Hidden Markov Models: The Best Models for Forager Movements?

Rocio Joo[1,2]*, Sophie Bertrand[1], Jorge Tam[2], Ronan Fablet[3]

1 Unité Mixte de Recherche (UMR) 212 Exploited Marine Ecosystems, Institut de Recherche pour le Développement (IRD), Sète, France, 2 Instituto del Mar del Perú, Callao, Perú, 3 UMR 6285 Information and Communication Science and Technology Laboratory (LabSTICC), Instut TELECOM/TELECOM Bretagne, Brest, France

Abstract

One major challenge in the emerging field of movement ecology is the inference of behavioural modes from movement patterns. This has been mainly addressed through Hidden Markov models (HMMs). We propose here to evaluate two sets of alternative and state-of-the-art modelling approaches. First, we consider hidden semi-Markov models (HSMMs). They may better represent the behavioural dynamics of foragers since they explicitly model the duration of the behavioural modes. Second, we consider discriminative models which state the inference of behavioural modes as a classification issue, and may take better advantage of multivariate and non linear combinations of movement pattern descriptors. For this work, we use a dataset of >200 trips from human foragers, Peruvian fishermen targeting anchovy. Their movements were recorded through a Vessel Monitoring System (~1 record per hour), while their behavioural modes (fishing, searching and cruising) were reported by on-board observers. We compare the efficiency of hidden Markov, hidden semi-Markov, and three discriminative models (random forests, artificial neural networks and support vector machines) for inferring the fishermen behavioural modes, using a cross-validation procedure. HSMMs show the highest accuracy (80%), significantly outperforming HMMs and discriminative models. Simulations show that data with higher temporal resolution, HSMMs reach nearly 100% of accuracy. Our results demonstrate to what extent the sequential nature of movement is critical for accurately inferring behavioural modes from a trajectory and we strongly recommend the use of HSMMs for such purpose. In addition, this work opens perspectives on the use of hybrid HSMM-discriminative models, where a discriminative setting for the observation process of HSMMs could greatly improve inference performance.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rocio.joo@ird.fr

Introduction

Movement paths result from the interaction between the behaviour of an organism and the spatial structuring patterns of its environment [1–5]. Those paths result from the succession of distinct types of behavioural modes (e.g., travelling from one area to another, searching for cues or preys, pursuing and eating a prey), each one associated with the fulfilment of a particular goal. The knowledge of these modes provides rich information on the processes underlying movement, but they are not directly accessible through the sole observation of the sequence of positions recorded by GPS or other position-logging artefacts. The inference of the behavioural modes from movement paths remains a challenging issue in the emerging field of movement ecology [6].

Hidden Markov models (HMMs) have become increasingly popular to address this issue (for examples in classifying activities such as foraging, searching, encamping, cruising, migrating and bedding, see [7–18]; in navigation strategies, see [19–21]; and in types of movement orientation, see [22]). HMMs rely on probabilistic inference of the behavioural modes, stated as hidden states, from the *in situ* observed series. Those series are typically sequences of positions or associated features such as distances, speeds or turning angles along the movement paths [8]. The key

feature of HMMs is to account for the temporal dynamics of the behavioural modes, mostly based on state transitions between steps (two consecutive positions define a step). Such first-order HMMs comprise computationally efficient inference procedures [23,24]. However, it may be unrealistic to consider that a forager takes a decision about changing its behavioural mode at each step, and regardless of any behaviour dating from more than one step back. In this respect, hidden semi-Markov models (HSMMs), recently investigated in movement ecology [8], may be more appealing. While HMMs characterize behaviour at the step scale, HSMMs characterize behaviour at the segment scale; a segment is composed of consecutive steps associated with a same state. HSMMs do account for transitions between consecutive but distinct states and for durations of state segments corresponding to one behavioural mode.

For most living organisms studied in ecology, groundtruthed datasets – samples of tracks or positions for which behavioural modes are known – are hardly available. Therefore, inference issues are generally stated within a non-supervised framework. Furthermore, rigorous model validation (e.g. by cross-validation as in [25,26]) cannot be performed. Model validation mainly relies on some expert-driven evaluation of the ecological or behavioural plausibility of the behavioural modes inferred. Fishermen had long

been the only foragers whose true behavioural modes were available. Actually, on-board observers can provide direct observations of the vessels' activities during fishing trips, allowing for model validation [11]. In pelagic ecosystems, water masses and fish schools are constantly moving [27] so that precise prey localization is unpredictable regardless of the predator, human or animal [28]. Foraging movements for all those predators aim at the same goal, i.e. dealing with uncertainty on prey localization and maximizing prey encounters. As a result, fishermen deploy similar foraging strategies to those of other animal predators [28], therefore the same statistical methods used for other foragers are applied to them (for instance, see [28,29] for a characterization on their diffusive movement through Lévy walks; and see [10–12] for applications of HMMs for identifying their behaviour). Nowadays, tracking data on non-human foragers can be enriched by concurrent deployment of additional devices that explicitly record activities, at least for a sample of individuals (e.g. time-depth recorders for diving as in [9,30–32], video-cameras as in [33,34] and tri-axial accelerometers as in [35]). Given such partially groundtruthed datasets, we are no more in a fully non-supervised context, but rather in supervised or semi-supervised ones [36,37].

Markovian models may apply in such supervised setting. Nevertheless, alternative models may also be considered, particularly because Markovian models are limited for handling multiple observed variables. Choosing and fitting the most appropriate multivariate distribution may be delicate. A simplifying hypothesis is commonly adopted to solve this issue: observed variables conditioned on states are assumed mutually independent, so the multivariate distribution becomes the product of the univariate conditional distributions of each variable. By contrast, discriminative models, such as random forests (RFs), artificial neural networks (ANNs) and support vector machines (SVMs), provide robust solutions for non-linear discrimination in high-dimensional spaces. They have been shown to be highly efficient for a wide range of applications [38–42]. Their availability in several software without the need of strong computational skills makes them attractive for applications to ecological datasets [43,44]. This includes a few studies dedicated to behavioural modes [45–47]. This context of technological advances for data collection enables a wide range of supervised models. Hence, evaluating and comparing models accuracy for inferring behavioural modes becomes necessary.

Here, we consider as study case the foraging movement of 50 Peruvian purse-seiners targeting anchovy. More than 200 of their fishing trips were documented by a Vessel Monitoring System and their behavioural modes simultaneously registered by on-board observers. This unique and large groundtruthed dataset allows performing, via cross-validation, a comprehensive evaluation and comparison of Markovian (HMMs and HSMMs) and discriminative models (random forests, artificial neural networks and support vector machines) for inferring the behavioural modes of a moving forager. We show that HSMMs provide the most accurate inference of the behavioural modes with 80% of global accuracy. We also show via simulation that this result could be greatly reinforced with position records of higher frequency.

Materials and Methods

The purse-seine Peruvian anchovy fishery is the world largest mono specific fishery [48]. Satellite tracking by Vessel Monitoring System (VMS) is mandatory for the whole Peruvian industrial fishing fleet (>1000 vessels) since 2000. Vessel positions (±100m of accuracy; ~1 record per hour) for tens of thousands of fishing trips are thus available for scientific purposes since (e.g., [28–30,46,47,49]). Although most records are given according to one-

hour intervals, some irregularities (e.g. 0.17, 0.99, 12) seldom occur. Since there is no straightforward optimal interpolation method for these cases [8], we work with the records as they are. Therefore, the considered VMS data consist in tracks (i.e., series of positions) with non-regular steps. For each VMS track, several observed variables are computed at each step: speed (sp), heading (θ), changes of speed and turning angles between the previous and the current step (Δsp_{-1} and $\Delta\theta_{-1}$) and between the current and the next step (Δsp_{+1} and $\Delta\theta_{+1}$).

In addition, IMARPE (Peruvian Marine Research Institute) runs a program of observers on-board for a ~1% sample of the fishing trips. They record the location and time of the different behavioural modes occurring during the trips: fishing, searching, cruising (i.e. travelling following a predetermined course), drifting, helping other vessels, and receiving or giving fish to other vessel. For the remaining 99% of the fishing trips, behavioural modes are unknown.

Based on the criteria described in [28,29,47], a groundtruthed dataset gathering tracking data and their corresponding behavioural modes is built. Overall we consider three behavioural modes, fishing, searching and cruising. Fishing trips involving 'helping', 'receiving/giving' and 'drifting' modes are discarded, due to the low number of occurrences of these modes. Together they represent less than 6% of the groundtruthed dataset. We work with a dataset corresponding to 2008, consisting of 242 fishing trips (~36000 fishing trips were performed in total that year). Fig. 1 shows an example of a trip with each VMS record associated with a behavioural mode.

For hidden state inference, two different approaches are investigated and evaluated. Markovian models, which take into account the sequential nature of data; and discriminative models, remarkably popular in the pattern recognition and machine learning domain [50–52]. Henceforth, we will denote by S_t the state variable at time t taking a discrete value s_t, which encodes a behavioural mode (fishing, searching or cruising). A state sequence starting at time 0 and ending at T is then denoted by $S_0^T \equiv S_0, S_1, \ldots, S_T$, taking discrete values s_0^T. Likewise, X_0^T represents the sequence of continuous observed variables taking values x_0^T. Under the two approaches, the goal is to infer s_0^T.

We perform a quantitative evaluation of the models performance using a classic cross-validation procedure. It proceeds as follows. The groundtruthed dataset is split into two sub-samples. The first partition is used for training the models, i.e., learning from the data and estimating the parameters. The second partition is used for validating the models, i.e., evaluating model performance. Training and validation partitions gather each 50% of the original sample of trips and are built by repeated random sub-sampling (20 repetitions). This parameter setting provides us with a trade-off between the performance evaluation and computational efficiency.

Markovian Models

HMM. HMMs are the classic models for inferring hidden state sequences from observed variables [53]. A HMM combines the two following processes. An underlying first-order Markov process of the hidden state sequence, where the probability of currently being at state s_t only depends on the immediately preceding state s_{t-1}. And a state-dependent observation process, where the probability of $X_t = x_t$ only depends on the current state s_t and not on previous states or observations. Assuming homogeneity, a HMM can be fully characterized by (1) the initial probabilities $\pi_i = P(S_0 = i)$, (2) the transition probabilities $p_{ij} = P(S_t = j | S_{t-1} = i)$, and (3) the state-dependent observation probability density functions (pdfs) $b_j(x_t) = f(x_t | S_t = j)$, where
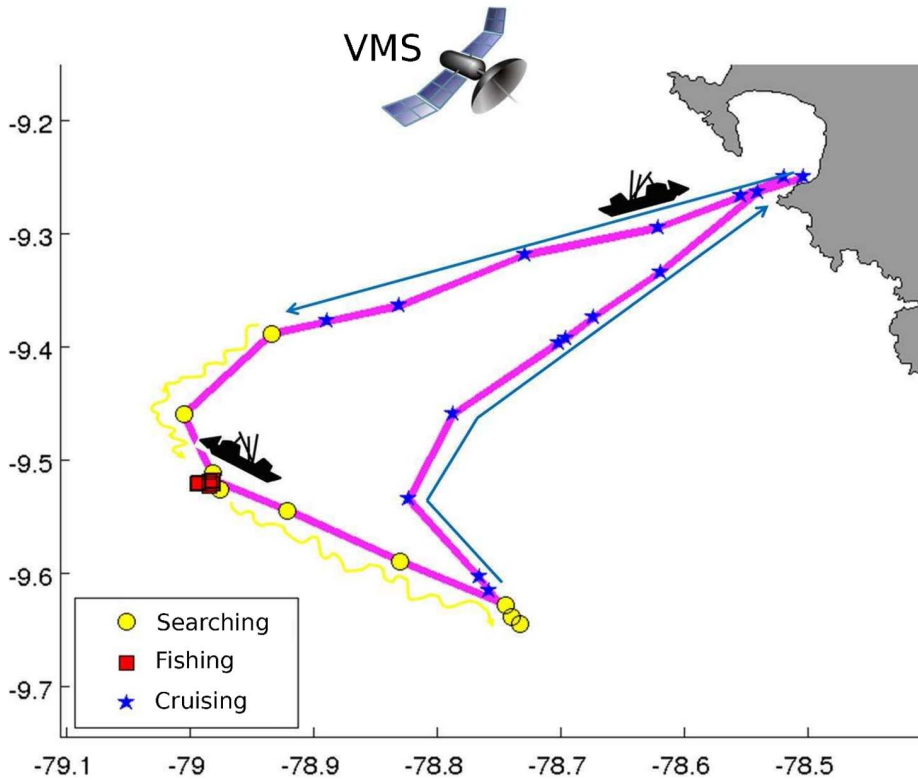
Figure 1. Fishing trip with VMS records and their corresponding behavioural modes.
doi:10.1371/journal.pone.0071246.g001

$f(x_t|S_t=j)$ denotes the conditional pdf of X_t at x_t given $S_t=j$. When observations are multivariate, under mutual independence

$$b_j(x_t)=\prod_{e=1}^{E}b_{je}(x_{te})$$

where E is the number of observed variables included in the model. The likelihood of a HMM can be written as

$$P(S_0^T=s_0^T,X_0^T=x_0^T)=\pi_{s_0}b_{s_0}(x_0)\left[\prod_{t=1}^{T}p_{s_{t-1}s_t}b_{s_t}(x_t)\right]$$

In our case study, several observed variables are available $(sp,\theta,\Delta sp_{-1},\Delta\theta_{-1},\Delta sp_{+1}$ and $\Delta\theta_{+1})$. Over all possible combinations of observed variables, the subset (combination) of variables giving the highest state-inference accuracy is chosen –the computation of accuracy as well as other performance indicators are described in section 'Indicators of model performance'. For each observed variable, we test several probability distributions based on a supervised maximum likelihood (ML) fit. When ML estimation cannot be derived analytically, a numerical optimization is used. Goodness-of-fit (GOF) is tested using the robust Cramér-von Mises statistic [54]. In cases where two or more distributions provided significant fits, the AIC criterion [55] is used for selection among them. All fishing trips start in cruising mode, so initial probabilities are set to one for cruising and zero for the other states. Given the training partition, the ML estimation of the transition probabilities resorts to computing the relative frequencies of the transitions between successive states [36]. Using all these elements, the inference of the sequence of hidden states s_0^T is

done by global decoding via the Viterbi algorithm [23]. Hidden Markov Model toolbox for Matlab [56] is used.

HSMM. A first-order Markov state process may not be, however, the most natural choice for the interpretation of movement patterns. It implicitly assumes that time spent at a given state is distributed according to a geometric distribution. This distribution is memoryless; it means that at a given time t, the waiting time for switching from one state to a distinct state is independent from the time already spent in the former state. However, in practice, a forager's behaviour is not memoryless. A semi-Markov process may therefore be more suitable. It explicitly models the state duration distribution and may consider any distribution function. HSMMs are thus generalizations of HMMs. They combine two processes: a state-dependent observation process as in HMMs, and an underlying semi-Markov state process. A semi-Markov process is determined by the duration distributions $d_j(u)=P(S_{t+u+1}\neq j,S_{t+2}^{t+u}=j|S_{t+1}=j,S_t\neq j)$ and transition probabilities between distinct states $p_{ij}=P(S_t=j|S_t\neq i,S_{t-1}=i)$. For the last visited state, a survival function of the duration is used: $D_j(u)=\sum_{v\leq u}d_j(v)$. The likelihood of a HSMM [57] can be written as

$$P(S_0^T=s_0^T,X_0^T=x_0^T)=\pi_{s_0}d_{s_0}(u_0)\left[\prod_{r=1}^{R-1}p_{s_{r-1}s_r}d_{s_r}(u_r)\right]p_{s_{R-1}s_R}$$

$$D(u_R)I\left(\sum_{r=0}^{R}u_r=T\right)\left[\prod_{t=0}^{T}b_{s_t}(x_t)\right]$$

where $R+1$ is the number of visited states, u_r is the duration at state s_r, and $I()$ denotes the indicator function.

Therefore, compared to HMMs, HSMMs provide a model of the state process at a higher scale: the segment scale (Fig. 2; [58,59]). This segment scale is potentially more relevant for interpreting and discriminating distinct behavioural modes in foraging movement.

The selection of observed variables, the fit of state-dependent observed variable distributions and the estimation of transition probabilities (between distinct states) follow the same criteria as for HMMs. Although state durations are inherently discrete, continuous distributions provide flexibility under certain irregularities on the frequencies of positioning of satellite records. They enable the incorporation of those data directly into the model. Extensive literature on the use of continuous distributions for modelling duration is available (e.g., [58–61]). Here we examine seven continuous probability distributions for modelling the duration of each of the three behavioural modes. Their parameters are estimated by maximum likelihood using the training dataset. Then, GOF is tested using Cramér-von Mises statistic and AIC criterion is used for selection among distributions not rejected by the test. Using all these elements, the inference of the sequence of hidden states s_0^T is done by global decoding via the forward-backward Viterbi algorithm [62]. A code in Matlab for this Viterbi algorithm is given in Text File S3.

Discriminative Models

Discriminative models are alternative approaches for inferring behavioural states within recorded trajectories. In contrast to Markovian approaches, discriminative models do not rely on the explicit modelling of the joint likelihood of observation and state sequences. The inference of the behavioural mode sequence s_0^T is stated as a classification issue, i.e. the determination of the class (behavioural mode) attached to any position along the trajectory. Within a supervised framework, discriminative models learn a classification rule to predict a class from an observed vector x_t. Random forests [63], support vector machines [64] and artificial neural networks [65] are among the state-of-the-art techniques in the machine learning domain [41]. These models differ in the way classification rules are stated and learned. For SVMs, the goal is to maximize the margin around the hyperplane that separates classes. For ANNs, the objective is to minimize the classification error. And for RFs, discrimination is achieved by the simultaneous minimization of the within-group variances and maximization of the between-group variances. The relative performances between these methods are application-dependent and vary according to the structure of the observation space [66]. A key feature of discriminative models is that they do not require any assumption
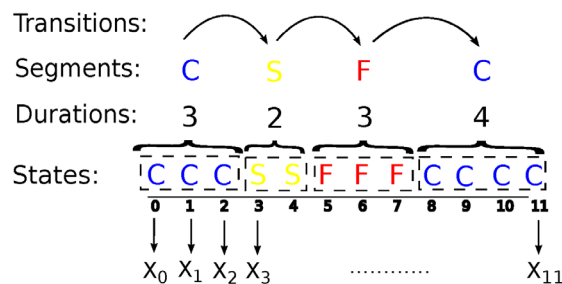
Figure 2. Schematic representation of a HSMM. At each step, an observed feature X is related to a state, which encodes a behavioural mode (C: cruising, F: fishing, S: searching). The state process is modelled at the segment scale and it is characterized by durations and transitions as shown above.
doi:10.1371/journal.pone.0071246.g002

on the nature of the observed variables, their distributions or covariances. To prevent over-fitting during the learning stage, a cross-validation procedure can be applied. Still, it requires sufficiently large and representative groundtruthed datasets.

As for HMMs and HSMMs, the subset of observed variables giving the highest inference accuracy is selected. The selected subsets may differ among the three discriminative models. Architecture and parametrization of each discriminative model is described below.

RFs. A random forest involves a set of N decision trees. A decision tree discriminates patterns recursively in a tree-like structure. At each tree node, m variables are randomly selected among the subset of observed variables. Data are split following certain conditions on those m variables, so that within-group variance is minimized and between-group variance is maximized. For each observed vector x_t, a tree's output is its classification in a behavioural mode. Consequently, a random forest's output is the statistical mode of the classification outputs of N trees. We test $N = \{50,100,500,1000\}$ and $m = \{1,2,\ldots,Y\}$, where Y is the size of the subset of observed variables. The Matlab implementation of the random forest library [68] is used.

SVMs. Support vector machines are based on linear discrimination. A Gaussian kernel is used here for mapping the originally observed vectors into a new space in which classes (i.e., behavioural modes) may be linearly separated. Tested values for the scale parameter of the Gaussian kernel are $\{10^{-4}, 10^{-3}, 0.1, 0.5\}$. SVMs also involve a regularization parameter C. Increasing the value of C increases the cost of misclassifying points and decreases generalization power of the model. We test $C = \{0.1, 1, 10, 100\}$. The Matlab^{TM} implementation of the Libsvm library [68] is used.

ANNs. Multilayer perceptrons (MLPs) are the most widely used architectures of ANNs. Neurons are organized in layers. The first layer is composed of the observed variables and the last layer is composed of the model classification output. Between those first and last layers, one or more hidden layers can exist. Here, we use a MLP with one hidden layer as in [47]. Considered options for the number of hidden neurons range from one to ten. The Matlab neural network toolbox is used for the analysis.

For each discriminative model, we determine the optimal parameter setting according to the classification accuracy.

Indicators of Model Performance

Overall, we aim at accurately reconstructing the sequence of states associated with each foraging trip. We consider two scales of analysis. First, we evaluate the accuracy of the inference at the step scale, and define the accuracy indicator as the percentage of individual steps where the inferred states correspond to the real ones. Second, we assess model performance at the segment scale (Fig. 2), which best characterizes behavioural modes. We use three indicators for each behavioural mode:

- The segment-level precision, defined as the percentage of inferred segments where the inferred behavioural mode corresponds to the true one.
- The segment-level recall, defined as the percentage of real segments where the true mode is correctly inferred.
- The F-measure or F1, which combines precision and recall performances [69]. It is defined as the harmonic mean of precision and recall, and reported here in terms of percentage similarly to precision and recall indicators.

Accuracy, precision, recall and F1 are standard performance evaluation measures in supervised contexts [70]. Beyond these

Table 1. Indicators of model performance.

Scale	Indicator
Step	$\text{Accuracy} = \dfrac{\text{inferred} \cap \text{true}}{\#\text{states}} \times 100\%$
Segment	$\text{Recall} = \dfrac{\text{inferred} \cap \text{true}}{\#\text{true}} \times 100\%$
	$\text{Precision} = \dfrac{\text{inferred} \cap \text{true}}{\#\text{inferred}} \times 100\%$
	$\text{F1} = \dfrac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$
	$\text{Duration} = \dfrac{\sum_{i=1}^{n}(F_i - G_i)^2}{n}$

Notes: F and G represent empirical cumulative distributions for the real and inferred durations of a given behavioural mode, respectively.
doi:10.1371/journal.pone.0071246.t001

performance measures, we also investigate the extent to which the considered models deliver a relevant global characterization of foraging patterns, particularly regarding the shape of the distributions of the behavioural mode durations. In this respect, we define a fourth indicator at the segment scale, called duration. This auxiliary indicator is computed as the mean squared difference between the empirical cumulative distribution functions of both real and inferred mode durations. Its values range from 0 to 1, where 0 refers to an error-free inference.

Formulas for the computation of all these indicators are shown in Table 1. Further details as well as an illustrative example on the computation of accuracy, precision, recall and F1 are described in Text File S1.

Results

The selected distributions for the state-dependent observation process (for HMMs and HSMMs) and for the duration of the states (for HSMMs) are shown in Table 2. For further details, AIC values of distributions with significant fits for observed variables and durations corresponding to the HSMM (Table 3) are indicated in Table S1.

For evaluating and comparing the two Markovian and the three discriminative models, we selected, for each model, the subset of observed variables which led to the greatest inference performance in terms of accuracy rate. Performance indicators at step and segment scales are reported for each of these models (Table 3). All models infer states with an accuracy greater than 75%. By a small

though significant difference ($p < 10^{-5}$ in paired-sample randomness tests; [71]), the HSMM's accuracy is the highest.

Regarding behavioural modes, cruising seems to be the easiest mode to identify. All models show greater F1 scores for the cruising mode (between 74% and 89%). Likewise, the greatest recall and precision values correspond to cruising for all models. Relevant F1 scores are also reached for fishing mode inference (between 73% and 77%). By contrast, the identification of the searching mode appears difficult for all models (F1 between 54% and 67%). This behavioural mode involves relatively large confusion rates with both fishing and cruising modes (between 15% and 19% of the searching states are classified as fishing, and between 25% and 34% are classified as cruising, among all models).

For each behavioural mode, the HSMM outperforms all the other models (greatest F1 scores of 77%, 67% and 89% for fishing, searching and cruising, respectively). The second best model is the HMM. Differences between F1 scores of the HSMM and the HMM are significant for all behavioural modes ($p < 10^{-5}$ in all cases). Among the discriminative models, the ANN is the best model, followed closely by the SVM.

The analysis of the distribution of the inferred durations for each behavioural mode leads to similar conclusions. In Fig. 3 it can be observed that all three discriminative models show higher empirical densities for low duration values than the Markovian models and the groundtruth. Discriminative models, i.e. RF, SVM and ANN, which do not consider state transitions nor durations, tend to under-estimate the duration of modes due to over-segmentation. By contrast, the Markovian models, particularly HSMM, provide more accurate estimates of these durations. Whereas the distribution of the durations for fishing and cruising modes are clearly better represented with the HSMM (duration statistics of 4×10^{-4} and 1×10^{-4} for fishing and cruising, respectively; Table 3), the HMM gives slightly better results for the searching mode (duration of 1×10^{-4}).

The over-segmentation problem is illustrated for one trajectory sample when comparing the sequences of behavioural modes inferred by the HSMM and the RF with the true sequence of modes (Fig. 4). There is strong over-segmentation in the sequences inferred by the RF, leading to under-estimation of the duration of the segments. By contrast, the HSMM achieves relevant representation of the mode sequences through time (Fig. 4, low panel) and thus also through space (Fig. 4, right panel).

Regarding computational cost, we compare all five models in Table 3 for one replica where 121 tracks were randomly selected for training and the remaining 121 for validation. The HMM shows the lowest computational time (16.78 seconds), followed by

Table 2. Distributions for each observed variable and duration conditioned on states.

| Observed Variable	Searching	Fishing	Cruising
sp	generalized Pareto	generalized extreme value	Gaussian mixture
θ	uniform	wrapped Cauchy	Laplace-Gaussian mixture
$\Delta\theta_{-1}$	Kumaraswamy	uniform	loglogistic
$\Delta\theta_{+1}$	Beta	uniform	loglogistic
Δsp_{-1}	Laplace	Gaussian mixture	Student's t
Δsp_{+1}	Laplace	Gumbel	Student's t
Duration	generalized extreme value	lognormal	generalized extreme value

Notes: When Beta and Kumaraswamy distributions are used, data is transformed to scale from 0 to 1.
doi:10.1371/journal.pone.0071246.t002

Table 3. Performance of all models for their corresponding best subsets of observed variables.

Mode	Model	HSMM	HMM	SVM	RF	ANN
	Subset	$sp, \Delta sp_{+1}$	$sp, \Delta sp_{+1}$	$sp, \Delta sp_{-1}, \Delta sp_{+1}$	$sp, \Delta sp_{-1}, \Delta sp_{+1}, \theta, \Delta\theta_{+1}$	$sp, \Delta sp_{-1}, \Delta sp_{+1}, \Delta\theta_{+1}$
	Accuracy	**80.3%**	79.1%	79.0%	76.4%	79.2%
F	Recall	86.6%	84.7%	88.3%	85.5%	88.5%
	Precision	69.3%	69.4%	65.8%	64.5%	65.7%
	F1	**77.0%**	76.3%	75.4%	73.5%	75.4%
	Duration	4	14	17	29	22
S	Recall	37.5%	64.9%	56.3%	62.0%	59.5%
	Precision	56.9%	56.1%	57.6%	47.5%	54.3%
	F1	**66.7%**	60.2%	56.9%	53.7%	56.9%
	Duration	7	4	33	41	26
C	Recall	91.0%	87.4%	89.6%	76.9%	88.7%
	Precision	87.3%	86.4%	71.9%	72.0%	75.2%
	F1	**89.1%**	86.9%	79.8%	74.3%	82.1%
	Duration	1	2	7	25	7

Notes: In bold, the highest values of accuracy and F1. F: fishing; S: searching; C: cruising. Duration values are scaled by (10^{-4}).
doi:10.1371/journal.pone.0071246.t003
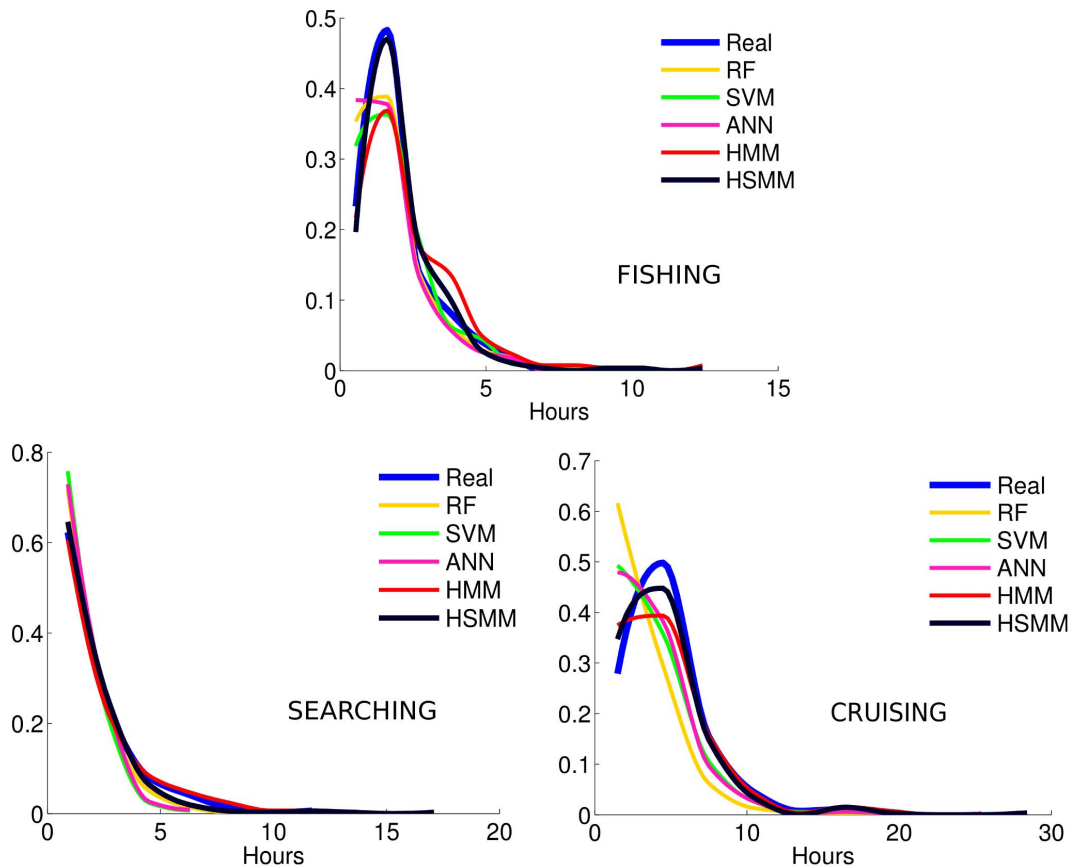
Figure 3. Distribution of the duration of each behavioural mode. For each model, an empirical distribution of the duration of each mode is estimated based on the duration of all inferred segments encoding the mode. RF: random forest. SVM: support vector machine. ANN: artificial neural network. HMM: hidden Markov model. HSMM: hidden semi-Markov model. Real: known behavioural modes.
doi:10.1371/journal.pone.0071246.g003
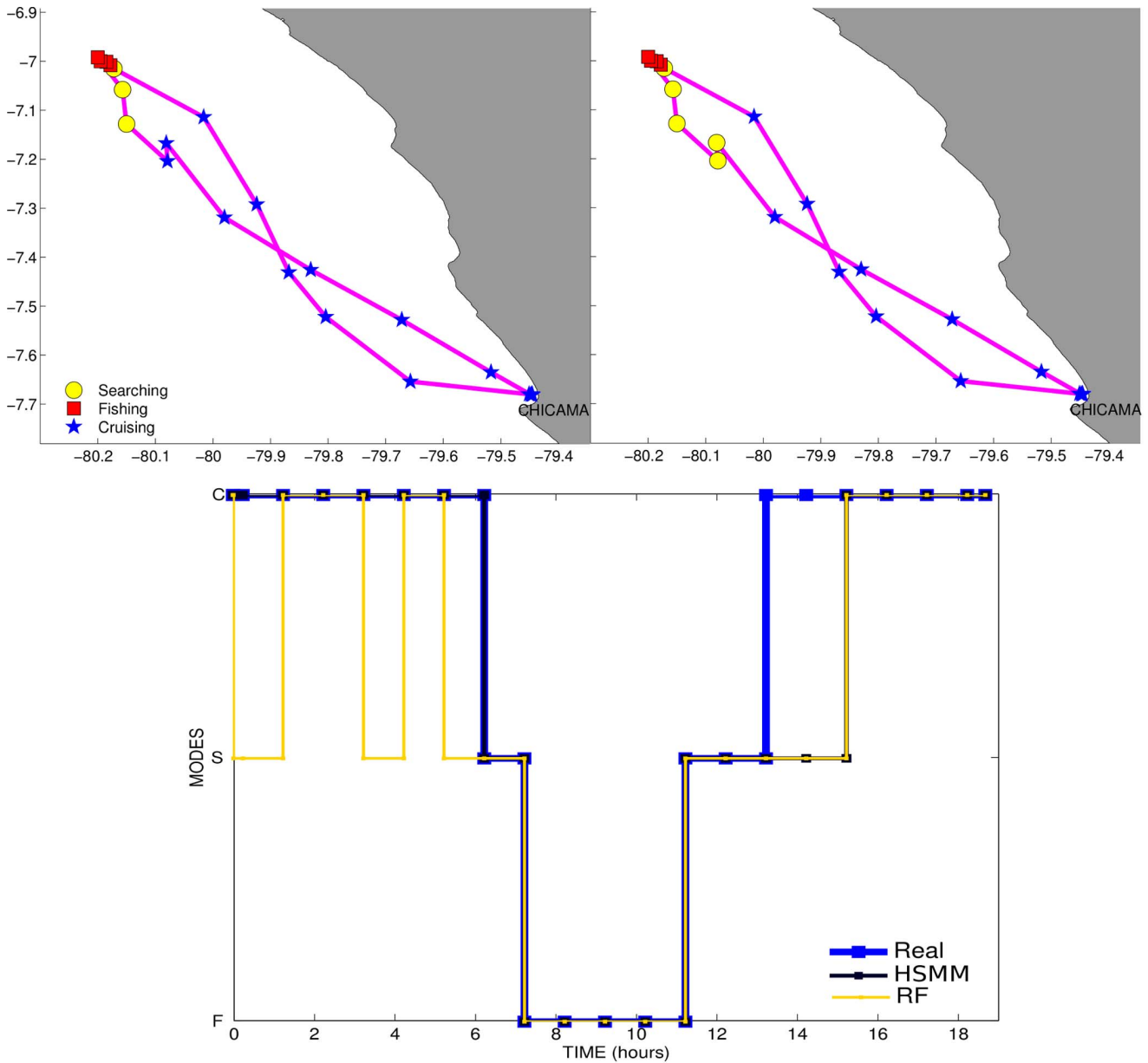
Figure 4. A fishing trajectory. Left upper panel: track with real behavioural modes. Right upper panel: track with inferred modes using the HSMM. Lower panel: temporal representation of the behavioural mode sequences, real and inferred, where 0 in the x-axis represents the beginning of the trip.
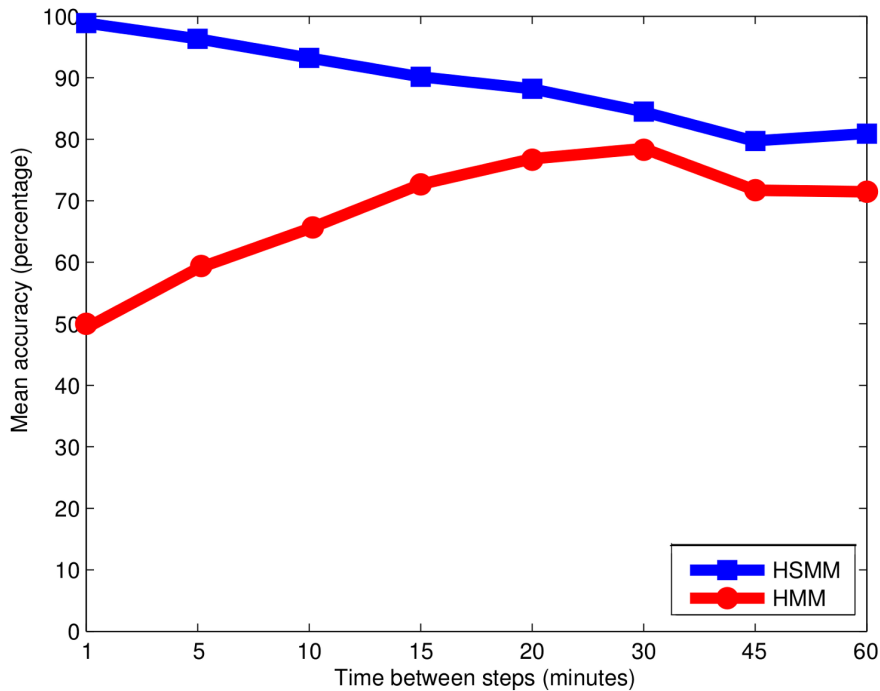
doi:10.1371/journal.pone.0071246.g004

the RF and the SVM models (22.09 and 23.07 seconds, respectively). Next it is the HSMM (64.04 seconds) and finally the most expensive one is the ANN (140.14 seconds). For the HMM and the HSMM, the computational time comprised the estimation of the probability density function parameters and Viterbi algorithm application. For the SVM, the RF and the ANN, it comprised the optimal parameter setting, as described in the Methods section. The high computational cost of the ANN could be greatly affected by the call to a graphical interface as automatically performed by the Neural Network toolbox of Matlab^{TM}. This computational analysis should only be regarded in relative terms. Optimized implementations of these models could be expected to provide important computational gains (by a factor of 10 or more).

Discussion

With a representative groundtruthed dataset composed of 242 fishing trips, we perform a comprehensive cross-validation evaluation of different Markovian and discriminative models for inferring behavioural modes from trajectory data. Our results show that the HSMM is the best model and enlighten several critical issues.

State Dynamics are Key Information

Markovian models have the strength of considering the sequential nature of the data: state transitions are explicitly modelled and the sequence of states is inferred as the most likely sequence given the performed trajectory. However, they present

Figure 5. Mean accuracy for simulated sequences for different sampling rates using HSMM and HMM.
doi:10.1371/journal.pone.0071246.g005

limitations for incorporating the information contained in the observed variables, especially in cases of non-Gaussian multivariate observation spaces. Practical applications of Markovian models often involve simplifications such as independence and/or Gaussianity assumptions for modelling the multivariate distribution of the observed features given the behavioural modes. In contrast, discriminative models state the inference of behavioural modes as a classification issue. They use powerful non-linear and multivariate classification rules. At the step scale, the HSMM surpassed the discriminative models by small differences (+1% of accuracy with respect to the ANN and the SVM, and +4% with respecto to the RF; Table 3). At the segment scale, the surpassing outperformance of the HSMM was clearer (differences in F1 scores between +1.6% and +9.8% regarding both the ANN and the SVM, and between +3.5% and 14.8% regarding the RF; Table 3). This evidences that the information contained in the state sequence is key for accurately inferring the behavioural modes.

HSMMs are Recommended for Behavioural Mode Inference

To our knowledge, our study presents the first application of HSMMs to foraging tracks using groundtruthed data on behavioural modes. For this study case, with steps of ~1 hour, the HSMM performed slightly better than the HMM. A simulation study on high-resolution data (one-second steps) is described in Text File S2. We applied HSMMs and HMMs to sub-sampled versions of these sequences. The performance of each model was assessed by the mean accuracy (MA), which is the average of the accuracy for each behavioural mode (Fig. 5). For one-minute steps, the HMM performed very poorly, whereas for 30-minute steps it was by far more relevant (50% vs 78% of MA). By contrast, MA rates for the HSMM remained above 80% for all time steps. The HSMM actually benefited from high-resolution sequences – when available – to significantly improve inference

performance (100% of MA for one-minute steps). These additional results clearly illustrate that the relevance of the first-order Markov state process embedded in HMMs greatly depends on the time steps of the trajectory data. By contrast, we show that the relevance of the HSMM does not decrease with smaller time steps. Likewise, [72] showed that reducing time steps severely decreased the performance of first-order Markov processes for estimating animal spatial distributions from tracking data.

Alternatively, higher-order (n^{th}-order) hidden Markov models account for additional complexity in the dynamics of the state sequence. They comprise a nth memory, i.e. the state value s_t depends on the state values taken at the n preceding states. They implicitly involve more general distributions on state segment durations than geometrical distributions. Therefore, they should outperform first-order HMMs for high-resolution sequences. However, in most practical problems the choice of the order of the hidden Markov model is not obvious and depends on both the time resolution of the data and the characteristic durations of the state segments. In addition, they are computationally expensive. HSMMs avoid the problem of choosing and fixing an order for the Markovian process. By considering transitions between distinct state segments and distributions on their durations, HSMMs model the scale of a homogeneous behavioural mode. By considering any distribution for modelling duration probability, HSMMs explicitly model the time an individual stays in a behavioural mode, rather than simply accepting the geometric decay of the duration distribution imposed by standard first-order HMMs [8]. Moreover, by considering continuous distributions, HSMMs can directly incorporate tracking data involving some cases with different time steps.

Overall the great flexibility of HSMMs makes them particularly attractive for the analysis of foraging movement patterns, since tracking data on animals are commonly available at high resolutions and are often acquired with irregular sampling rates [73].

Real Behavioural Modes and the Relevance of Model Validation

The technological and methodological advances enable access to larger amounts of data and lead to continuously elaborating and applying new flexible modelling approaches for animal movement [73]. While following this trend, model validation and evaluation are often disregarded. [74] discuss this issue as a challenge in the future of statistics in general. It is also a challenge in movement modelling, particularly due to the conceptual and practical difficulties for obtaining groundtruthed data on animal behaviour.

Hence, when validating models with groundtruthed data, not only the models should be discussed but the data as well. In this work, we had access to a groundtruthed dataset, where behavioural modes were not chosen by us. Instead, they were previously defined by the predators themselves (fishermen) together with the on-board observers. This meant that states were not chosen in a way that they would be *a priori* easily recognizable (based on path geometry). On the other hand, it gave a great opportunity for evaluating the models performance for inferring real and complex behavioural mode sequences.

We reported 80% of global accuracy and 77%, 67% and 89% of F1 for fishing, searching and cruising, respectively, using the fitted HSMM model. Whereas the general performance is satisfactory, the searching mode appears difficult to identify. It might be explained by the nature of this behavioural mode. Interviewed fishermen anticipated that geometrical patterns in their tracks related to searching might vary greatly depending on several factors, especially whether or not they presume the inspected zone to be of high prey density. According to the fishermen, observed patterns for fishing and cruising are more stable. The low F1 score for searching may also be due to the time resolution of the data. As for fishing, the activity lasts ~2 hours on average. However, 30-minute searching modes between two fishing modes were also reported by on-board observers. Such short state segments result in mixed signatures at the one-hour steps of the VMS data and can hardly be analysed. Higher-resolution tracking data should clearly contribute to a better identification of such searching modes, and would decrease the confusion rates with fishing and cruising; thus improve the inference accuracy of all behavioural modes. Moreover, as shown by the simulation study, HSMMs would increase their inference power if data resolution increases.

Beyond Validation: Inference in Supervised and Semi-supervised Contexts

In supervised contexts, inferring behavioural modes is not only useful for achieving model validation. Supervised contexts do not necessarily imply that groundtruthed data on the behavioural modes of the whole population of tracks are available. Known behavioural modes may only be available for a subset of the tracks. In the case of fishermen, for instance, there may not be enough resources for on-board observers to register activities from all fishing trips of the entire population of vessels with tracking devices. For the Peruvian anchovy fishing fleet, more than 30000 fishing trips are tracked by VMS per year, but behavioural modes of only ~300 of those trips are registered by on-board observers. Likewise, financial limitations could make possible tagging more individuals with GPS than with time-depth recorders (e.g. [9,30]). Other limitations such as the memory card capacity for video-camera devices or daily diaries (e.g. [33–35]) may enable access to behavioural modes only at the beginning of the tracks. For all those cases, models trained and validated over the groundtruthed samples could be used for inferring behavioural modes over the remaining tracks or segments of tracks.

On the other hand, the non-supervised observed data could be used for updating the trained and validated models. In the machine learning domain, this is generally referred to as a semi-supervised setting. During the last years, numerous semi-supervised strategies have been proposed (see [37] for an extensive classification and revision). Among them, Markovian models naturally extend from the supervised case to the semi-supervised one, using the EM algorithm [75]. This appears as a particularly promising research direction for ecological studies, including the estimation of the resources (i.e. number of on-board observers, animal-borne electronic devices and analyses) to be allocated for gathering an optimal groundtruthed dataset.

Modelling Extensions for Improving Inference Power

We have shown and discussed the advantages of Markovian models for taking into account the sequential nature of the data, while discriminative models typically achieve an independent inference of each state. Introducing past information on the observed variables may improve the inference performance of the discriminative models. We tested this possibility by introducing the immediate past values of the observed variables as new observed variables for the discriminative models. That meant adding four observed variables: speed at the previous step (sp_{-1}), heading at the previous step (θ_{-1}), change of speed between the two previous steps (Δsp_{-2}) and turning angle between the two previous steps ($\Delta\theta_{-2}$). The immediate past values of Δsp_{+1} and $\Delta\theta_{+1}$ are Δsp_{-1} and $\Delta\theta_{-1}$, respectively. As indicated in the Methods section, for each model, from all the possible combinations of observed variables, we retained the subset of variables giving the greatest accuracy rate. Only for ANNs, a different subset of variables ($sp,sp_{-1},\Delta sp_{-1},\Delta sp_{+1},\theta,\theta_{-1},\Delta\theta_{-1},\Delta\theta_{+1}$) gave a higher accuracy. The new subset of observed variables involves the subset of variables from Table 3 plus four more observed variables. It improves inference of cruising modes (+0.5% in F1) and the general accuracy of the ANN model (+0.3%), although it decreases the performance over fishing and searching modes (−0.8% and −0.9% in F1, respectively).

Of course, more memory (past and future) in the observed variables could be added. But then, we would come across with the same memory-order dilemma than the one discussed for states in HMMs. Moreover, when we consider n^{th} order past (or future) of an observed variable, the first (or last) n records will have missing values. This could be particularly annoying for classification using discriminative models. Another possibility would be to incorporate binary probabilities of the past states (i.e. presence or absence of a behavioural mode in the past states) for incremental training of discriminative models [76]. Incremental training involves training the model one time-step at a time, updating the model at each step. Nonetheless, this may result in over-fitting and large generalization errors. Besides, the direct application of this strategy may lead to drift effect. It means that inference at time t may be biased as it is driven by the effect of the inference at time $t-1$. By contrast, Markovian models rely on a global inference, i.e. retrieving the state sequence that maximizes the posterior likelihood given the observed series. This global inference involves a forward-backward procedure which guarantees that the inference of any given state equally depends on past and future features along the trajectory.

Hence, combining the Markovian setting, which accounts for the sequential nature of the states, and the discriminative setting, which can achieve improved classification performance in high-dimensional non-Gaussian observation spaces, seems highly

appealing. Such hybrid models have been investigated for different applications, especially speech recognition (e.g.[77–80]). They are stated as Markovian models that rely on the definition of an observation likelihood from the output of the chosen discriminative model (e.g. the discrimination SVM function for hybrid SVM-Markov models; [79]). However, the parametrization of the observation likelihood and the training of the hybrid model remain complex issues, which should be investigated in a future work.

Another attractive extension would be to model the observation process at the segment scale, i.e. at the same scale than that of the semi-Markov state process. That way, at each segment, one observation feature would be related to one state segment, which at the same time, would depend on the immediately preceding state segment. This modelling approach presents some potential advantages: it would imply modelling at the behavioural mode scale not only the state process but the observation process as well, and it could significantly improve the robustness to the presence of low-informative observation features.

The incorporation of informative priors could also play an important role in improving behavioural mode inference. For instance, predators may know *a priori* that the probability of foraging success increases/decreases with daylight. Since this knowledge affects their behaviour, hour-dependent state transition priors can be incorporated to the model. Likewise, priors on competition/association, as well as local climate conditions restricting mode transitions and durations could also be introduced in the model.

Synthesis

We have shown a pioneer evaluation and comparison of Markovian and discriminative models for inferring behavioural modes within movement tracks in a supervised framework. The surpassing performance of Markovian models over the discriminative models highlights the importance of modelling state dynamics for accurately inferring the behavioural mode sequences. HMMs have been the most common approach in movement ecology. However, semi-Markov processes represent better the behavioural mode sequences than first-order Markov processes, since they explicitly model state duration and consider transitions at a segment scale. The HSMM performance on the ground-truthed dataset is slightly better than that of the HMM. As discussed above, this result responds to the nature of these particular behavioural modes as well as to the low resolution of the data. The ~1 hour time steps are slightly below the characteristic durations of fishing and searching segments. Hence, regarding time steps, it is a favourable scenario for HMM. Through a simulation experiment, it was shown that increasing time resolution may decrease the accuracy obtained with HMMs and conversely increase the accuracy of HSMM inference. In foraging movement analysis, where (1) each type of behaviour contained in a track is typically characterized by a distinct duration, (2) tracking data are increasingly available at high resolutions, and (3)

irregularity in sampling rates is not uncommon, we highly recommend the use of HSMMs. In addition, this work opens perspectives on the use of hybrid HSMM-discriminative models, where a discriminative setting for the observation process of HSMMs could greatly improve inference performance.

Supporting Information

Figure S1 Example of a sequence with its real and inferred behavioural modes. 1's and 0's in recall/precision represent a positive or null recall/precision corresponding to each behavioural mode, respectively. C = cruising, S = searching and F = fishing.
(TIF)

Table S1 Distributions with significant fits under Cramér von Mises test for each observed variable and duration conditioned on states, for the HSMM in Table 3 of the manuscript. AIC values are in parentheses.
(DOC)

Text File S1 Details on calculations of accuracy, precision, recall and F1 indicators.
(DOC)

Text File S2 Details on the simulation study.
(DOC)

Text File S3 Matlab^{TM} code for HSMM Viterbi algorithm.
(TXT)

Author Contributions

Conceived and designed the experiments: RJ RF SB. Performed the experiments: RJ. Analyzed the data: RJ RF. Contributed reagents/materials/analysis tools: RF JT. Wrote the paper: RJ SB RF JT.

References

1. Johnson AR, Wiens JA, Milne BT, Crist TO (1992) Animal movements and population dynamics in heterogeneous landscapes. Landscape Ecology 7: 63–75.
2. With KA (1994) Using fractal analysis to assess how species perceive landscape structure. Landscape Ecology 9: 25–36.
3. Nams V (1996) The VFractal : a new estimator for fractal dimension of animal movement paths. Landscape Ecology 11: 289–297.
4. Bergman CM, Schaefer JA, Luttich SN (2000) Caribou movement as a correlated random walk. Oecologia 123: 364–374.
5. Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, et al. (2008) A movement ecology paradigm for unifying organismal movement research. PNAS 105: 19052–19059.
6. Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-space models of individual animal movement. Trends in Ecology and Evolution 23: 87–94.
7. Hart T, Mann R, Coulson T, Pettorelli N, Trathan P (2010) Behavioural switching in a central place forager: patterns of diving behaviour in the macaroni penguin (Eudyptes chrysolophus). Marine Biology 157: 1543–1553.
8. Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, et al. (2012) Flexible and practical modeling of animal telemetry data : hidden Markov models and extensions. Ecology 93: 2336–2342.
9. Dean B, Freeman R, Kirk H, Leonard K, Phillips RA, et al. (2012) Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov

model reveals the distribution of at-sea behaviour. Journal of the Royal Society, Interface/the Royal Society 20120570.

10. Peel D, Good NM (2011) A hidden Markov model approach for determining vessel activity from vessel monitoring system data. Canadian Journal of Fisheries and Aquatic Sciences 68: 1252–1264.

11. Walker E, Bez N (2010) A pioneer validation of a state-space model of vessel trajectories (VMS) with observers' data. Ecological Modelling 221: 2008–2017.

12. Vermard Y, Rivot E, Mahévas S, Marchal P, Gascuel D (2010) Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian hidden Markov models. Ecological Modelling 221: 1757–1769.

13. Bestley S, Patterson TA, Hindell MA, Gunn JS (2010) Predicting feeding success in a migratory predator: integrating telemetry, environment, and modeling techniques. Ecology 91: 2373–2384.

14. Pedersen MW, Patterson TA, Thygesen UH, Madsen H (2011) Estimating animal behavior and residency from movement data. Oikos 120: 1281–1290.

15. Patterson TA, Basson M, Bravington MV, Gunn JS (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. The Journal of animal ecology 78: 1113–1123.

16. Jonsen ID, Myers RA, James MC (2007) Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology Progress Series 337: 255–264.

17. Franke A, Caelli T, Hudson RJ (2004) Analysis of movements and behavior of caribou (Rangifer tarandus) using hidden Markov models. Ecological Modelling 173: 259–270.

18. Franke A, Caelli T, Kuzyk G, Hudson RJ (2006) Prediction of wolf (Canis lupus) kill-sites using hidden Markov models. Ecological Modelling 197: 237–246.

19. Roberts S, Guilford T, Rezek I, Biro D (2004) Positional entropy during pigeon homing I: application of Bayesian latent state modelling. Journal of theoretical biology 227: 39–50.

20. Guilford T, Roberts S, Biro D, Rezek I (2004) Positional entropy during pigeon homing II: navigational interpretation of Bayesian latent state models. Journal of theoretical biology 227: 25–38.

21. Lau KK, Roberts S, Biro D, Freeman R, Meade J, et al. (2006) An edge-detection approach to investigating pigeon navigation. Journal of theoretical biology 239: 71–8.

22. Mann RP, Perna A, Strömbom D, Garnett R, Herbert-Read JE, et al. (2013) Multi-scale Inference of Interaction Rules in Animal Groups Using Bayesian Model Selection. PLoS Computational Biology 9: e1002961.

23. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77: 257–286.

24. Gimpel K, Rudoy D (2008) Statistical Inference in Graphical Models. Technical report, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts.

25. Tan CO, Özesmi U, Beklioglu M, Per E, Kurt B (2006) Predictive models in ecology: Comparison of performances and assessment of applicability. Ecological Informatics 1: 195–211.

26. Hijmans RJ (2012) Cross-validation of species distribution models : removing spatial sorting bias and calibration with a null model. Ecology 93: 679–688.

27. Peraltilla S, Bertrand S In situ measurements of the speed of peruvian anchovy schools.

28. Bertrand S, Bertrand A, Guevara-Carrasco R, Gerlotto F (2007) Scale-invariant movements of fishermen: The same foraging strategy as natural predators. Ecological Applications 17: 331–337.

29. Bertrand S, Burgos JM, Gerlotto F, Atiquipa J (2005) Lévy trajectories of Peruvian purse-seiners as an indicator of the spatial distribution of anchovy (Engraulis ringens). ICES Journal of Marine Science 62: 477–482.

30. Bertrand S, Joo R, Smet CA, Tremblay Y, Weimerskirch H (2012) Local depletion by a fishery can affect seabird foraging. Journal of Applied Ecology.

31. Cook TR, Hamann M, Pichegru L, Bonadonna F, Grémillet D, et al. (2012) GPS and time-depth loggers reveal underwater foraging plasticity in a ying diver, the Cape Cormorant. Marine Biology 159: 373–387.

32. Dragon A, Bar-Hen A, Monestiez P, Guinet C (2012) Comparative analysis of methods for inferring successful foraging areas from Argos and GPS tracking data. Marine Ecology Progress Series 452: 253–267.

33. Heaslip SG, Iverson SJ, Bowen WD, James MC (2012) Jellyfish support high energy intake of leatherback sea turtles (Dermochelys coriacea): video evidence from animal-borne cameras. PloS one 7: e33259.

34. Thiebault A, Tremblay Y (in press) Splitting animal trajectories into fine-scale behaviorally consistent movement units: breaking points relate to external stimuli in a foraging seabird.

35. Wilson R, Shepard E, N L (2008) Prying into the intimate details of animal lives: use of a daily diary on animals. Endang Species Res 4: 123–137.

36. Dietterich TG (2002) Machine learning for sequential data: A review. In: Structural, Syntactic, and Statistical Pattern Recognition. Springer-Verlag, 15–30.

37. Chapelle O, Schölkopf B, Zien A (2006) Semi-Supervised Learning. Cambridge, Massachusetts.: The MIT Press.

38. Zhang G (2000) Neural Networks for Classification: A Survey. IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews 30: 451–462.

39. Byun H, Lee SW (2002) Applications of support vector machines for pattern recognition: A survey. In: Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines. London, UK: Springer-Verlag, SVM '02, 213–236. URL http://dl.acm.org/citation.cfm?id=647230.719394.

40. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, et al. (2007) Random forests for classification in ecology. Ecology 88: 2783–2792.

41. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference and prediction. Springer, 2 edition. URL http://www-stat.stanford.edu/tibs/ElemStatLearn/.

42. Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66: 247–259.

43. Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83: 171–193.

44. Crisci C, Ghattas B, Perera G (2012) A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling 240: 113–122.

45. Morales JM, Fortin D, Frair JL, Merrill EH (2005) Adaptive models for large herbivore movements in heterogeneous landscapes. Landscape Ecology 20: 301–316.

46. Bertrand S, Diaz E, Lengaigne M (2008) Patterns in the spatial distribution of Peruvian anchovy (Engraulis ringens) revealed by spatially explicit fishing data. Progress In Oceanography 79: 379–389.

47. Joo R, Bertrand S, Chaigneau A, Ñiquen M (2011) Optimization of an artificial neural network for identifying fishing set positions from VMS data: An example from the Peruvian anchovy purse seine fishery. Ecological Modelling 222: 1048–1059.

48. Chavez F, Bertrand A, Guevara-Carrasco R, Soler P, Csirke J (2008) The northern Humboldt Current System: Brief history, present status and a view towards the future. Progress in Oceanography 79: 95–105.

49. Bertrand S, Dewitte B, Tam J, Díaz E, Bertrand A (2008) Impacts of Kelvin wave forcing in the Peru Humboldt Current system: Scenarios of spatial reorganizations from physics to fishers. Progress in Oceanography 79: 278–289.

50. Nallapati R (2004) Discriminative models for information retrieval. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, SIGIR '04, 64–71. doi:10.1145/1008992.1009006. URL http://doi.acm.org/10.1145/1008992.1009006.

51. He X, Deng L, Chou W (2008) Discriminative Learning in Sequential Pattern Recognition. IEEE Signal Processing Magazine 14: 14–36.

52. Mjolsness E, DeCoste D (2001) Machine learning for science: state of the art and future prospects. Science (New York, NY) 293: 2051–5.

53. Bengio Y (1999) Markovian models for sequential data. Neural Computing Surveys 2: 129–162.

54. Schwarz G (1978) Estimating the dimension of a model. The Annals of Statistics 6: 461–464.

55. Akaike H (1981) Likelihood of a model and information criteria. Journal of Econometrics 16: 3–14.

56. Murphy K (1998). Hidden markov model toolbox for matlab. URL http://www.ai.mit.edu/murphyk/Software/hmm.html/.

57. Guédon Y (2003) Estimating hidden semi-Markov chains from discrete sequences. Journal of Computational and Graphical Statistics 12: 604–639.

58. Dong M, He D (2007) A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. Mechanical Systems and Signal Processing 21: 2248–2266.

59. Yu Sz (2010) Hidden semi-Markov models. Artificial Intelligence 174: 215–243.

60. Levinson S (1986) Continuously variable duration hidden markov models for automatic speech recognition. Computer Speech & Language 1: 29–45.

61. Beyreuther M, Wassermann J (2011) Hidden semi-Markov Model based earthquake classification system using Weighted Finite-State Transducers. Nonlinear Processes in Geophysics 18: 81–89.

62. Guédon Y (2007) Exploring the state sequence space for hidden Markov and semi-Markov chains. Computational Statistics & Data Analysis 51: 2379–2409.

63. Breiman L (2001) Random forests. Machine Learning 45: 5–32.

64. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2: 121–167.

65. Warner B, Misra M (1996) Understanding neural networks as statistical tools. The American Statistician 50: 284–293.

66. Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. Neurocomputing 55: 169–186.

67. Jaiantilal A (2009). Classification and regression by randomforest-matlab. URL http://code.google.com/p/randomforest-matlab.

68. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2: 27: 1–27: 27.

69. McSherry F, Najork M (2008) Computing information retrieval performance measures efficiently in the presence of tied scores. In: Macdonald C, Ounis I, Plachouras V, Ruthven I, White R, editors, Advances in Information Retrieval, Springer Berlin/Heidelberg, volume 4956 of Lecture Notes in Computer Science. 414–421. URL http://dx.doi.org/10.1007/978-3-540-78646-7_38.

70. Kohavi R, Provost F (1998) Glossary and terms. Maching Learning 30: 271–274.

71. Siegel S (1956) Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

72. Whitehead H, Jonsen ID (2013) Inferring animal densities from tracking data using markov chains. PloS one 8: e60901.

73. Jonsen I, Basson M, Bestley S, Bravington M, Patterson T, et al. (2013) State-space models for bio-loggers: A methodological road map. Deep-Sea Research Part II 88–89: 34–46.

74. Sacks J, Ylvisaker D (2012) After 50+ Years in Statistics, An Exchange. Statistical Science 27: 308–318.

75. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39: 1–38.

76. Heskes T, WigerinckW(1996) A theoretical comparison of batch-mode, on-line, cyclic, and almostcyclic learning. IEEE transactions on neural networks 7: 919–25.

77. Bourlard H, Morgan N (1994) Connectionist speech recognition: A hybrid approach. Kluwer Academic Publishers.

78. Bourlard H, Morgan N (1998) Hybrid HMM/ANN systems for speech recognition : Overview and new research directions. In: Adaptive Processing of Sequences and Data Structures, ser. Lecture Notes in Artificial Intelligence.

79. Ganapathiraju A, Hamaker J, Picone J (2000) Hybrid SVM/HMM architectures for speech recognition. ICSLP 4: 504–507.

80. Stadermann J, Rigoll G (2004) A hybrid SVM/HMM acoustic modeling approach to automatic speech recognition. In: ICSLP - INTERSPEECH. Jeju Island, Korea.