Scalable clustering of trajectories within a continuous time framework

Abstract

In this paper, we propose a new modelling framework to cluster sequences of geographical positions through time (such as GPS data) – or trajectories – at potentially irregular frequencies. The model is specified within a continuous time framework, being robust to irregular sampling in records and accounting for possible heterogeneous movement patterns within a single trajectory. The clustering is performed using non parametric Bayesian methods, namely the hierarchical Dirichlet process, allowing a clustering of both individuals' movement patterns and trajectories in the data. In order to propose a scalable method, the inference is done using stochastic variational inference in an easy-to-distribute framework. The performances are studied on both simulated data sets and a large trajectory data set from the Automatic Identification System (AIS), used to monitor the world marine traffic.

1 Introduction

For the last thirty years, the tracking of movements of individuals or objects has been eased by the increasing development of several tracking devices, such as the Global Positioning System (GPS) that allows recording geographical positions through time. Such devices have led to large movement databases that store trajectory data, in which each point represents a position in a space at a given time. These databases contain a great deal of knowledge and require analysis [5] such as the extraction of movement patterns across diverse trajectories in order to perform human mobility or traffic monitoring [14, 32], motion prediction [23], human action recognition [29] or detection of abnormal behavior in maritime routes [26]. A striking example of large movement database comes from the worldwide marine traffic surveillance. This monitoring relies on several sources of data, in a rising context of maritime big data [7]. Among these sources lies the Automatic Identification System (AIS), which automatically collects messages from vessels around the world, at a high frequency. AIS data basically consist in GPS-like data, together with the instantaneous speed and heading, and some vessel specific static information [4].

Extracting significant movement patterns within trajectories is then of prime interest but particularly challenging because of their diversity as they may (1) be collected at different frequencies (2) have different lengths (3) be regularly sampled or not, and (4) represent very different behaviors. Approaches to cluster those trajectories have to be defined, being able to handle such complex data and being efficient on large databases.

In this purpose, existing trajectory clustering methods follow two main trends. The first class of methods is a density-based clustering framework that relies on the definition of typical motions in the trajectories [6, 13, 20]. Similarities between full trajectories or sub-sequences are considered to take into account local details in the segments. As a consequence, these methods only use the geometry of the trajectories, and can consider extra information, as the speed of the trajectory for instance, but with difficulty. The second class of methods is a rule-based clustering framework, that relies on a pre-conceived decision rules to build each cluster. Trajectories within a cluster are then used to estimate a normality model using kernel density methods over observations (*e.g.* [18] in the clustering of AIS data context). The main advantage is of these methods is their simplicity, but they suppose the prior knowledge of decision rules, and afterwards make strong independence assumption on observed points, that might not hold in practice. They also assume that all observations within a cluster are identically distributed, omitting the heterogeneous nature of movement data [15]. In a more parametric approach, [28] proposed the definition of a normality model for trajectories through an unsupervised clustering framework inspired by topic models [22]. For an in-depth review of spatial trajectory clustering algorithms, one can refer to [31].

In text analysis, topic models originally aim at classi-

Preliminary work. Under review by AISTATS 2019. Do not distribute.

fying documents depending on their main topics. The clustering approach then estimates the different topics in each document as well as clustering documents together regarding their topics. In their transposition of topic models to trajectory analysis, [28] consider trajectories as documents, whose words are quantized GPS observations (position and velocity) and it is assumed that each point of the trajectory belongs to a semantic region (a topic), called hereafter a *movement mode*. A movement mode is therefore a specific distribution to be estimated from trajectory data. Trajectories are then considered as a mixture of movement modes, and a trajectory cluster is a set of trajectories with a common mixture distribution. Estimation of both movement modes and trajectory clusters can be performed on a discretized space within a Bayesian framework using Gibbs sampling. If this approach still assumes independence of GPS observations (conditionally to their semantic region), it strongly differs from [18] by taking into account different heterogeneity levels for movement data. However, it requires a quantization of the space whose influence over results is not investigated, and relies on Gibbs sampling for inference, which is known to be hardly scalable [3].

Following these works, we here define a novel normality model, based on topic models, for clustering large and complex trajectory data sets in a scalable framework. This model distinguishes from existing work in the definition of GPS observations distributions as it is now assumed that (1) the distribution lives in a continuous space, avoiding the quantization of the (position/velocity) space, and (2) within a movement mode (or a topic), observations are not independent. Each movement mode is modeled through the bivariate velocity as a continuous time Gaussian process, namely the Ornstein Uhlenbeck Process (OUP) [25], that has been proposed in movement ecology as a flexible framework to model the fundamental units of movement [9]. Thanks to its continuous time formulation, the OUP allows handling irregularly sampled data. Moreover, as a Gaussian process, it accounts for observation autocorrelation. The unsupervised clustering is made in a non parametric Bayesian context using a Hierarchical Dirichlet Process framework [24]. To perform movement mode estimation in a scalable approach, we use asymptotic properties of the OUP together with stochastic variational inference [10]. Contrary to MCMC methods that obtain samples from the posterior distribution, the variational approach solves an optimization problem, allowing the use of distributed computation and stochastic optimization to scale and speed up inference [3].

The remainder of this paper is organized as follows: in Section 2, the hierarchical parametric framework to model trajectory data is fully specified, and the proposed scalable approach to estimate model parameters from data is then described. The method is evaluated in Section 4, both on simulated data sets, to assess inference performances, and on actual AIS data, to show the interest of our approach in a realistic context.

2 Movement model

In this section, we define a parametric framework to model trajectory data, *i.e.* sequences of geographical positions recorded through time. The modelling framework aims to account for two levels of heterogeneity possibly present in trajectory data: (1) heterogeneity of an individual's movement within a single trajectory, and (2) heterogeneity between observed trajectories of several individuals.

In the following, the fundamental modelling of movement is made through the individual's velocity process. Velocity is observed directly (as in the AIS context) or is at least computed from the successive geographical locations of the individuals. The positions and velocities of an individual are modeled as two continuous time processes in $\mathbb{R}^2$, denoted by $(X_t)_{t\geq 0}$ and $(V_t)_{t\geq 0}$. Following a common paradigm, we assume that a moving individual's trajectory might be the collection of heterogeneous patterns [15], namely the *movement modes*. Different movement modes along a trajectory refer to different ways of moving in terms of velocity distribution, reflecting different behaviors, activities, or routes. It is assumed that the set of possible movement modes of tracked individuals during their trajectories is countable, and that a given movement mode can be adopted by several individuals.

2.1 Fundamental unit of movement

Similarly to [9], a movement mode is assumed to be characterized by a specific correlated velocity model. Formally, during a time segment $[\tau_1;\tau_2]$, if an individual adopts the movement mode k, then its velocity process $(V_t)_{\tau_1\leq t\leq\tau_2}$ is assumed to be the solution of the following Stochastic Differential Equation (SDE) (see [17] for a formal definition of SDEs):

$$\mathrm{d}V_t = -\Gamma_k\left(V_t-\mu_k\right)\mathrm{d}t + \Sigma_k \mathrm{d}W_t, \qquad V_{\tau_1} = v_{\tau_1}, \quad \tau_1 \leq t \leq \tau_2, \quad (1)$$

where:

- $\mu_k \in \mathbb{R}^2$ is the asymptotic mean velocity of the k-th movement mode;
- Γ_k is a 2×2 real-valued matrix, and is an autocorrelation parameter, modelling the recall force of the process towards the mean μ_k;

- Σ_k is a 2×2 real-valued matrix, and is a diffusion term, modelling the variability of the process around the mean μ_k;
- W_t is a bivariate standard Brownian motion, modelling the stochasticity of the individual's velocity process;
- v_{τ_1} is the initial condition of the SDE: the individual's velocity at time τ_1.

The solution to Equation (1) is a well known continuous time stochastic process, the Ornstein Uhlenbeck Process (OUP) [25], also known as the *mean reverting* process. This mean reverting property is due to the recall force Γ_k and makes the OUP suitable to describe movement modes of individuals, that are often characterized by a mean velocity regime, which is reached by rapid and brutal accelerations (or decelerations, see Figure 1). The OUP satisfies the three following properties: (1) $(V_t)_{\tau_1 \leq t \leq \tau_2}$ is a continuous time Markov process, (2) $(V_t)_{\tau_1 \leq t \leq \tau_2}$ is a Gaussian process, *i.e.* the law of the random variable $V_t | V_{\tau_1} = v_{\tau_1}$ is Gaussian, with explicit mean and covariance (given in Appendix A) and (3) if both eigenvalues of Γ_k are positive, then $(V_t)_{t \geq \tau_1}$ is an asymptotically stationary Gaussian Process, *i.e.*, the law of $V_t | V_{\tau_1} = v_{\tau_1}$ satisfies:

$$V_t | V_{\tau_1} \overset{t \to \infty}{\longrightarrow} \mathcal{N}(\mu_k, \Lambda_k), \tag{2}$$

where Λ_k is the asymptotic covariance matrix of the OUP (given in Appendix A).

As discussed in [9], the OUP offers a flexible framework to model wide range of autocorrelated velocity processes, such as highly directed movements or rotational ones. Knowing the individual's velocity process, its position process, starting at position x_{τ_1}, is obtained by deterministic integration over time. The resulting process $(X_t)_{\tau_1 \leq t \leq \tau_2}$ is known as an Integrated Ornstein Ulhenbeck Process (IOUP), which remains a Gaussian Process.

2.2 Within trajectories heterogeneity

To account for heterogeneous ways of moving during a single trajectory, we assume that an individual's trajectory in a time segment $[0, T]$ is a sequence of IOUP. Formally, there exists a sequence of times $\tau_0 = 0 < \tau_1 < \cdots < \tau_K = T$ such that the trajectory is a sequence of K successive IOUP. A simulated example of such a sequence is shown on Figure 2. Therefore, a trajectory is characterized by (1) the movement modes adopted within the trajectory, (2) the time spent by the tracked individual within each movement mode and (3) the transitions from one movement mode to another. Here, we don't impose specific modelling for the last two points. Our first goal is then to estimate the different movement modes within trajectories present in a given data set.

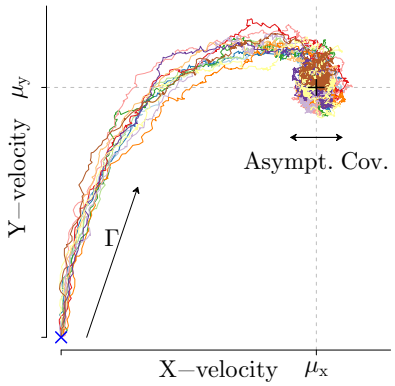

Figure 1: Realizations of bivariate Ornstein Ulhenbeck Processes of parameters μ, Γ and Σ, solution of the SDE Eq.(1), starting at the blue cross.

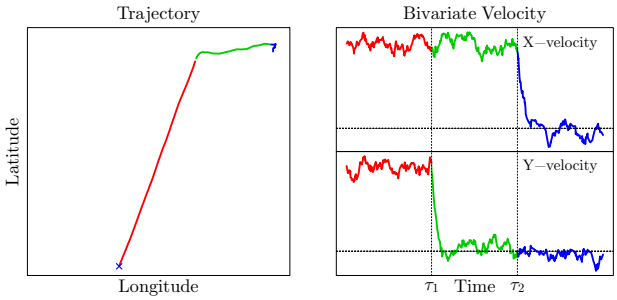

Figure 2: (Left) Simulation of a sequence of Integrated Ornstein Uhlenbeck Processes defining a trajectory. The blue cross marks the starting position. The trajectory is made of three movement modes (identified by their colors), characterized by three different sets of parameters for the velocity process from Eq. (1). (Right) Bivariate velocity processes, together with the instants τ_1 and τ_2 at which parameters change (vertical dotted lines). Horizontal dotted lines mark the 0 value.

2.3 Between trajectories heterogeneity

In addition to this first clustering problem, we assume that the data set is composed of different (unknown) clusters of trajectories. A cluster is characterized by similar distributions over movement modes. In other words, two trajectories belong to a same cluster if they are composed of the same movement modes in similar proportions. This second clustering problem is illustrated in Figure 3. Our second goal is to cluster together such trajectories.

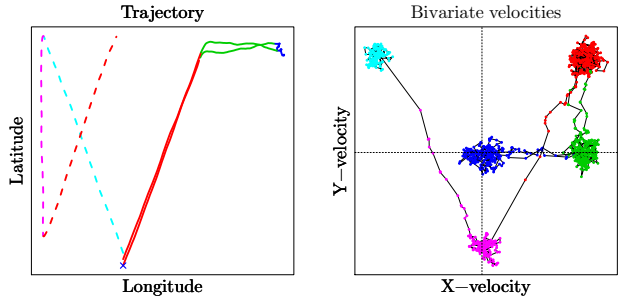

Figure 3: Simulation of three different trajectories with the same model as in Figure 2. (Left) The blue cross marks their shared starting point. The two plain trajectories are made of the same movement modes (red, green and blue) so they belong to the same cluster. The dashed line trajectory has a different distribution over movement modes (light blue, purple and red) so it belongs to another cluster. (Right) The trajectories are plotted on the bivariate velocity space, showing the five possible movement modes. One of them (red) is shared by the two trajectory clusters. The vertical and horizontal lines show the 0 value.

To summarize, the issues to be addressed in an unsupervised and scalable approach are (1) characterizing different movement modes present in a large data set, (2) for each point of the trajectory, estimating in which movement modes it belongs and (3) clustering together trajectories having the same distribution over movement modes.

3 Scalable two step clustering

In the previous section, we defined a realistic framework to model movement data, introducing a continuous time autocorrelated model over velocity process. The resulting framework therefore consists in a continuous time regime switching diffusion model. Estimating both parameters and movement mode assignation of data is in this case a statistical challenge, even in the case of a single trajectory [19]. For instance, [2] use thinning Poisson process and MCMC methods to perform parameter estimation from movement data, resulting in highly time demanding approach. A big challenge for estimation methods therefore remains their scalability, especially in the dual clustering context that we aim here.

In order to perform scalable parameter inference and clustering of both trajectories and GPS observations (into movement modes), we adopt a pragmatic two step approach that takes advantage of the inherent properties of the OUP and are described with more details hereafter:

1. A first dual clustering is performed based on a simpler independent Gaussian mixture model, in order to estimate potential movement modes and trajectory clusters: it allows getting rid of within mode autocorrelation in the inference, and therefore eases the computations. The Gaussian hypothesis comes from the fact that the OUP stationary distribution is Gaussian.

2. Among the estimated movement modes, only those meeting a temporal consistency constraint are kept. Parameters of these consistent movement modes are then estimated, and used to reassign observations that were assigned to inconsistent movement modes. It ensures that only trajectory segments for which this stationary distribution was reached are kept to estimate movement modes.

3.1 Step 1: simplified model based on an independence assumption

3.1.1 Hierarchical model

The data set is composed of J independent trajectories that we aim to cluster in C groups, and each trajectory j gathers n_j velocities $(V_i^j)_{i=1,\dots,n_j}$. In this simplified model, we assume (as done in [28]) that conditionally to their movement mode, all the velocities are mutually independent. This strong hypothesis eases the computation and ensures the method scalability as a first step, and will be relaxed in the second step. Moreover, we suppose that there exists K (possibly infinite) movement modes. The set of unknown movement parameters $\{(\mu_k, \Lambda_k)\}_{k=1\cdots K}$ is denoted $\mathcal{T}$.

We assume that trajectories in the data set are drawn from a mixture of clusters with mixture distribution with weights $\boldsymbol{w} = [w_1, \cdots w_C]$. In each trajectory j, the within trajectory heterogeneity is modelled through a Gaussian mixture whose parameters are drawn from $\mathcal{T}$ with mixture weights $\boldsymbol{\pi}^j = \left[\pi_1^j, \cdots, \pi_K^j\right]$. It is then assumed that two trajectories j_1 and j_2 belong to a same trajectory cluster c if $\boldsymbol{\pi}^{j_1} = \boldsymbol{\pi}^{j_2} = \boldsymbol{\pi}^c$. The movement parameter set $\mathcal{T}$, the trajectory clusters weights $\boldsymbol{w}$ and, for each cluster c, the movement mode weights $\boldsymbol{\pi}^c$ are the *global variables* (reusing the terminology used in [10]) of the model that have to be estimated.

Moreover, for each trajectory j, the cluster label F_j has to be estimated. Finally, for each observation V_i^j, the movement mode label Z_i^j has to be estimated. Sets $\boldsymbol{F} = \{F_j\}_{j=1,\dots,J}$ and $\boldsymbol{Z} = \left\{Z_i^j\right\}_{\substack{j=1,\dots,J \\ i=1,\dots,n_j}}$ are the *local variables* of the model to be estimated.

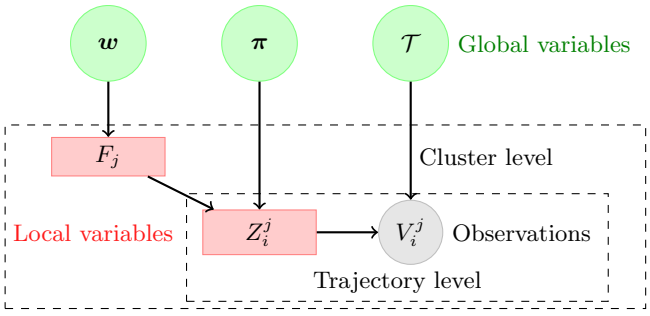

Figure 4: Graphical representation of the hierarchical structure of the simplified model of Section 3.1.

The resulting hierarchical structure of this model is shown on Figure 4.

3.1.2 Bayesian estimation of the parameters using stochastic variational inference

In the first step, the inference of the local and global variables is made within a Bayesian context, considering hidden variables as random variables. The scalable approach to estimate both local and global variables from GPS data is now detailed.

A first problem to address is the number of possible movement modes and trajectory clusters. The chosen approach here is to use a Bayesian non parametric approach, that "side-steps [the model selection problem] by allowing the data to determine the complexity of the model" [8]. More precisely, we use a Hierarchical Dirichlet Process framework [24], movement modes and trajectory clusters both being drawn from an infinite mixture distribution. The constructive stick breaking distribution approach [21] is used to get the infinite mixture distribution priors: an infinite vector of weights (with positive values and summing to 1) is defined from a set of independent beta distributed random variables. The prior distributions for the other global and local variables are the classical conjugate priors in Gaussian mixture models. This conjugacy property is essential in our scalable approach objective. All the details about prior specifications are given in Appendix B. In a Bayesian context, the inference target is the posterior density $p(\mathcal{T}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{F}, \boldsymbol{Z}|\boldsymbol{V})$, where $\boldsymbol{V}$ is the set of all observed velocities. However, for such model as ours, the computation of this density is not tractable. Instead of seeking the expression of this posterior density, an alternative is to obtain samples from this posterior distribution. This can be done using MCMC methods, that can be implemented for Dirichlet processes [16], and were used in [28] for the inference in their discrete-space framework.

A more scalable alternative to this simulation-based approach is the use of variational inference (see [3] for a recent review). Variational methods reduce the inference to an optimization problem by minimizing a divergence (typically the Kullback-Leibler divergence [12]) between the target posterior distribution and the members of a simpler family of distributions, the *variational* family. An appropriate and common variational family $\mathcal{Q}$ is the one which satisfies the mean field assumption, *i.e.* the set of q distributions that can be fully factorized. Our variational family is therefore such that:

$$q(\mathcal{T}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{F}, \boldsymbol{Z}) = q(\boldsymbol{w}) \prod_c q(\boldsymbol{\pi}^c) \prod_k q(\mu_k, \Lambda_k) \prod_j q(F_j) \prod_{i,j} q(Z_i^j). \quad (3)$$

The mean field inference problem reduces to find q^* such that:

$$q^* = \underset{q \in \mathcal{Q}}{\operatorname{argmax}} \{\mathbb{E}_q [\log p(\mathcal{T}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{F}, \boldsymbol{Z}, \boldsymbol{V})] - \mathbb{E}_q [\log q(\mathcal{T}, \boldsymbol{w}, \boldsymbol{\pi}, \boldsymbol{F}, \boldsymbol{Z})]\}, \quad (4)$$

where $\mathbb{E}_q[\cdot]$ denotes the expectation with respect to the p.d.f. q. The right hand side of (4) is known as the evidence lower bound (ELBO), and can be computed for appropriate family of distributions $\mathcal{Q}$, and thus can be maximized.

In addition to this mean field property, the variational approximation of the posterior distribution is restricted to finite sets of parameters. Formally, the posterior distributions of the infinite cluster (resp. movement mode) weights are approximated by a distribution on a finite set of weights having C_{max} (resp. K_{max}) elements. Here, C_{max} and K_{max} are variational parameters given by the user. This variaional approximation is known as the truncated stick breaking distribution [10] (see Appendix B for details).

A known algorithm to compute q^* defined by (4) is the coordinate ascent variational inference (CAVI [1]).

This iterative algorithm starts from an initial guess $q^{(0)}$ for the optimal variational distribution, and successively updates each of its components by supposing the others known, and computes an expectation with respect to their distribution. For the model presented in section 3.1 and the chosen prior distributions, all needed expectations can be computed explicitly (and are shown in Appendix B).

As described in [10], this algorithm can be seen as a gradient ascent, and is therefore suitable for stochastic approximations, resulting in the stochastic variational inference (SVI). The estimation algorithm therefore reduces to (1) sample uniformly a batch from the data set, (2) compute the expectations computed in

appendix using only this batch (3) update variational distributions using these expectations, as described in [10]. This procedure can be performed *online*, therefore with no need of stocking the data [27]. Performances of these stochastic methods are widely discussed in [27] and [10].

It is worth noting here that all the needed computations (shown in appendix) can naturally be distributed (due to the independance simplification), as they are essentially a sum over simple operations involving single observations. Therefore, the SVI algorithm proposed here for the simplified model is widely scalable, unlike Gibbs sampling procedures.

As the optimization is done in a high dimensional space, and the algorithm only guarantees convergence to a local optimum, it is crucial to initialize the algorithm from different starting points, to ensure a good exploration of the space [1]. Again, different runs of the algorithm can be distributed.

3.2 Step 2: Estimation of OUP parameters from clustering outputs

This first step gives as an output the set of optimal variational distributions, namely, $q^*_{\mathcal{T}}$, $q^*_{\boldsymbol{w}}$ and $q^*_{\boldsymbol{\pi}}$, the posterior distribution of all models global variables, $q^*_{\boldsymbol{F}} = \left\{q^*_{F_j}\right\}_{j=1,\dots J}$, for each trajectory j, the posterior probability of being in each possible trajectory cluster, and $q^*_{\boldsymbol{Z}} = \left\{q^*_{Z_i^j}\right\}_{\substack{j=1,\dots J \\ i=1,\dots n_j}}$, for each observation i of trajectory j, the posterior probability of being in each possible movement mode. From the last two distributions, one can get estimators of clusters and movement modes present in the data. A classical estimator would be the maximum a posteriori (MAP), *i.e.* the cluster (resp. movement mode) label giving the maximum weight of the posterior multinomial distribution $q^*_{F_j}$ (resp. $q^*_{Z_i^j}$) .

In order to estimate the OUP parameters of the k-th movement mode, as defined in section 2, a filter is applied on movement mode sequences based on their temporal consistency. Formally, for a trajectory j, let us consider a sequence of m successively recorded velocities $V^j_{i_1}, \dots, V^j_{i_m}$, that were all estimated to belong to a same movement mode k. This sequence is said to be temporally consistent if its length $t_{i_m} - t_{i_1}$ is larger than δ, a user chosen parameter, representing the minimal time lag for a movement mode. All consistent sequences estimated in a same movement mode are considered as independent realizations of OUP with common parameters. From the Markov and Gaussian properties of the OUP, the likelihood related to this data set can be easily maximized to obtain estimates of parameters μ_k, Γ_k and Σ_k (see Appendix A). If an estimated movement mode k (*i.e.* a movement mode containing at least one observation) has no consistent sequence, this movement mode is considered as inconsistent. Finally, to refine the movement modes classification, observation sequences belonging to inconsistent movement modes are reallocated to the consistent movement mode whose parameters maximize their likelihood.

This second step therefore depends on a parameter δ, whose value is surely data dependent, but has a clear and easy interpretation. The resulting consistent movement mode concept allows one i) to have a good estimation of OUP parameters within a movement mode (as a consistent sequence will often be related to a large amount of points) and ii) to filter out "noise" movement modes gathering few observations in a temporally inconsistent manner.

4 Experiments

4.1 Experiments on simulated data

To ensure that the inference proposed in Section 3 suits to the model defined in Section 2, a simulated experiment is performed.

Simulation set up. A data set of 40 trajectories containing overall 22 000 observations is simulated, in a model containing 2 trajectory clusters, each cluster being composed of 2 movement modes.

Within each movement mode, velocities are drawn from an OUP whose parameters are movement mode dependent. The sequence of movement modes is drawn using a continuous time hidden semi-Markov chain [30] (see Appendix C for details) whose transition parameters depend on the trajectory cluster. Simulated trajectories are depicted in Figure 5.

Movement mode estimation: results after the first step. A first clustering is performed using the variational inference approach within a hierarchical Dirichlet process, as described in Section 3.1. The variational approximation of the stick breaking distributions are made with truncated breaking distribution with $C_{max} = 5$ and $K_{max} = 12$. The optimization is made independently from 200 randomly chosen starting points, performing 1000 iterations of the SVI, with a decreasing gradient step along iterations. The best estimate is chosen as the final point maximizing the ELBO defined in eq. (4). The estimated movement

[1] In this purpose also, it should be pointed out that any stochastic approach is better than the fully deterministic CAVI algorithm, as discussed in [11], for instance.

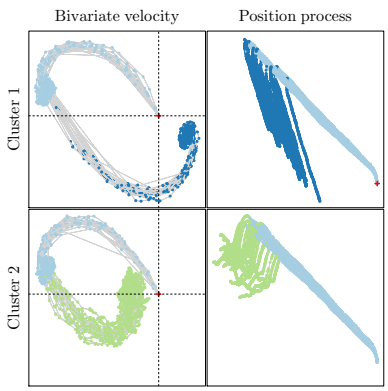

Figure 5: Simulated trajectory data. Three movement modes are shared by two clusters. Each cluster corresponds to two different roads. The "light blue" movement mode (top of the trajectories) is shared by the two clusters.

mode for a velocity V_i^j is computed as the MAP of the weights distribution $q^*_{Z_i^j}$. From the estimation, it results that among the $K_{max} = 12$ possible movement modes, only 4 of them contain at least one observation. Overall, 96.9% of the points are attributed to the correct cluster (Table S1 in Appendix D). However, an extra movement mode is estimated, corresponding to the transition phases towards the light blue movement mode and leaving from it (see Figure 6).

It is worth noting here that this first phase is a continuous space version of the work of [28]. The presented approach, however, do not suffer from choosing a space quantization grid.

Movement mode estimation: results after the second step. Including the second step described in Section 3.2, movement modes for inconsistent sequences are re-estimated. Using k-consistent sequences, OUP parameters are estimated and used to reallocate inconsistent ones. This results in a good reallocation of the problematic transitory phases having now an overall 99.7% good classification rate, as shown in Figure 6 (and Table S2 in Appendix D). One can see here that coming back to the OU property enables correct classification of transitory phases of the movement.

Trajectory clusters estimation. As expected on such a simple example, the cluster assignation is 100% right. It is worth noting here that it would not be the case on trickier examples (not shown here), for instance, when the clusters are distinguished by the order of movement modes sequence, which cannot be captured by the HDP for Gaussian mixtures used here.

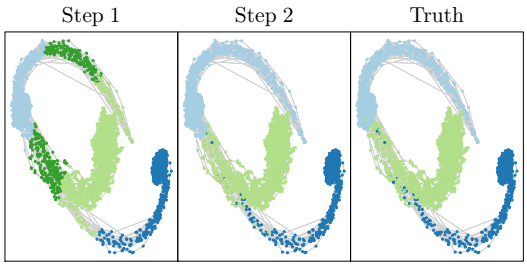

Figure 6: Estimated movement modes (in the bivariate velocity space) after one step (left) and two steps (middle). The ground truth is shown on the right.

4.2 Experiments on marine traffic data

Our clustering approach was performed on 6 months of AIS data of vessels steaming in the Ouessant rail (Brittany, west of France). The data set consists in 18 603 trajectories, gathering overall more than 7 millions GPS observations. The whole trajectory data set is shown on Figure 7.

Model fit was performed considering a maximum of 90 movement modes and 30 trajectory clusters, using non-informative priors. 200 runs of 50 000 iterations for SVI were performed independently, each run taking approximately 8 hours. The run leading to the best ELBO was kept as the estimate, the results are shown hereafter.

Two step movement modes estimation. The SVI leads to the estimation of 81 different movement modes containing (in the sense of maximum a posteriori probability) at least one observation, with 50 movement modes containing 95% of observations. The shapes (mean and covariance) of these movement modes are shown on Figure 8. One can note that a lot of movement modes lead to a same steaming direction, but at different speeds. In this context, putting more informative prior or adding a penalty on movement modes covariance parameters in the ELBO optimization could lead to a model having less movement modes.

Trajectory clusters estimation. In addition to the movement modes estimated above, trajectory cluster were estimated. 29 clusters were estimated as non-empty, three of them are represented on Figure 9. The first two of them are the two more important clusters (gathering respectively 18% and 12% of trajectories), and are composed of two preponderant movement modes. Those two clusters represent the two main marine roads of the Ouessant rail, the first one entering the English Channel, the second one exiting it. The third shown cluster gathers 5% of trajectories,

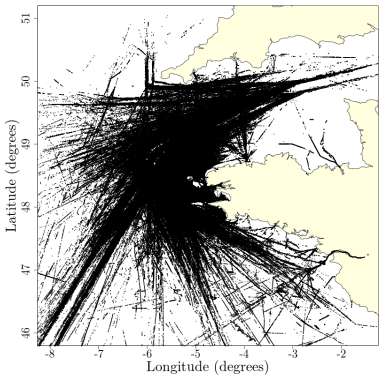

Figure 7: AIS data in the Ouessant rail (Brittany, France), gathered during 6 months (7M. GPS obs.).

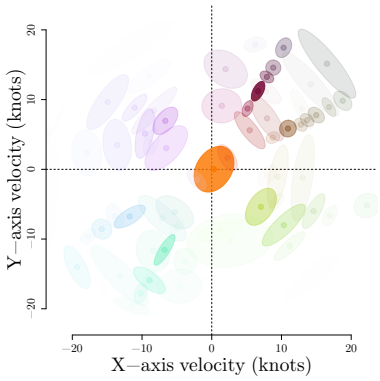

Figure 8: Movement modes estimated distributions (on the velocity space). Each ellipsoid represents a movement mode distribution, the dot being the expected mean, and the contour the 50% centered quantile. The transparency shows the estimated relative weight. The dotted lines show the 0 values.

and mostly gathers trips that are performed at low speed (below 5 knots), gathering both sailing boats and fishing vessels.

Interest of the post-filtering step. In the present case, the second step to re-estimate inconsistent sequences enables avoiding the estimation of transitory movement modes. Figure 10 shows a trajectory belonging to the central cluster of Figure 9. The transitory phase of the vessel, corresponding to its change of direction, is then no longer estimated as a dedicated movement mode, but rather belongs to one of the main movement modes.

5 Conclusions

In this work, we defined a generic framework for the clustering of large trajectory data sets. The proposed framework is specified in continuous time and space,

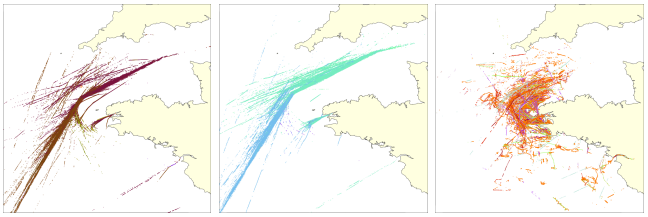

Figure 9: Three estimated trajectory clusters, see the text for details. Colors correspond to those of Fig. 8.

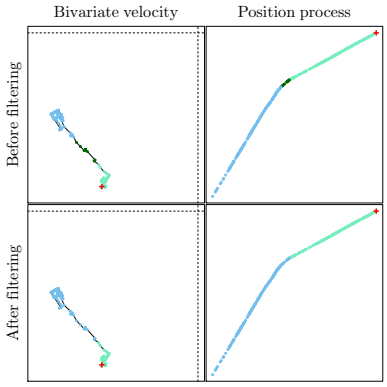

Figure 10: Example of result of the proposed second step in movement mode estimation. Colors corresponds to different movement modes. The green phase in the middle of the trajectory is inconsistent (as defined in the text). It is a transitory phase in the vessel's velocity process. After the second step, it is assigned as the beginning of the second movement mode.

and has therefore a formulation insensitive to GPS sampling, unlike a discrete time and/or space formulation. The model inference is done in a two step scalable approach using stochastic variational inference for a conjugate hierarchical Dirichlet process. In addition to the conjugacy property and the stochastic approach, scalability is guaranteed by an easy to distribute framework, and a linear complexity in the number n of observations. It should be pointed out that our approach relies on a variational approximation of the stick breaking distribution, that imposes to set two parameters K_{max} and C_{max}, as the maximal number of movement modes and trajectory clusters present in the data. One should notice that this number should increase with n. Therefore, the scalability is subject to a reasonable increase of these two parameters with respect to n.[2] Our approach is illustrated in the AIS data context, which is the most obvious example of large trajectory data set. However, the proposed

[2] As a remark, it is worth noting that for our prior specification, a Dirichlet process of parameter α, the expected number of classes in a n data set is $\alpha \log n$.

framework is generic and is, in our opinion, suited to a wide range of trajectory data sets.

References

[1] C. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.

[2] P. Blackwell, M. Niu, M. S. Lambert, and S. D. LaPoint. Exact bayesian inference for animal movement in continuous time. *Methods in Ecology and Evolution*, 7(2):184–195, 2016.

[3] D. Blei, A. Kucukelbir, and J. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112:859–877, 2017.

[4] F. Clazzer, A. Munari, M. Berioli, and F. L. Blasco. On the characterization of ais traffic at the satellite. In *OCEANS 2014-TAIPEI*, pages 1–9. IEEE, 2014.

[5] U. Demšar, K. Buchin, F. Cagnacci, K. Safi, B. Speckmann, N. Van de Weghe, D. Weiskopf, and R. Weibel. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, 3(1):5, 2015.

[6] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[7] B. Garnier and A. Napoli. Exploiting the potential of the future "maritime big data". In *Maritime Knowledge Discovery and Anomaly Detection Workshop*, 2016.

[8] S. J. Gershman and D. M. Blei. A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12, 2012.

[9] E. Gurarie, C. H. Fleming, W. F. Fagan, K. L. Laidre, J. Hernández-Pliego, and O. Ovaskainen. Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement ecology*, 5(1):13, 2017.

[10] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[11] H. H. Hoos and T. Stützle. *Stochastic local search: Foundations and applications*. Elsevier, 2004.

[12] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[13] J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604. ACM, 2007.

[14] X. Li, J. Han, J.-G. Lee, and H. Gonzalez. Traffic density-based discovery of hot routes in road networks. In *International Symposium on Spatial and Temporal Databases*, pages 441–459. Springer, 2007.

[15] R. Nathan. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19050–19051, 2008.

[16] R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

[17] B. Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[18] G. Pallotta, M. Vespe, and K. Bryan. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, 15(6):2218–2245, 2013.

[19] T. Patterson, A. Parton, R. Langrock, P. Blackwell, L. Thomas, and R. King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *Advances in Statistical Analysis*, 101(4):399–438, 2017.

[20] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3-4):225–239, 2008.

[21] J. Sethuraman. A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

[22] M. Steyvers and T. Griffiths. Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7):424–440, 2007.

[23] C. Sung, D. Feldman, and D. Rus. Trajectory clustering for motion prediction. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 1547–1552. IEEE, 2012.

[24] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[25] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical review*, 36(5):823, 1930.

[26] M. Vespe and F. Mazzarella. Maritime knowledge discovery and anomaly detection workshop. European commission JRC conference, July 2016.

[27] C. Wang, J. Paisley, and D. Blei. Online variational inference for the hierarchical dirichlet process. In *Proceedings of AISTATS*, pages 752–760, 2011.

[28] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson. Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *International journal of computer vision*, 95(3):287–312, 2011.

[29] T. Yao, Z. Wang, Z. Xie, J. Gao, and D. D. Feng. Learning universal multiview dictionary for human action recognition. *Pattern Recognition*, 64:236–244, 2017.

[30] S.-Z. Yu. Hidden semi-markov models. *Artificial intelligence*, 174(2):215–243, 2010.

[31] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang. A review of moving object trajectory clustering algorithms. *Artificial Intelligence Review*, 47(1):123–144, 2017.

[32] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM, 2008.

A Characteristics of the Ornstein Uhlenbeck process

Notations:

- I_d is the $d \times d$ identity matrix;
- For a matrix A, A^T is the transposed matrix of A;
- For two 2×2 matrices A and B, $A \oplus B$ denotes the Kronecker sum of A and B, defined by : $A \oplus B = A \otimes I_2 + I_2 \otimes B$, where $\otimes$ denotes the Kronecker product. Note that in our case, $A \oplus B$ is a 4×4 matrix;
- For a matrix $A = \begin{pmatrix} a_1 & a_3 \\ a_2 & a_4 \end{pmatrix}$, $\mathbf{vec}(A)$ denotes the vectorization of A, *i.e.* the vector $\mathbf{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$.

Transition density Suppose that the process $(V_t)_{\tau_1 \leq t \leq \tau_2}$ is solution to the equation (1), then,

$$\forall \tau_1 \leq t \leq \tau_2, \quad V_t | V_{\tau_1} = v_{\tau_1} \sim \mathcal{N}(m_{k,\Delta}^{v_{\tau_1}}, \Lambda_{k,\Delta})$$

where $\Delta = t - \tau_1$ and

$$\begin{aligned} m_{k,\Delta}^{v_{\tau_1}} &= \mathrm{e}^{-\Gamma_k \Delta} v_{\tau_1} + (I_2 - \mathrm{e}^{-\Gamma_k \Delta}) \mu_k \\ \mathbf{vec}(\Lambda_{k,\Delta}) &= (\Gamma_k \oplus \Gamma_k)^{-1} \left(I_4 - \mathrm{e}^{-(\Gamma_k \oplus \Gamma_k)\Delta} \right) \mathbf{vec}\left(\Sigma_k \Sigma_k^T \right) \end{aligned}$$

Likelihood Let $\mathbf{v}$ be a sequence of observations $v_{t_1} = v_{\tau_1}, v_{t_2}, \ldots, v_{t_n} = v_{tau_2}$, discrete time observations of a OUP starting at V_{τ_1} and such that V_{τ_1} is a random variable with p.d.f. (possibly depending on μ_k, Γ_k and Σ_k, $p_{\tau_1}(\cdot)$. Then the likelihood of the observed sequence is given by:

$$L(\mu_k, \Gamma_k, \Sigma_k; \mathbf{v}) = p_{\tau_1}(v_{\tau_1}) \prod_{i=1}^{n-1} \phi_{k,\Delta_i}(v_{t_{i+1}}),$$

where $\Delta_i = t_{i+1} - t_i$ and $\phi_{k,\Delta_i}(v_{t_{i+1}})$ is the p.d.f. of a Gaussian distribution with mean $m_{k,\Delta_i}^{v_{t_i}}$ and covariance Λ_{k,Δ_i}

Stationary distribution From above, one can see that if both $\mathrm{e}^{-\Gamma_k \Delta}$ and $\mathrm{e}^{-(\Gamma_k \oplus \Gamma_k)\Delta}$ vanishes to 0 when Δ increase, then the process $(V_t)_{t \geq \tau_1}$ is asymptotically stationary, and

$$V_t | V_{\tau_1} = v_{\tau_1} \overset{distrib.}{\longrightarrow} \mathcal{N}(\mu_k, \Lambda_k)$$

where

$$\mathbf{vec}(\Lambda_k) = (\Gamma_k \oplus \Gamma_k)^{-1} \mathbf{vec}\left(\Sigma_k \Sigma_k^T \right)$$

The vanishing condition is satisfied when both eigenvalues of Γ_k are positive.

B Computation for variational inference

B.1 Prior specification

Overall, the variables prior distribution are defined as follows:

$$\begin{cases} (\mu_k, \Lambda_k^{-1}) & \overset{i.i.d.}{\sim} & GW(m_0, \rho_0, \gamma_0, W_0) & k \geq 1, \\ \boldsymbol{w} & \sim & GEM(\alpha) & \\ \boldsymbol{\pi}^c & \overset{i.i.d.}{\sim} & GEM(\beta) & c \geq 1 \\ F_j | \boldsymbol{w} & \overset{i.i.d.}{\sim} & Mult(\boldsymbol{w}) & 1 \leq j \leq J \\ Z_i^j | \boldsymbol{\pi}, F_j & \overset{i.i.d.}{\sim} & Mult(\boldsymbol{\pi}^{F_j}) & 1 \leq j \leq J,\ 1 \leq i \leq n_j \end{cases}$$

where $GW(\cdot)$ denotes the Gaussian Wishart distribution, depending on 4 hyperparameters, $GEM(\cdot)$ denotes the stick breaking distribution depending on 1 hyperparameter, and $Mult(\cdot)$ denotes the multinomial distribution (with weights as parameters). The choice of these distributions is convenient as it leads to a conjugate framework for conditional distributions, *i.e.*, the conditional distribution of a hidden variable given the observations and the other hidden variables.

B.2 About the stick breaking construction

To build a prior for an infinite sequence of weights summing to one, we use the stick breaking construction. The topic weights w_c are build using an hyperparameter α:

$$\nu_c \overset{i.i.d}{\sim} \mathcal{B}eta(1,\alpha) \qquad c = 1,2,\dots$$
$$w_c = \nu_c \prod_{i=1}^{c-1}(1-\nu_i) \qquad \textbf{Init: } w_1 = \nu_1$$

Weights π_k^c are sampled similarlrly, with an hyperparameter β:

$$\eta_k^c \overset{i.i.d}{\sim} \mathcal{B}eta(1,\beta) \qquad k = 1,2,\dots$$
$$\pi_k^c = \eta_k^c \prod_{i=1}^{k-1}(1-\eta_i^c) \qquad \textbf{Init: } \pi_1^c = \eta_1^c$$

In the following, we'd rather deal with the sequence of $\boldsymbol{\nu}$ and $\boldsymbol{\eta^c}$, as there are i.i.d. samples. We have, for each c and each k:

$$p(\nu_c) = (1-\nu_c)^{\alpha-1}$$
$$p(\eta_k^c) = (1-\eta_k^c)^{\beta-1}$$

B.3 Likelihood

Keeping the notation of the main text, the likelihood of an observation $\mathbf{v}_i^j$ is given by:

$$p(\mathbf{v_i^j}|Z_i^j,\mu,\Lambda) = \prod_{k=1}^{\infty}\varphi(\mathbf{v_i^j}|\mu_k,\Lambda_k^{-1})^{z_{i,k}^j},$$

which results, for the complete data set $\mathbf{V}$

$$p(\mathbf{V}|Z,\mu,\Lambda) = \prod_{j=1}^{M}\prod_{i=1}^{n_j}\prod_{k=1}^{\infty}\varphi(\mathbf{v_i^j}|\mu_k,\Lambda_k^{-1})^{z_{i,k}^j}$$

The distribution of latent allocation vectors (or local variables) F_j and Z_j^i is given by an (infinite) multinomial distributions depending on weights $\boldsymbol{w}$ and $\boldsymbol{\pi}$:

$$p(Z_i^j|F_j,\boldsymbol{\pi}) = \prod_{c=1}^{\infty}\left(\prod_{k=1}^{\infty}(\pi_k^c)^{z_{i,k}^j}\right)^{f_{j,c}} \tag{5}$$

$$p(F_j|\boldsymbol{w}) = \prod_{c=1}^{\infty} w_c^{f_{j,c}} \tag{6}$$

B.4 Variational approximations of the posterior distributions

Let $\mathcal{L} = \left\{\{Z_i^j\},\{F_j\},\boldsymbol{\eta},\boldsymbol{\nu},\boldsymbol{\mu},\boldsymbol{\Lambda}\right\}$ be the set of (local and global) hidden variables. For any hidden variable U in $\mathcal{L}$ the corresponding $q^*(U)$ is given (Bishop) by

$$\ln q^*(U) = \mathbb{E}_{\mathcal{L}\setminus U}\left[\ln p(\mathbf{V},\mathbf{Z},\mathbf{F},\boldsymbol{\nu},\boldsymbol{\eta},\boldsymbol{\mu},\boldsymbol{\Lambda})\right] + \text{ constant} \tag{7}$$

The complete joint distribution can be split in simpler terms:

$$p(\mathbf{V},\mathbf{Z},\mathbf{F},\boldsymbol{\nu},\boldsymbol{\eta},\boldsymbol{\mu},\boldsymbol{\Lambda}) = p(\mathbf{V}|\mathbf{Z},\boldsymbol{\mu},\boldsymbol{\Lambda})p(\mathbf{Z}|\mathbf{F},\boldsymbol{\eta})p(\mathbf{F}|\boldsymbol{\nu})p(\boldsymbol{\mu},\boldsymbol{\Lambda})p(\boldsymbol{\eta})p(\boldsymbol{\nu}) \tag{8}$$

B.4.1 Optimal variational distribution for $\boldsymbol{\nu}$

The variational approximation of the trajectory cluster is a truncated stick breaking distribution. It is therefore define by C_{max} random variables, such that the first $C_{max}-1$ are beta distributed, and the last one equals to one almost surely. A simple computation using equations (7) and (8) shows that for the $C_{max}-1$ first terms, the variational distribution of ν_c is given by:

$$q^*(\nu_c) \sim \mathcal{B}eta\left(1+\sum_{j=1}^{J}\mathbb{E}_{q_{\boldsymbol{F}}}[\mathbf{1}_{F_j=c}], \alpha+\sum_{j=1}^{J}\mathbb{E}_{q_{\boldsymbol{F}}}[\mathbf{1}_{F_j>c}]\right),$$

where $\mathbf{1}$ denotes the classical indicator function.

B.4.2 Optimal variational distribution for $\boldsymbol{\eta}$

Similarly to the clustering weights, a truncated stick breaking distribution is used. For the cluster c and the movement mode $1 \leq k \leq K_{max}-1$, we have:

$$q^*(\eta_k^c) \sim \mathcal{B}eta\left(1+\sum_{j=1}^{J}\mathbb{E}_{q_{F_j}}[f_{j,c}]\sum_{i=1}^{n_j}\mathbb{E}_{q_{Z_i^j}}[\mathbf{1}_{Z_i^j=k}], \beta+\sum_{j=1}^{J}\mathbb{E}_{q_{F_j}}[f_{j,c}]\sum_{i=1}^{n_j}\mathbb{E}_{q_{Z_i^j}}[\mathbf{1}_{Z_i^j>k}]\right)$$

where we have written $F_j = (f_{j,1}, \ldots, f_{j,C_{max}})$.

B.4.3 Optimal variational distribution for $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$

This computation is done in [1]. For eack movement mode $1 \leq k \leq K_{max}$, the optimal variational distribution is a $\mathcal{GW}(m_k, \rho_k, \gamma_k, W_k)$ with:

$$\begin{aligned}
\rho_k &= \rho_0 + N_k \\
m_k &= \frac{\rho_0 m_0 + N_k \bar{\mathbf{v}}_{\mathbf{k}}}{\rho_k} \\
\gamma_k &= \gamma_0 + N_k \\
W_k &= \left(W_0^{-1} + N_k \mathbf{S}_{\mathbf{k}} + \frac{\rho_0 N_k}{\rho_0 + N_k}(\bar{\mathbf{v}}_{\mathbf{k}}) - m_0)(\bar{\mathbf{v}}_{\mathbf{k}}) - m_0)\right)
\end{aligned}$$

where

$$\begin{aligned}
N_k &= \sum_{j=1}^{J}\sum_{i=1}^{n_j}\mathbb{E}_{q_{Z_i^j}}[z_{i,k}^j] \\
\bar{\mathbf{v}}_{\mathbf{k}} &= \frac{1}{N_k}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\mathbb{E}_{q_{Z_i^j}}[z_{i,k}^j]\mathbf{v}_{\mathbf{i}}^{\mathbf{j}} \\
\mathbf{S}_{\mathbf{k}} &= \frac{1}{N_k}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\mathbb{E}_{q_{Z_i^j}}[z_{i,k}^j](\mathbf{v}_{\mathbf{i}}^{\mathbf{j}} - \bar{\mathbf{v}}_{\mathbf{k}})(\mathbf{v}_{\mathbf{i}}^{\mathbf{j}} - \bar{\mathbf{v}}_{\mathbf{k}})^T
\end{aligned}$$

B.4.4 Optimal variational distribution for F_j

The computation results in $q^*(F_j)$ being a multinomial distribution on the set $[[1, C_{max}]]$ with the c-th weight proportional to:

$$\mathbb{E}_{q_{\boldsymbol{\nu}}}[\log(w_c)] + \sum_{i=1}^{n}\sum_{k=1}^{K_{max}}\mathbb{E}_{q_{Z_i^j}}[z_{i,k}^j]\mathbb{E}_{q_{\boldsymbol{\eta}}}[\log \pi_k^c]$$

Here, one can see that

$$\mathbb{E}_{q_{\boldsymbol{\nu}}}[\log(w_c)] = \mathbb{E}_{\boldsymbol{\nu_c}}[\log(\nu_c)] + \sum_{i=1}^{c-1}\mathbb{E}_{\boldsymbol{\nu_i}}[\log(1-\nu_i)]$$

This expectation can therefore be computed using the two following properties of the beta distribution. If $X \sim \mathcal{B}eta(\alpha_1, \alpha_2)$, then

1. $(1 - X) \sim \mathcal{B}eta(\alpha_2, \alpha_1)$;
2. $\mathbb{E}[\log X] = \psi(\alpha_1) - \psi(\alpha_1 + \alpha_2)$, where ψ is the digamma function.

Of course, the same remarks hold for $\mathbb{E}_{q_{\boldsymbol{\eta}}}[\log \pi_k^c]$.

B.4.5 Optimal variational distribution for Z_i^j

The computation results in $q^*(Z_i^j)$ being a multinomial distribution on the set $[[1, K_{max}]]$ with the k-th weight proportional to:

$$\sum_{c=1}^{C_{max}} \mathbb{E}_{q_{F_j}}[f_{j,c}]\mathbb{E}_{q_{\boldsymbol{\eta}}}[\log \pi_k^c] + \frac{1}{2}\mathbb{E}_{q_{\Lambda_k}}\left[\log(\det \Lambda_k)\right] - \frac{1}{2}\mathbb{E}_{q_{\mu_k,\Lambda_k}}\left[(\mathbf{v}_i^j - \mu_k)^T \Lambda_k (\mathbf{v}_i^j - \mu_k)\right] - \log 2\pi$$

C Simulation design for experiments

We consider a framework with $K = 12$ movement modes, and $C = 10$ trajectory cluster. Each trajectory cluster c is characterized by:

- Its initial movement mode density (the law of the first movement mode chosen during the cluster) denoted by $\boldsymbol{\pi}_0^c = (\boldsymbol{\pi}_{0,1}^c, \ldots, \boldsymbol{\pi}_{0,K}^c)$;
- Its transition densities between movement modes (the laws for the next movement mode after having been in a cluster k) denoted by denoted by $\boldsymbol{\pi}_0^c = (\boldsymbol{\pi}_{k,1}^c, \ldots, \boldsymbol{\pi}_{k,K}^c)$ for all $1 \leq k \leq K$;

A movement mode k is characterized by:

- Its movement parameters $\mu_k, \Gamma_k, \Sigma_k$, *i.e.* the OUP parameters;
- Its sojourn time distribution, *i.e.* the distribution of time spent in this movement mode, denoted by d_k. In our example, d_k is the p.d.f. is the p.d.f. of gamma distribution $\mathcal{G}(\Delta_k, \frac{1}{2})$ where Δ_k is a movement mode-specific parameter.

For each trajectory $1 \leq j \leq 500$, we independently repeat the following procedure:

- **Final time** Sample[3] a final time T^j;
- **Number of observations** Sample[4] a number of observed points n_j;
- **Observation times** Sample independently $\upsilon_1, \ldots, \upsilon_{n_j}$ points from a uniform $U[0, T^j]$, and sort them to set times as $t_1 := \upsilon_{(1)}, \ldots t_{n_j} = \upsilon_{(n_j)}$;
- **Trajectory cluster** Sample a trajectory cluster c with known probabilities $\boldsymbol{w} = (w_1, \ldots, w_{10})$;
- **First movement mode** Sample k_0 with probabilities $\boldsymbol{\pi}_0^c$;
- **First duration** Set $\tau_0 = 0$; Sample δ_0 from d_{k_0}, set $\tau_1 = \tau_0 + \delta_0$;
- **First OUP sampling** Set an initial value V_0. Sample a OUP $(V_t)_{\tau_0 \leq t \leq \tau_1}$ starting at v_0 at all observation times between τ_0 and τ_1 and at τ_1;
- Set $i = 1$;

[3] In our example, it was sampled uniformly in the continuous interval $[80, 500]$

[4] In our example, it was sampled uniformly in the discrete interval $[[50, 450]]$

- **While** $\tau_i < T^j$
  - **Movement mode** Sample k_i with probabilities $\boldsymbol{\pi}^c_{k_{i-1}}$;
  - **Duration** Sample δ_i from d_{k_0}, set $\tau_{i+1} = \tau_i + \delta_i$;
  - **OUP sampling** Sample a OUP $(V_t)_{\tau_i \leq t \leq \tau_{i+1}}$ starting at V_{τ_i} at all observation times between τ_i and τ_{i+1}, and at τ_{i+1};
  - Set $i = i + 1$;

The resulting process is a continuous time hidden semi-Markov model [30] whose emission densities are autoccorelated OU processes.

D Supplementary tables and figures

Table S1: Contingency table (counting the number of points) between true movement modes (in row) and estimated movement modes (in columns, the label assignment was made *a posteriori*), after the first step.

True label / Est. label	1	2	3	4
1	11066	0	116	186
2	2	5261	33	14
3	46	0	5011	276

Table S2: Contingency table (counting the number of points) between true movement modes (in row) and estimated movement modes (in columns, the label assignment was made *a posteriori*), after the second refinement step.

True label / Est. label	1	2	3
1	11363	5	0
2	2	5295	13
3	46	0	5287