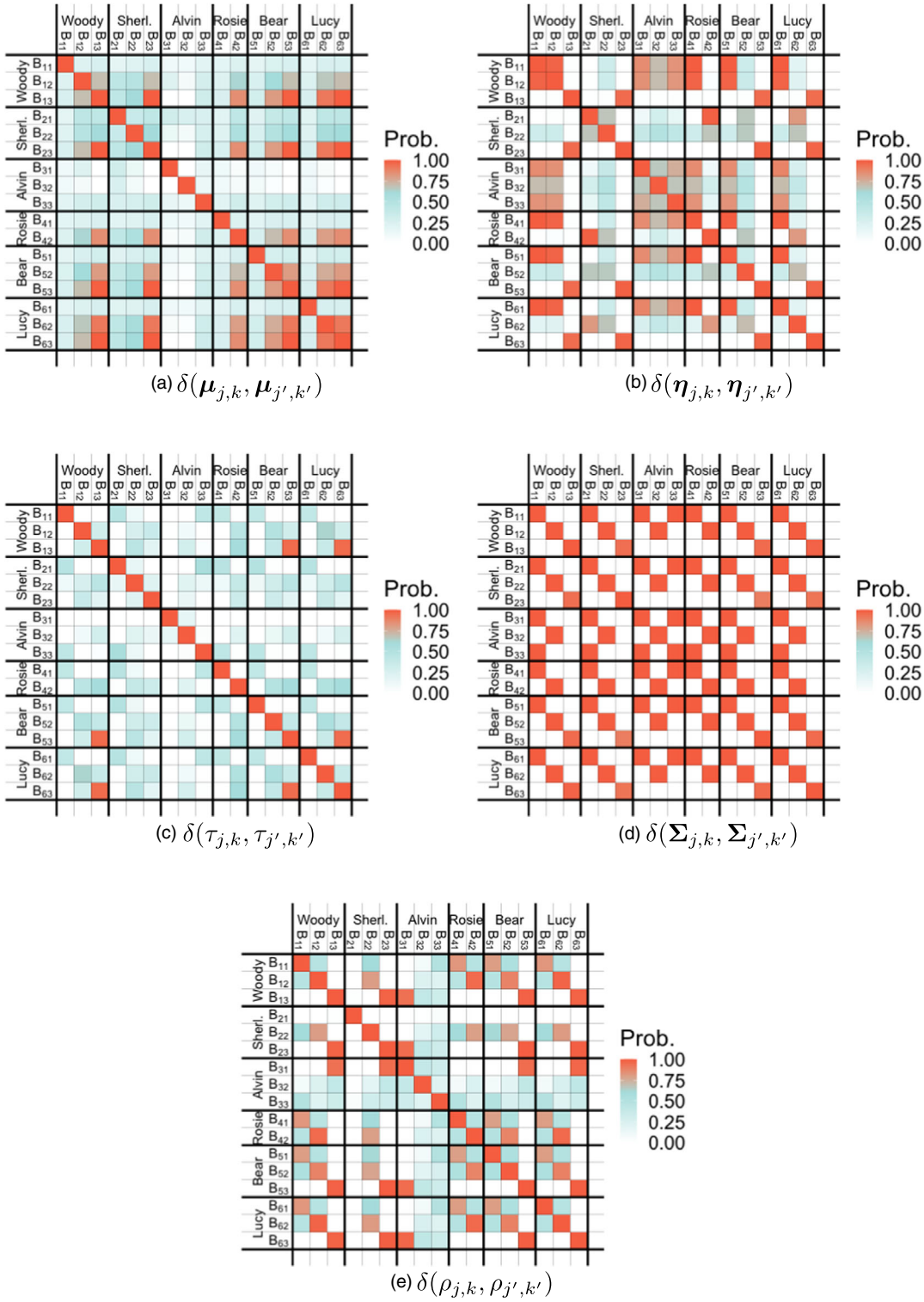


FIGURE 5 Graphical representation of the posterior mean of $\delta(\cdot,\cdot)$, which represents the probability of the parameters in its argument ((a) $\boldsymbol{\mu}_{j,k}$, (b) $\boldsymbol{\eta}_{j,k}$, (c) $\tau_{j,k}$, (d) $\boldsymbol{\Sigma}_{j,k}$, (e) $\rho_{j,k}$) having the same value for two behaviours

TABLE 2 The Adjusted Rand Index for all the pairs of dogs

	Woody	Sherlock	Alvin	Rosie	Bear	Lucy
Woody	1.000	0.094	0.006	0.030	0.181	0.414
Sherlock	0.094	1.000	−0.004	0.015	0.125	0.081
Alvin	0.006	−0.004	1.000	0.047	−0.002	0.003
Rosie	0.030	0.015	0.047	1.000	0.021	0.021
Bear	0.181	0.125	−0.002	0.021	1.000	0.159
Lucy	0.414	0.081	0.003	0.021	0.159	1.000

Figure 5 for all possible combinations of behaviours and animals. To take into account that identifiability for $\boldsymbol{\mu}_{j,k}$ and $\boldsymbol{\eta}_{j,k}$ is only granted if $\rho_{j,k} < 1$ and $\rho_{j,k} > 0$, respectively (see Equation 1), we assume $\delta(\boldsymbol{\mu}_{j,k}, \boldsymbol{\mu}_{j',k'}) = 0$, if $\rho_{j,k} = 1$ or $\rho_{j',k'} = 1$, and $\delta(\boldsymbol{\eta}_{j,k}, \boldsymbol{\eta}_{j',k'}) = 0$ if $\rho_{j,k} = 0$ or $\rho_{j',k'} = 0$, for $(i,j) \neq (i',j')$.

Similarities between the MAP behaviours. One of the assumptions of our proposal is that the temporal evolution of the behaviours are independent. To have a posterior confirmation that this hypothesis is true, we computed the Adjusted Rand Index (Gates & Ahn, 2017) for the MAP behaviours of each pair of animals. The Adjusted Rand Index, which is a measure of similarity (or agreement) between two clusterizations, has a value close to one, if there is a strong agreement, while its value is close to zero (even negative) if the clusterization is very dissimilar. It should be noted that we are able to compute the index because the animals are observed in the same temporal-window. The results in Table 2 show that the values of the index are very low, with the exception of dogs 1 and 6, where the index is 0.414. We can conclude that our hypothesis is reasonable.

The dogs in the cohesive group. We can clearly see, from Figures 3 and 4, that the four dogs that form a cohesive group (dog 1 Woody, dog 2 Sherlock, dog 5 Bear and dog 6 Lucy) have similar behaviours. In behaviour B_{j1}, the length of $\vec{\mathbf{F}}_{j,t_{j,i}}$ is ≈ 0, which means that the distribution of $\mathbf{s}_{j,t_{j,i+1}}$ is centered on the previous location ($\mathbf{s}_{j,t_{j,i}}$), there is no a preferable direction, since the ellipses are very close to a circle, the speed is very low (see the size of the ellipses), and there is no attractor. Hence, the movement is only determined by the covariance matrix $\boldsymbol{\Sigma}_{j,1}$, which is the same for all the first behaviours (B_{j1}), see Figure 5d. The first behaviour of the cohesive group can easily be interpreted as boundary-patrolling- or scent-marking-behaviour, which is a common behaviour in this dog breed, and it has already been observed and with similar movement characteristics (Mastrantonio, 2020; McGrew & Blakesley, 1982).

As in B_{j1}, the length of $\vec{\mathbf{F}}_{j,t_{j,i}}$ is ≈ 0 in B_{j2}, there is no attractor but, since the ellipses rotate according to the previous directions, there is directional persistence. The major and minor axes of the ellipses have different lengths, with the major one in the same direction as $\phi_{j,t_{j,i-1}}$, which means that we can expect to observe more movements on a straight line, that is, in the same direction as the previous bearing angle $\phi_{j,t_{j,i-1}}$, or in direction $\phi_{j,t_{j,i-1}} - \pi$. The strength of directional persistence, measured by parameter $\rho_{j,2}$, is very similar for all the B_{j2}, as we can see from Figure 5e, where the probability values are close to 1. The movement speed in B_{j2} increases compared to B_{j1}, and B_{j2} is fully characterized by $\boldsymbol{\Sigma}_{j,2}$ and $\rho_{j,2}$. The strong directional persistence, the movement along a straight line, and the higher speed can lead to B_{j2} being interpreted as a defending-behaviour, where the dog defends the territory and livestock from predators, that is, mostly wild dogs and foxes that are present in the area (see Brook et al., 2012; Walton et al., 2017),

or an explore-behaviour (van Bommel & Invasive Animals Cooperative Research Centre, 2010; Mastrantonio, 2020).

The third behaviour is a BRW, since the CIs of $\rho_{j,3}$ are very close to 0 and the ellipses are independent of the previous direction (see Figures 3 and 4). The CIs of $\tau_{j,3}$, which are in $[0.1, 0.28]$, indicate a moderate attraction to $\boldsymbol{\mu}_{j,3}$, see Tables B.1, B.2, B.5 and B.6 in the Web-based Supporting Materials. The four dogs have the same spatial attractor, $\boldsymbol{\mu}_{j,3}$, the same covariance matrix, $\boldsymbol{\Sigma}_{j,3}$, the same parameter $\rho_{j,k}$ and, with the exception of the second dog, the same $\tau_{j,k}$, as we can see from Figures 3–5. The spatial attractor, due to the large variance of the movement (the ellipses size), can be considered as a tendency of these dogs to move to the central patch of the space and, since we can see from figure c2 of van Bommel and Johnson (2012) that the attractor is close to where the livestock is, it is easy to interpret this behaviour as the dog attending livestock.

It is interesting to note that the parameters that define the three behaviours, that is, $\boldsymbol{\Sigma}_{j,k}$ in B_{j1}, $(\boldsymbol{\Sigma}_{j,k}, \rho_{j,k})$ in B_{j2}, $(\boldsymbol{\mu}_{j,k}, \tau_{j,k}, \boldsymbol{\Sigma}_{j,k}, \rho_{j,k})$ in B_{j3}, have a high probability of being the same between dogs, see Figure 5, which means that they behave in a similar manner. In transition matrix terms, we can see, from Tables B.1, B.2, B.5, B.6 in the Web-based Supporting Materials, that $\hat{\pi}_{j,1,2} > \hat{\pi}_{j,1,3}$ and $\hat{\pi}_{j,3,2} > \hat{\pi}_{j,3,1}$. Then, after patrolling (B_{j1}), it is more probable that the dog begins to explore the space, or to defend the property from predators spotted during patrolling (B_{j2}), than to guard livestock (B_{j3}). After attending livestock, it is more probable that the dog switches to B_{j2}.

The socially excluded dog and the old one. The socially excluded dog ($j = 3$) has 3 behaviours, one of which, B_{31}, is different from all the other dogs' behaviours, while the other two, B_{32} and B_{33}, are similar to the ones of the cohesive group, for example, B_{32} is similar to B_{j2} and B_{33} is similar to B_{j1}, with $j = 1, 2, 5, 6$, see Figures 3 and 4. B_{33} is only characterized by the covariance matrix $\boldsymbol{\Sigma}_{j,3}$, since like B_{j1} in the cohesive group, $\vec{\mathbf{F}}_{3,t_{3,i}}$ is ≈ 0 and there is no attractor or directional persistence. Therefore, due to the high probability of $\boldsymbol{\Sigma}_{3,3}$ having the same value as $\boldsymbol{\Sigma}_{j,3}$, with $j = 1, 2, 5, 6$, see Figure 5d, B_{33} may be interpreted in the same way as the first behaviours of the cohesive group, that is, a boundary-patrolling behaviour.

B_{32} is similar to the second behaviour of the cohesive group in terms of covariance matrix, see Figure 5d, but there is a lack of directional persistence, and a slight bias toward an attractor located in the central area, see the direction of $\vec{\mathbf{F}}_{3,t_{3,i}}$. With this behaviour, the dog is exploring but, at the same time, staying in the proximity of the sheep paddock, see Figure 2c.

The CI of $\tau_{3,1}$ is very close to 1 and B_{31} therefore has a strong attraction to the coordinates $\hat{\boldsymbol{\mu}}_{3,1} = (0.575, -381)'$, as we can see from Table B.3 of the Web-based Supporting Materials. The CIs of $\mu_{3,1,1}$ and $\mu_{3,1,2}$ are very small and equal to $(0.575, 0.576)$ and $(-0.381, 0.380)$, respectively, which means that the attractor is well localized in space. We can see from figure c2 of van Bommel and Johnson (2012), that $\hat{\boldsymbol{\mu}}_{3,1}$ is close to the owner's homestead. This behaviour cannot be interpreted as the dog attending livestock since, once it reaches the spatial attractor, it does not move very much, that is, the sizes of the ellipses are very small. Hence, this is probably a behaviour in which the dog stays close to the owner's house, and rests.

For the last dog, that is the old one, the model has found only two behaviours. B_{41} has the same characteristics as the first behaviour of the cohesive group, with the same values of the covariance matrix, while B_{42} is similar to the second behaviour of the cohesive group, with the same covariance matrix and similar directional persistence, see Figures 3–5. Hence, they can be interpreted as the first and second behaviours of the cohesive group. It is interesting to note that, since this dog is very old, she is no longer able to attend livestock, which is the activity that requires the most energy (higher speed), but she is still working, that is, checking the boundaries (B_{41}), exploring the space, and defending the territory (B_{42}).

5 | FINAL REMARKS

In this work, we have proposed a new HMM which can be used to model trajectory-tracking data of multiple animals and which, according to the classification given by Scharf and Buderman (2020), is part of the indirect approach. Our model allows subsets of parameters to be shared between animals and behaviours, and the number of latent behaviours to be selected during the model fitting. The emission distribution is the STAP, but other distributions can be used by changing the model formalization accordingly. The model was used to help understand the behaviour of 6 Maremma Sheepdogs, observed in a property in Australia. The results show that there are many common features between the animals, such as the attractive point, and most of them share the same number of behaviours as well as the same parameter values. The obtained results are easily interpretable, and the rich output offers an insight into the similarities between animals.

As a possible extension, we are currently exploring the use of covariates to model the probability that behaviours share parameters, and we are working on a different formalization that makes the model able to detect whether different animals tend to follow the same behaviours at the same time-points.

ACKNOWLEDGEMENTS

The authors thank the Editor-in-Chief, the Associate Editor and the two anonymous reviewers for their comments that have greatly improved the manuscript. The work of the author has partially been developed under the MIUR grant Dipartimenti di Eccellenza 2018—2022 (E11G18000350001), conferred to the Dipartimento di Scienze Matematiche—DISMA, Politecnico di Torino.

DATA AVAILABILITY STATEMENT

The dataset is freely available from the movebank repository https://www.datarepository.movebank.org/handle/10255/move.395

ORCID

Gianluca Mastrantonio https://orcid.org/0000-0002-2963-6729

REFERENCES

Anderson, C.R. & Lindzey, F.G. (2003) Estimating cougar predation rates from GPS location clusters. *The Journal of Wildlife Management*, 67(2), 307–316.

Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2017) Julia: a fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98.

Biernacki, C., Celeux, G. & Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719–725.

Blackwell, P. (1997) Random diffusion models for animal movement. *Ecological Modelling*, 100(1), 87–102.

van Bommel, L. & Invasive Animals Cooperative Research Centre. (2010). Guardian dogs: best practice manual for the use of livestock guardian dogs. *Invasive Animals Cooperative Research Centre*.

van Bommel, L. & Johnson, C.N. (2012) Good dog! Using livestock guardian dogs to protect livestock from predators in Australia's extensive grazing systems. *Wildlife Research*, 39(3), 220–229.

van Bommel, L. & Johnson, C. (2014a) Data from: where do livestock guardian dogs go? *Movement Patterns of Free-Ranging Maremma Sheepdogs*, https://doi.org/10.5441/001/1.pv048q7v

van Bommel, L. & Johnson, C.N. (2014b) Where do livestock guardian dogs go? Movement patterns of free-ranging maremma sheepdogs. *PLoS ONE*, 9(10), 1–12.

van Bommel, L. & Johnson, C.N. (2016) Livestock guardian dogs as surrogate top predators? How Maremma sheepdogs affect a wildlife community. *Ecology and Evolution*, 6(18), 6702–6711.

Brook, L.A., Johnson, C.N. & Ritchie, E.G. (2012) Effects of predator control on behaviour of an apex predator and indirect consequences for mesopredator suppression. *Journal of Applied Ecology*, 49(6), 1278–1286.

Buderman, F.E., Hooten, M.B., Alldredge, M.W., Hanks, E.M. & Ivan, J.S. (2018) Time-varying predatory behavior is primary predictor of fine-scale movement of wildland-urban cougars. *Movement Ecology*, 6(1), 22.

Calabrese, J.M., Fleming, C.H., Fagan, W.F., Rimmler, M., Kaczensky, P., Bewick, S. et al. (2018) Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746), 20170007.

Celeux, G., Forbes, F., Robert, C.P., Titterington, D.M., Futurs, I. & Rhône-alpes, I. (2006) Deviance information criteria for missing data models. *Bayesian Analysis*, 4, 651–674.

Christ, A., Hoef, J.V. & Zimmerman, D.L. (2008) An animal movement model incorporating home range and habitat selection. *Environmental and Ecological Statistics*, 15(1), 27–38.

Codling, E.A., Plank, M.J. & Benhamou, S. (2008) Random walk models in biology. *Journal of the Royal Society Interface*, 5(25), 813–834.

Dunn, J.E. & Gipson, P.S. (1977) Analysis of radiotelemetry data in studies of home range. *Biometrics*, 33(1), 85–101.

Fortin, D., Morales, J.M. & Boyce, M.S. (2005) Elk winter foraging at fine scale in Yellowstone National Park. *Oecologia*, 145(2), 334–342.

Fox, E.B., Sudderth, E.B., Jordan, M.I. & Willsky, A.S. (2011) A sticky HDPHMM with application to speaker diarization. *The Annals of Applied Statistics*, 5(2A), 1020–1056.

Frühwirth-Schnatter, S. & Malsiner-Walli, G. (2019) From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification*, 13(1), 33–64.

Gates, A.J. & Ahn, Y.-Y. (2017) The impact of random models on clustering similarity. *Journal of Machine Learning Research*, 18(87), 1–28.

Gehring, T.M., VerCauteren, K.C. & Cellar, A.C. (2017) Good fences make good neighbors: implementation of electric fencing for establishing effective livestock-protection dogs. *Human-Wildlife Interactions*, 5(1), 106–111.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2013). *Bayesian data analysis*, 3rd edition. Boca Raton, FL: Chapman and Hall/CRC.

Gnedin, A., Gnedin, E. & Kerov, S. (2001) A characterization of GEM distributions. *Combinatorics, Probability and Computing*, 10, 213–217.

Hastie, D.I. & Green, P.J. (2012) Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica*, 66(3), 309–338.

Hebblewhite, M. & Merrill, E. (2008) Modelling wildlife and uman relationships for social species with mixed-effects resource selection models. *Journal of Applied Ecology*, 45(3), 834–844.

Hooten, M., Johnson, D., McClintock, B. & Morales, J. (2017). *Animal movement: statistical models for telemetry data*. Boca Raton, FL: CRC Press.

Hooten, M.B., Scharf, H.R., Hefley, T.J., Pearse, A.T. & Weegman, M.D. (2018) Animal movement models for migratory individuals and groups. *Methods in Ecology and Evolution*, 9(7), 1692–1705.

Jammalamadaka, S.R. & Kozubowski, T.J. (2004) New families of wrapped distributions for modeling skew circular data. *Communications in Statistics - Theory and Methods*, 33(9), 2059–2074.

Jonsen, I. (2016) Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Scientific Reports*, 6(1), 20625.

Jonsen, I.D., Flemming, J.M. & Myers, R.A. (2005) Robust state-space modeling of animal movement data. *Ecology*, 86(11), 2874–2880.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. & Morales, J.M. (2012) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11), 2336–2342.

Leos-Barajas, V., Gangloff, E.J., Adam, T., Langrock, R., van Beest, F.M., Nabe-Nielsen, J. et al. (2017) Multi-scale modeling of animal movement and general behavior data using hidden Markov Models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3), 232–248.

Maruotti, A., Punzo, A., Mastrantonio, G. & Lagona, F. (2016) A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. *Stochastic Environmental Research and Risk Assessment*, 30, 1725–1740.

Mastrantonio, G. (2018) The joint projected normal and skew-normal: a distribution for poly-cylindrical data. *Journal of Multivariate Analysis*, 165, 14–26.

Mastrantonio, G. (2020) Modelling animal movement with directional persistence and attractive points. *arXiv*. 2012.03248.

Mastrantonio, G., Jona Lasinio, G., Pollice, A., Teodonio, L. & Capotorti, G. (2021) A Dirichlet process model for change-point detection with multivariate bioclimatic data. *Environmetrics*, 33, e2699.

McClintock, B.T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B.J. & Morales, J.M. (2012) A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, 82(3), 335–349.

McClintock, B.T., Russell, D.J.F., Matthiopoulos, J. & King, R. (2013) Combining individual animal movement and ancillary biotelemetry data to investigate populationlevel activity budgets. *Ecology*, 94(4), 838–849.

McGrew, J.C. & Blakesley, C.S. (1982) How Komondor dogs reduce sheep losses to coyotes. *Journal of Range Management*, 6(35), 693–696.

Merrill, S.B. & David Mech, L. (2000) Details of extensive movements by Minnesota wolves (*Canis lupus*). *The American Midland Naturalist*, 144(2), 428–433.

Michelot, T., Langrock, R. & Patterson, T.A. (2016) moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, 7(11), 1308–1315.

Michelot, T., Langrock, R., Bestley, S., Jonsen, I.D., Photopoulou, T. & Patterson, T.A. (2017) Estimation and simulation of foraging trips in land-based marine predators. *Ecology*, 98(7), 1932–1944.

Milner, J.E., Blackwell, P.G. & Niu, M. (2021) Modelling and inference for the movement of interacting animals. *Methods in Ecology and Evolution*, 12(1), 54–69.

Niu, M., Frost, F., Milner, J.E., Skarin, A. & Blackwell, P.G. (2020) Modelling group movement with behaviour switching in continuous time. *Biometrics*, 1–14. Available from: https://doi.org/10.1111/biom.13412

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008) State-space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2), 87–94.

Pohle, J., Langrock, R., van Beest, F.M. & Schmidt, N.M. (2017) Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3), 270–293.

Scharf, H.R. & Buderman, F.E. (2020) Animal movement models for multiple individuals. *WIREs Computational Statistics*, 12, e1506.

Teh, Y.W., Jordan, M.I., Beal, M.J. & Blei, D.M. (2006) Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476), 1566–1581.

Wade, S. & Ghahramani, Z. (2018) Bayesian cluster analysis: point estimation and credible balls (with discussion). *Bayesian Analysis*, 13(2), 559–626.

Walton, Z., Samelius, G., Odden, M. & Willebrand, T. (2017) Variation in home range size of red foxes *Vulpes vulpes* along a gradient of productivity and human landscape alteration. *PLoS ONE*, 12(4), 1–14.

Westley, P.A.H., Berdahl, A.M., Torney, C.J. & Biro, D. (2018) Collective movement in ecology: from emerging technologies to conservation and management. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746), 20170004.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mastrantonio, G. (2022) The modelling of movement of multiple animals that share behavioural features. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–19. Available from: https://doi.org/10.1111/rssc.12561