Short running title: Hidden Markov models for accelerometer data

Analysis of animal accelerometer data using hidden Markov models

Vianey Leos-Barajas[a*], Theoni Photopoulou[b], Roland Langrock[c], Toby A. Patterson[d], Yuuki Y. Watanabe[e,f], Megan Murgatroyd[g] and Yannis P. Papastamatiou[h,i]

[a]Department of Statistics
Iowa State University
Ames, IA, USA

[b]Centre for Statistics in Ecology, Environment and Conservation
Department of Statistical Sciences
University of Cape Town
Cape Town, South Africa

[c]Department of Business Administration and Economics
Bielefeld University
Bielefeld, Germany

[d]CSIRO, Oceans and Atmosphere
Hobart, Tasmania, Australia

[e]National Institute of Polar Research
Tachikawa, Tokyo, Japan

[f]SOKENDAI (The Graduate University for Advanced Studies)
Tachikawa, Tokyo, Japan

[g]Animal Demography Unit
Department of Biological Sciences

University of Cape Town
Cape Town, South Africa

[h]School of Biology, Scottish Oceans Institute
University of St Andrews
St Andrews, UK

[i]Department of Biological Sciences
Florida International University
North Miami, Fl, USA

Word Count: –

*corresponding author:
e-mail: vianey@iastate.edu
address: Snedecor Hall, Ames IA, Zip:50011
phone number: 515-509-5388

Abstract

1. Use of accelerometers is now widespread within animal biotelemetry as they provide a means of measuring an animal's activity in a meaningful and quantitative way where direct observation is not possible. In sequential acceleration data there is a natural dependence between observations of behaviour, a fact that has been largely ignored in most analyses.

2. Analyses of acceleration data where serial dependence has been explicitly modelled have largely relied on hidden Markov models (HMMs). Depending on the aim of an analysis, an HMM can be used for state prediction or to make inferences about drivers of behaviour. For state prediction, a supervised learning approach can be applied. That is, an HMM is trained to classify unlabelled acceleration data into a finite set of pre-specified categories. An unsupervised learning approach can be used to infer new aspects of animal behaviour when biologically meaningful response variables are used, with the caveat that the states may not map to specific behaviours.

3. We will provide the details necessary to implement and assess an HMM in both the supervised and unsupervised learning context and discuss the data requirements of each case. We outline two applications to marine and aerial systems (shark and eagle) taking the unsupervised learning approach, which is more readily applicable to animal activity measured in the field. HMMs were used to infer the effects of temporal, atmospheric and tidal inputs on animal behaviour.

4. Animal accelerometer data allow ecologists to identify important correlates and drivers of animal activity (and hence behaviour). The HMM framework is well suited to deal with the main features commonly observed in accelerometer data, and can easily be extended to suit a wide range of types of animal activity data. The ability to combine direct observations of animal activity with statistical models, which account for the features of accelerometer data, offers a new way to quantify animal behaviour, energetic expenditure and deepen our insights into individual behaviour as a constituent of populations and ecosystems.

Keywords: animal behaviour; activity recognition; latent states; serial correlation; time series

1 Introduction

Accelerometers are becoming more prevalent in the fields of animal and human bio-logging (Bao & Intille, 2004; Ravi *et al.*, 2005; Shepard *et al.*, 2008; Altun *et al.*, 2010). The potential of accelerometers lies in the fact that they provide a means of measuring activity in a meaningful and quantitative way where direct observation is not possible (Shepard *et al.*, 2008; Nathan *et al.*, 2012; Brown *et al.*, 2013). While these instruments are cheap and compact, recording acceleration at a high temporal resolution and in up to three dimensions quickly results in terabytes of data that present various challenges regarding transmission, storage, processing and statistical modelling.

Much of the focus in the analysis of acceleration data has been on identifying patterns in the observed waveforms that correspond to a known behaviour or movement mode. This can be achieved by employing statistical classification methods and can entail observing the animal, manually assigning labels corresponding to behaviours to segments of the data and training a model using the labelled

data in order to subsequently classify remaining unlabelled data. Many studies that have shown the effectiveness of various machine learning algorithms for classification of human acceleration data (Bao & Intille, 2004; Ravi *et al.*, 2005; Altun *et al.*, 2010; Mannini & Sabatini, 2010). Algorithms such as support vector machines (SVM), classification trees, random forests, among others, have also recently been used for classification of animal acceleration data (Nathan *et al.*, 2012; Carroll *et al.*, 2014; Graf *et al.*, 2015). For example, Nathan *et al.* (2012) compared the effectiveness of five machine learning algorithms to distinguish between eating, running, standing, active flight, passive flight, general preening and lying down, for griffon vultures.

Most machine learning algorithms assume independence between individual observations. However, in sequential acceleration data there is a natural dependence between observations of behaviour — once initiated, particular animal behaviours often last for periods longer than the sampling frequency. This fact has been largely ignored in most applications of classification approaches. The studies where serial dependence has been explicitly modelled have mostly relied on hidden Markov models (HMMs) (Ward *et al.*, 2006; He *et al.*, 2007; Mannini & Sabatini, 2010, 2011). HMMs are stochastic time series models which assume that the observed time series, the so-called state-dependent process, is driven by an unobservable state process. In this scenario, the former corresponds to the acceleration data and the latter to the behavioural classes. Typically, and in common with the aforementioned machine learning approaches, in the training stage, the states of the HMM were known a priori, requiring corresponding data derived from direct observations.

There are two main difficulties with such a supervised learning approach. First, while there has been much success in classification of human acceleration data, where training data can usually be obtained with minimal effort, this may not be feasible for some animals. Humans can easily be observed in a laboratory setting, given instructions or monitored in more realistic settings, such as walking outdoors or in their home (e.g. Leenders *et al.*, 2000). In certain cases, animals can also be monitored in a laboratory setting (Wilson *et al.*, 2008), but movement patterns recorded in the lab from free-ranging animals may not appear exactly the same as in data collected while in more natural settings. Conversely, many behaviours can only be observed in natural settings, although there has been success using surrogate species for classification of behavioural modes (Shepard *et al.*, 2008; Nathan *et al.*, 2012; Campbell *et al.*, 2014; Brown *et al.*, 2013).

Second, human acceleration data has commonly been used as a tool for health monitoring and other situations where the focus is on (state) prediction, as opposed to learning how external factors drive the behaviours. Classification of behaviours *alone*, while certainly of interest in many scenarios, may not lead to biologically interesting inference. Once the classification has been done, the task of relating these states to environmental (and other) covariates in order to identify drivers in behaviours remains. Moreover, it is difficult to make appropriate inferential statements as the classifications are not without error, propagating the state uncertainty through to the modelled effect of the covariates.

In the supervised learning context, i.e. when classification is the main purpose of an analysis, we train

the HMM to recognize specific behaviours. Alternatively, HMMs can also be used in an unsupervised learning context, i.e. when there are no labelled data. In an unsupervised learning context the states are not pre-defined to represent a specific behaviour. Instead, the states will be allocated such that the model captures as much as possible of the marginal distribution of the observations, i.e. the distribution of an observation at a randomly chosen time point, not conditional on the previous history of the process, as well as their correlation structure. If biologically meaningful response variables from the acceleration data are considered, then the HMM states will usually represent interpretable activity levels or even proxies of behavioural modes. Being data-driven the states can be as, if not more, informative in the unsupervised learning setting than the alternatives. We can then incorporate exogenous or, where available, endogenous variable(s) of interest, to make inferential statements. HMMs and related state-switching models, in particular state-space models, have successfully been implemented to identify drivers of movement based on tracking data (Patterson *et al.*, 2009), and can similarly be applied in the context of accelerometer data. For example, Phillips *et al.* (2015) applied HMMs in an unsupervised learning context to understand the behaviour of free swimming tuna from vertical movement data collected by data-storage tags. We will implement an unsupervised learning approach for another difficult to observe marine species, the blacktip reef shark, and a volant species, the black eagle.

In this paper we review HMM-based approaches to the analysis of animal accelerometer data. In Section 2 we will provide an overview of accelerometer data and connect the data processing step to the HMM-based approaches described in Section 3. We will typically refer to the term *behavioural* class, rather than differentiate between identification of specific movements (e.g. wing flapping) or behaviours (e.g. foraging). In Section 4 we demonstrate the use of HMMs with real data examples from marine and aerial systems.

2 Accelerometer data

Accelerometer devices measure in up to three axes, which can be described relative to the body of the animal; longitudinal (surge), lateral (sway) and dorso-ventral (heave). Acceleration recorded along one or two axes can be used to measure movement in parts of the body, e.g. the mandible (Suzuki *et al.*, 2009; Naito *et al.*, 2010; Iwata *et al.*, 2015), or aspects of whole body acceleration, e.g. longitudinal surge (Sakamoto *et al.*, 2009). Currently, acceleration is most commonly recorded in three axes and, to a lesser degree, in two axes (Brown *et al.*, 2013), to measure locomotion.

2.1 Data Processing for Classification

While the observed acceleration data can be used to identify specific movements in animals, HMMs and other machine learning algorithms require more information to accurately classify the unlabelled data. These methods require appropriate features, i.e. summary statistics, from a window (or sliding window)

of observations. The derived features should be driven by the classes of movements that have been defined and chosen in such a way to accentuate the differences in observed acceleration measurements. There are many commonalities between the features used in applications of classification of acceleration data, though naturally no one optimal set exists (Bao & Intille, 2004; Martiskainen *et al.*, 2009; Nathan *et al.*, 2012; Brown *et al.*, 2013). For instance, Nathan *et al.* (2012) used thirty-eight features in order to distinguish between eating, running, standing, active flight, passive flight, general preening and lying down, for griffon vultures, while Graf *et al.* (2015) used eight features to distinguish between standing, walking, swimming, feeding, diving and grooming of Eurasian beavers. In each case, means and variances of each of the three axes are used, as well as overall dynamic body acceleration (ODBA), the sum of dynamic body acceleration from the three axis, among others.

2.2 Connecting Measures to Behaviours

When the aim is to classify the acceleration data, data processing is driven by identifying a set of features that can be used to distinguish between specific behaviours, even if those features are not themselves interpretable as a specific behaviour when considered on their own. However, there are metrics derived from accelerometer data that, on their own, can be used as proxies for behaviour and as input to an HMM. Repeating patterns in at least one axis tend to arise from behaviours such as stroking (Sakamoto *et al.*, 2009), flapping, running or walking (Shepard *et al.*, 2008), whereas sudden changes, corresponding to bursts of acceleration, are often associated with prey pursuits or capture (Suzuki *et al.*, 2009; Simon *et al.*, 2012; Ydesen *et al.*, 2014; Heerah *et al.*, 2014), as well as predator avoidance or conflict.

In addition to behaviour, several measures can be used to summarise effort or exertion and relate acceleration to activity levels, such as ODBA (Wilson *et al.*, 2006; Gleiss *et al.*, 2011; Elliott *et al.*, 2013; Gleiss *et al.*, 2013) and vectorial dynamic body acceleration (VeDBA) (Qasem *et al.*, 2012). Minimum specific acceleration (MSA) (Simon *et al.*, 2012) can be used to disentangle the gravitational component of acceleration (static acceleration) from the movement signal or specific acceleration (also dynamic acceleration). One of the simplest and most unambiguous interpretations of static acceleration data is body posture, which in many cases can be directly interpreted as a specific behaviour (Wilson *et al.*, 2008; Shepard *et al.*, 2008).

Both ODBA and MSA are used to reduce the dimensionality of 3D acceleration data while retaining important information (e.g. Wilson *et al.* (2008); Simon *et al.* (2012)). They remove the gravitational component from the acceleration signature and produce an overall value of the dynamic acceleration experienced by the animal. ODBA is derived by smoothing over a time period, e.g. 1 sec, making it useful for continuous data, whereas MSA is calculated point-wise (as the norm of the three vectors minus 1 for the effect of gravity) and is more suited to lower resolution acceleration data.

3 Analysis of accelerometer data

We will first provide a brief overview of the HMM framework (Section 3.1). Subsequently, in Section 3.2, we will review how HMMs can be used for state prediction, i.e. classification of animal accelerometer data. In Section 3.3, we focus on the implementation of HMMs in a setting where the meaning of the states is driven entirely by the data and the focus lies on general inference rather than classification only.

3.1 Hidden Markov models

An HMM is a stochastic time series model involving two layers: an observable *state-dependent process*, denoted by $\{Y_t\}_{t=1}^T$ (in the univariate case), and an unobservable *state process*, denoted by $\{C_t\}_{t=1}^T$. The state-dependent process models the observations, while the state process is a latent factor influencing the distribution of the observations. In our case, the observations are the accelerometer metrics considered, and the latent states are closely related to the animal's behavioural state. More specifically, the state process $\{C_t\}$ takes on a finite number of possible values, $1, \ldots, M$, and its value at time t, c_t, selects which of M possible component distributions generates observation y_t. The Markov property is assumed for $\{C_t\}$, i.e. the (behavioural) state at time t only depends on the (behavioural) state at time $t-1$, such that evolution of the process over time is completely characterized by the one-step state transition probabilities. These models are natural and intuitive candidates for modelling animal accelerometer data, for two reasons: 1) they directly account for the fact that any corresponding observation will be driven by the underlying behavioural state, or general activity level, of the animal, and 2) they accommodate serial correlation in the time series by allowing states to be persistent. HMMs seek to capture the strong autocorrelation in accelerometer data in a mechanistic way, rather than either neglecting this feature completely or only including it in a nuisance error term. HMMs can therefore be used for inference on complex temporal patterns, including the behavioural state-switching dynamics and how these are driven by environmental variables (Patterson *et al.*, 2009; McKellar *et al.*, 2015).

To complete the basic HMM formulation, we first summarize the probabilities of transitions between the different states in the $M \times M$ transition probability matrix (t.p.m.) $\mathbf{\Gamma} = (\gamma_{ij})$, where $\gamma_{ij} = \Pr\big(C_{t+1} = j | C_t = i\big)$ (for any t), $i, j = 1, \ldots, M$. Note that here we are assuming that the state transition probabilities are constant over time; this assumption will be relaxed in Section 3.3. The initial state probabilities are summarized in the row vector $\boldsymbol{\delta}$, where $\delta_i = \Pr(C_1 = i)$, $i = 1, \ldots, M$.

Second, we need to specify state-dependent distributions (sometimes called emission distributions), $p(y_t | C_t = m)$, or more succinctly $p_m(y_t)$, for $m = 1, ..., M$. These distributions can be discrete or continuous, and possibly also multivariate (in which case we write $\mathbf{y}_t = (y_{1t}, \ldots, y_{Rt})$). Usually, the same parametric distribution is assigned to all M states, such that each state differs in terms of its associated values of the parameters. Selection is driven by the data itself, e.g. count data or continuous observations.

3.2 State prediction

HMMs provide a solid framework for the classification of data with strong serial dependence, such as sequential acceleration data, which are often processed to represent movements over a few seconds, or less, at a time (Ward *et al.*, 2006; He *et al.*, 2007; Mannini & Sabatini, 2010). In this section, we will cover the implementation and testing of an HMM when the focus of the analysis is state prediction. A full example and R code implementing this approach is provided in the Supplementary material.

State prediction can be accomplished in three manners, commonly referred to as supervised, semi-supervised, or unsupervised learning. We will discuss the implementation of an HMM in the supervised learning case, such that each state will correspond to one behaviour of interest, and briefly comment on the other two cases at the end of the section. Hastie *et al.* (2001) detail how to split the labelled time series into training, validation, and testing data, in order to estimate the prediction error. Other approaches to estimating prediction error, such as a leave-one-out cross-validation (here treating a time series as an observation), are also provided in detail.

Since the states are known, the maximum likelihood estimates (MLEs) of the HMM parameters are obtained by maximizing the complete-data likelihood, which conveniently splits into several independent parts, each of which is fairly straightforward to maximize (details provided in the Appendix). First, the m-th entry of $\hat{\boldsymbol{\delta}}$ is simply the proportion of the time series that start in state m. Second, the entries of the t.p.m. are estimated by

$$\hat{\gamma}_{ij} = \frac{\#\text{ transitions from state } i \text{ to state } j}{\text{total } \#\text{ transitions from state } i},$$

for $i, j = 1, ..., M$. (Note this is the MLE conditional on the initial state, c_1.) Finally, for each $m = 1, \ldots, M$, the parameters of the state-dependent distribution given state m are estimated using only the observations allocated to state m. As a multivariate normal distribution (MVN) is a common choice in these cases, we cover the steps to fit the HMM with MVN state-dependent distributions in the Appendix and Supplementary material. Given a fitted HMM, we can use the Viterbi algorithm to decode the most likely state sequence, thereby assigning each observation to a state, at low computational effort. Full details for state decoding are provided in Zucchini *et al.* (2016). The state predictions can be compared to the known states, and the proportion of correctly decoded states serves as an estimate of the prediction accuracy.

As mentioned previously, there are two other approaches to state-prediction: semi-supervised and unsupervised learning. In a semi-supervised approach, classes are pre-defined, as in the supervised learning context, but there is additional flexibility provided in that the data do not have to be assigned to one of the pre-defined classes. Instead, multiple additional states can be estimated from the data. In an unsupervised learning approach, classes are not pre-defined in any manner. In these two cases, one objective can be to identify the number of distinct movement patterns exhibited by the animal, with the resulting estimated HMM states depending on the features selected for interpretation. However, as

multiple movement modes can correspond to the same behaviour (e.g foraging), interpretation of the estimated states should be made with caution. In the next section, we will detail the implementation of the unsupervised learning approach where the focus is to construct biologically relevant classes of animal behaviour in order to make inferential statements.

3.3 Inference

So far, we have mostly focused on the case where there is a training sample, i.e. acceleration data together with the associated behavioural states. Corresponding analyses involve training the HMM based on such labelled data and then using that HMM to categorize incoming new, unlabelled data. While certainly of interest in some settings, in practice, more often than not, labelled data will not be available but only the accelerometer data. In such unsupervised learning settings, the HMM framework can be equally useful, but is typically applied for different purposes than in classification. More specifically, the meaning of the states in such cases is often not of interest *per se*. Instead, an HMM is used simply as an approximate representation of the real data-generating process, and this may or may not entail that the nominal HMM states are biologically meaningful. (However, metrics derived from the accelerometer data, as described in Section 2, have been shown to provide insight into activity levels or correspond to classes of behaviours, such that when used as response variables in the HMM these can lead to biologically interpretable states.) Unsupervised learning of HMMs for accelerometer data has the advantage that the states are estimated in a data-driven manner. In particular, for many of the metrics described in Section 2 that are connected to behaviours, assignment of classes is difficult, to say the least, especially for animals where behaviours are not well-defined. These include animals which cannot be directly observed for long periods such as aquatic organisms.

There are three different possible purposes of having an approximate representation of the real process: (i) a mathematical description of the dynamics of the system (e.g. in order to have a concise description of how accelerometer measurements evolve over time, in terms of a small number of interpretable parameters and associated stochastic distributions); (ii) extraction of information (e.g. a hypothesis test on whether or not some environmental covariate increases the probability of an animal switching to a particular behavioural state); (iii) prediction of future or missing values (e.g. behavioural state prediction given accelerometer data) — see Konishi & Kitagawa (2008). In the ecological literature on animal movement modelling, HMMs are used primarily to address (i) and (ii), the former in the sense that concise descriptions of movement patterns are sought, the latter in the sense that inference on the interaction of animals with their environment is drawn. In general, the ability to make inferential statements provides an avenue to answer questions about the behavioural processes, movement patterns and transitions between behaviours under different in relation to covariates.

Addressing a research question related to aim (ii) usually involves the incorporation of covariates into the statistical model. In the HMM setting, this is commonly done at the level of the hidden states. For the general case of time-varying covariates, we define the corresponding time-dependent transition

probability matrix $\mathbf{\Gamma}^{(t)} = (\gamma_{ij}^{(t)})$, where $\gamma_{ij}^{(t)} = \Pr(C_{t+1} = j | C_t = i)$. The transition probabilities at time t, $\gamma_{ij}^{(t)}$, can then be related to a vector of environmental (or other) covariates, $\left(\omega_1^{(t)}, \ldots, \omega_p^{(t)}\right)$, via the multinomial logit link:

$$\gamma_{ij}^{(t)} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{N} \exp(\eta_{ik})}, \qquad \text{where} \qquad \eta_{ij} = \begin{cases} \beta_0^{(ij)} + \sum_{l=1}^{p} \beta_l^{(ij)} \omega_l^{(t)} & \text{if } i \neq j; \\ 0 & \text{otherwise.} \end{cases}$$

Essentially there is one multinomial logit link specification for each row of the matrix $\mathbf{\Gamma}^{(t)}$, and the entries on the diagonal of the matrix serve as reference categories.

While with labelled data the likelihood of interest is the complete-data likelihood, for unlabelled data the likelihood of interest is the density of the observations only, $L = p(\mathbf{y}_1, \ldots, \mathbf{y}_T)$, the evaluation of which requires the consideration of all possible state sequences that might have given rise to these data. The powerful forward algorithm, detailed in the Appendix, can be applied to accomplish this, opening up a straightforward and usually feasible avenue to MLEs, namely direct numerical maximization of the likelihood. In practice, one needs to consider multiple starting values in order to make sure to have found the global maximum. The Expectation-Maximization algorithm provides a popular alternative route to MLEs, despite being much more technically involved and having no clear practical advantages (MacDonald, 2014). Since it is our view that users are better off focusing on the simpler direct maximization approach, it is only this approach that we present in detail in the Appendix and Supplementary material (for a more comprehensive introduction to maximum likelihood estimation for HMMs, see Zucchini *et al.*, 2016).

Model selection techniques, in particular information criteria, can be used to choose an adequate family of state-dependent distributions, to select an appropriate number of states or to determine whether or not a covariate should be included in the model. However, users should not blindly follow such information criteria, especially with regard to the selection of the number of states. For animal behaviour data, it is our experience that such formal model selection approaches tend to favour models with more states than would be expected based on biological intuition, often to an extent such that selected models become near-impossible to interpret and very difficult to work with in practice (Langrock *et al.*, 2015). One explanation for this is that often additional states are included to compensate for a model formulation that ignores some pattern in the data. These patterns can be due to the influence of an unobserved covariate, within-day variation or individual heterogeneity which is not accounted for, a violation of the Markov assumption or outliers — which usually cannot be avoided in data structures as complex as those studied here, and which may not be pertinent to the ultimate aim of the study. Further, accelerometer data is directly connected to the movement of an animal, such that an HMM with a large number of states may reflect multiple movement modes, or general classes of movement, connected to the same behavioural class, e.g. foraging or active behaviour. In such cases a healthy dose of pragmatism is required. If the choice of the number of states turns out to be difficult, then it is often useful to carefully examine all plausible models (with lower and higher numbers of states), e.g. using

model checking tools, in order to understand what exactly it is that the more complex models capture that is not already captured by the simpler models. Langrock *et al.* (2015) discuss this issue in detail, demonstrating many of the points made above in a real data example.

The HMM framework encompasses various other useful tools for drawing inference. In particular, incorporating random effects into the model formulation will be crucial when there is substantial heterogeneity across multiple individuals observed. There are various ways in which this can be accomplished within the class of HMMs — see McKellar *et al.* (2015) and Chapter 13 in Zucchini *et al.* (2016) for comprehensive overviews, including discussions on the importance of acknowledging any potential heterogeneity. Furthermore, the dependence structure can be modified in various ways, e.g. allowing for more complex memory in the state process without losing the ability to efficiently calculate the likelihood using the forward algorithm (Langrock *et al.*, 2012). Assessment of the model adequacy, i.e. model checking, is commonly done using (pseudo-)residuals, which can reveal any notable lack of fit (Zucchini *et al.*, 2016).

4 Real data examples

4.1 Modelling activity in a soaring raptor

Large soaring birds, like raptors, depend on favourable meteorological conditions, as well as the underlying topography, for generation of updrafts required for low-energy flight (Pennycuick, 2008). Lift availability is known to be driven largely by wind speed and temperature, as well as their interaction with the underlying topography, though other factors also contribute. Lift adequate for soaring flight is generated by two mechanisms; (1) by upward thermal convection of air warmed by solar radiation (Ákos *et al.*, 2010) (thermal soaring), and (2) by the movement of air over slopes and ridges in the landscape (orographic or ridge soaring).

Recently, empirical studies relating bird activity patterns to weather conditions have become possible due to advances in bio-logging technology that allows for collection of high-resolution movement (e.g. acceleration) data. In particular, acceleration data can be used to distinguish between different movement modes or, more simply, as a proxy of overall activity level, even if they do not correspond clearly to different behaviours (Williams *et al.*, 2015).

An adult Verreaux's eagle *(Aquila verreauxii)* was instrumented with a remotely downloadable multi-sensor data-logger (UvABiTS, University of Amsterdam, The Netherlands, Bouten *et al.* (2013)) in the Western Cape, South Africa, in 2013. The data-logger recorded 3D acceleration (at 20 Hz) for 1 second directly after recording GPS location. The GPS location sampling rate depended on the solar-powered battery charge and thus was higher during the mid-day. Data were collected over 9 consecutive days, with a variable amount of acceleration data sampled each day and none collected overnight.

We were primarily interested in identifying potential drivers of activity level. As such, we extracted the MSA, which serves as an index of activity, over each 1 second sample of acceleration data recorded.

On average, each day produced approximately 135 observations (s.e. 23.32). Before fitting an HMM to the time series of MSA values, we first needed to resolve the irregular sampling of the acceleration data, as this is a clear violation of the HMM assumptions. The time series of MSA across days were taken to be independent and, within a day, the acceleration data was subsampled to produce one value of MSA every 112 seconds. Only 1 consecutive missing value was allowed before splitting the daily MSA time series into two or more segments.

The histogram of MSA values revealed two peaks close to zero, which may reflect general low-active behaviours such as roosting and preening. As we did not wish to discriminate between these two general types of behaviours, we fit a 2-state HMM with state 1 represented by a mixture of gamma distributions and a gamma distribution for state 2. The fitted state-dependent densities are shown in Figure 2, which we *post-hoc* interpreted as low-activity and high-activity behaviour. Although we do not connect state 2 to a specific flight behaviour, such as orographic soaring, we expect that behaviours requiring more energy are reflected by larger MSA values.

In order to examine the effect of wind speed and temperature on the state-switching dynamics between the two activity levels, we obtained hourly observations from the South African Weather Services (Lambert's Bay Station). The station is approximately 30 km from the general area in which the eagle was tracked, which lead to a slight spatial and temporal mismatch between the available weather data and the conditions actually experienced by the eagle. The range of temperatures and wind speeds experienced by the eagle during the study period was between 12.3–31.5 °C, and 0–7.4 m/sec, respectively. We allowed the entries of the t.p.m. to be a function of up to wind speed, temperature and their interaction. The wind-only model is written as $\text{logit}(\gamma_{ij}(t)) = \beta_{0i} + \beta_{1i}x_{1t}$, for $i = 1, 2$, $j \neq i$, $t = 1, \ldots, T$, with the intercept term $\beta_{0,i}$ reflecting the t.p.m. when wind speed is at 0 m/sec. The model including wind speed alone was favoured by the Bayesian Information Criterion (BIC) and the full model, with temperature and the interaction term, favoured by the Akaike Information Criterion (AIC) (Table 1). After examination of the (pseudo-)residuals of the models selected by AIC and BIC, we selected the model favoured by BIC as there was a similar lack-of-fit evident in both models. Further, we may not have captured a large enough range of temperatures in order to make general inferences about its effect on the activity levels of the eagle, and as such were cautious of over-fitting or over-interpreting the model results. We present confidence intervals and a plot of the (pseudo-)residuals for assessment of goodness-of-fit in the Appendix for the model with only wind speed included. R code to simulate MSA data and fit a 2-state HMM with the t.p.m. entries as functions of wind speed is included in the Supplementary material.

The estimated state transition probabilities suggest that, as wind speed increases, (i) the eagle has a very slightly increased chance of switching to the high-activity state when in the low-activity state, and (ii) spends much longer periods of time, on average, in the active state. As a consequence, the equilibrium (stationary) distribution for fixed wind speeds (Patterson *et al.*, 2009) indicates that the eagle spends more time in the active state overall as wind speed increases (Figure 3). Windier conditions favour

orographic soaring, as demonstrated by studies on migrating golden eagles *Aquila chrysaetos*, which is a more active behaviour (Lanzone *et al.*, 2012). There is also theoretical evidence to suggest that, in general, flying is more energetically demanding in high winds (Pennycuick, 1972).

4.2 Diel activity changes in a reef-associated shark

Many species of shark are upper trophic level predators may serve an important role in marine ecosystems. However determining the intensity of their predatory behaviour requires modelling the temporal component as their activity levels are likely to follow a diel and/or tidal cycle (e.g. Gleiss *et al.*, 2013; Papastamatiou *et al.*, 2015). Acceleration sensors provide a direct measure of activity, however, many species of shark swim continuously making it difficult to define specific behaviours (e.g. they are never truly at rest), making conventional classification methods problematic. HMMs can identify changes in behavioural states and how these may be related to time of day, tidal state, swimming depth, or water temperature. To demonstrate this, we applied HMMs to accelerometry data obtained from a free-ranging blacktip reef shark (*Carcharhinus melanopterus*) at Palmyra Atoll in the central Pacific ocean (data taken from Papastamatiou *et al.*, 2015). A multi-sensor package was attached to the dorsal fin of a 117 cm female shark. The multi-sensor data-logger (ORI400-D3GT, Little Leonardo, Tokyo, Japan) recorded 3D acceleration (at 20 Hz), depth and water temperature (at 1 Hz) and was embedded in a foam float which detached from the shark after four days (see Papastamatiou *et al.*, 2015). The package also contained a VHF transmitter allowing recovery at the surface after detachment.

In order to examine active behaviour, we calculated the average ODBA of the shark over 1 second intervals, which resulted in 321,815 observations (after removing the first four hours of data). Figure 4 displays the ODBA time series of one day. Compared to metrics such as tail-beat frequency, ODBA has the advantage of measuring change in behaviour in all axes. For example, if the shark is nose down at the seafloor, attempting to capture prey, its tail-beat frequency may be low but it is still active. As we are interested in the times of day the shark was more active, as well as tide effects, we applied a 2-state HMM with one state *post-hoc* interpreted as representing less active behaviour and the other more active behaviour.

Although there are clear spikes in ODBA that point to higher energetic activities, various combinations of parametric distributions for state 1 and 2 led to vastly different state-dependent densities. Further, the ODBA values had many extreme values that needed to be accommodated, which further increased the difficulties of selecting appropriate state-dependent distributions. As ODBA is not a metric that can easily be divided into active/inactive behaviours in sharks, we estimated the state-dependent densities nonparametrically, in both states, in order to minimize the bias introduced by assigning inadequate parametric distributions (Langrock *et al.*, 2015). Figure 5 displays the fitted distributions.

To examine potential diel and tide effects on activity levels, we let the entries of the t.p.m. be functions of up to two covariates: time of day and tide level (*ebb, flood, low, and high*). Tide data was obtained from the NOAA tides and currents website for Palmyra Atoll and was processed by denoting

high or low tide as ±1 hour from reported high or low tide times. Time of day is represented by two trigonometric functions with period 24 hours, cos(2πt/86400) and sin(2πt/86400) (86,400 is the number of seconds in a day). We use three indicator variables, x_{1t}, x_{2t} and x_{3t}, for tide levels *high*, *flood*, and *ebb*, respectively, such that $x_{1t} = 1$ when tide level is high and $x_{1t} = 0$ otherwise, and so on, which gives the entries of the t.p.m. the following form

$$\text{logit}(\gamma_{ij}(t)) = \beta_{0i} + \beta_{1i}\cos(2\pi t/86400) + \beta_{2i}\sin(2\pi t/86400) + \beta_{3i}x_{1t} + \beta_{4i}x_{2t} + \beta_{5i}x_{3t}$$

for $i = 1, 2$, $j \neq i$, $t = 1, \ldots, 86400$. The intercept term $\beta_{0,i}$ corresponds to *low* tide.

Based on the selected model (cf. Table 2), with confidence intervals and (pseudo-)residuals provided in the Appendix, the shark's activity levels were, on average, lowest from approximately 9:00 – 13:00 and highest from 21:00 – 1:00. In Figure 6, we see that the shark was more active during high tide in general when compared to flood, low or ebb tide. While the equilibrium (or stationary) distribution associated with low and ebb tide overlap, the state-dwell probabilities, i.e. the diagonal entries of the t.p.m. corresponding to the probability of remaining in the same state, are higher during ebb tide than in low tide. Naturally in a short time series the tide levels will be correlated with certain times of the day, but a longer time series or a joint modelling of multiple time series, with tide levels observed during all times of day, can provide robust estimates of the effect of tide on activity level using the HMM formulation provided here.

Using the Viterbi algorithm, we decoded the optimal state sequence to underlie the ODBA time series. To further understand the effect of vertical habitat on behaviour, we related the decoded state sequence to a grid of depth and temperature values, shown in Figure 7. The shark spent most of its time over the nearly five day period in depths of about 3-6 metres and between 28-29 °C, with some higher counts also in shallower waters, which is reflected in the state 2 counts. However, the percentages of state 2 observations reveals that the shark was generally more active when near the surface in waters of 28-29 °C. There was generally less active behaviour exhibited when the individual was in very shallow warm water (> 29 °C).

5 Discussion

We detailed two approaches for analysing animal accelerometer data with HMMs: a supervised learning approach for state prediction, such that classification is of primary interest, and an unsupervised learning approach, where the states reflect biologically meaningful classes of behaviour, in order to infer drivers of animal behaviour. The aim of a study and the type of data available will determine which of the two is to be preferred. When the objective is to do classification and there is a set of pre-defined behaviours of interest, then the model's ability to correctly predict and categorize behaviours is of main interest. In this instance, a supervised learning approach may be applied. One of the benefits of such an approach is that the behavioural classes are exactly defined, making interpretation relatively straightforward.

Alternatively, if the objective is to infer (or, colloquially speaking, to 'learn') new aspects of animal behaviour, then the unsupervised learning approach provides an excellent framework. The latter comes with the implicit caveat that the states will not necessarily map directly to specific animal behaviours. Any post-hoc behavioural interpretation of the estimated states is directly connected to the metric(s) used, and must draw from background biological knowledge of the species of interest. In many cases, behaviours such as foraging may not be exclusive to one state or another. Nonetheless, if the model is able to identify bouts of behaviour which consistently re-appear, then it is often likely that these signify something important in the animal's behavioural repertoire and are worthy of further investigation.

Even when classification is the goal of an analysis, there are certainly practical scenarios which preclude the use of an HMM, e.g. if the training data do not reflect the transitions between behaviours or if there is insufficient data. Moreover, multiple studies have shown that other machine learning algorithms, e.g. support vector machines (SVM) or random forests, can work well for classification of animal acceleration data (Martiskainen *et al.*, 2009; Nathan *et al.*, 2012; Carroll *et al.*, 2014; Graf *et al.*, 2015). However, disregarding the serial dependence in the acceleration data usually is an unrealistic assumption, which often goes unmentioned or is treated as an afterthought. Adopting the assumption of independence is particularly risky if inferential statistics are applied to the output of say a machine learning algorithm. In these cases, secondarily applied statistical tests will implicitly assume that the machine learning categorizations contain more information content than is warranted, potentially leading to spurious results. This is not just a statistical nuance and can be a crucial point. Such tests are often applied as decision making tools to sort out "what matters" and setting the direction for much further research effort. Also, in assuming independence, one allows for classifications that may not be biologically realistic or must filter the classifications to properly identify a specific behaviour. For instance, Carroll *et al.* (2014) used a SVM where one of the primary interests was to identify prey handling/capture for penguins at sea. To confirm a prey capture event, they ruled that if the SVM classified three consecutive observations as prey-handling this counted as a true prey capture. In contrast, an HMM would have bypassed the need to filter through the classification results by accounting for the serial dependence in observations corresponding to prey handling. In general, many behaviours persist over longer stretches of time than those at which the data is processed, also necessitating the use of a model that can account for the serial dependence. It may be difficult for any machine learning algorithm that assumes independence to properly classify a sequence of observations into the same class, unless the boundaries between classes are well-defined. In the context of recognition tasks, e.g. speech or pattern recognition, HMMs have proven to be extremely successful tools for classification precisely because they do account for the serial dependence in the signal of interest (Rabiner, 1989).

In the literature, inference on behavioural state-switching dynamics has sometimes been made using two-stage (or even three-stage) analyses, where HMMs (or other machine learning algorithms) are used to decode the behaviours underlying given observations, and subsequently a logistic regression is conducted for relating the decoded behaviours to covariates (see, e.g., Hart *et al.*, 2010; Broekhuis *et al.*, 2014). The

appeal of such an approach lies in the ease of implementation: fairly basic HMMs, without covariates, are fitted to the accelerometer data and used to decode the states, and, subsequently, standard regression software packages can be used to conduct a regression of the behavioural states on covariates. However, it is our view that such a multi-stage analysis is less suited to relating accelerometer data to covariates than the *joint* modelling approach presented in Section 3.3, for two reasons: (i) in the multi-stage analyses, the uncertainty in state estimates is usually not propagated through the different stages of analysis, and (ii) a regression analysis on decoded states needs to take into account the high serial correlation in those states. Rather than ignoring these issues or trying to address them within a multi-stage analysis (which will render such an approach technically challenging), a direct joint modelling approach, where neither of the problems arise, seems preferable.

Using a direct joint modelling approach in Section 4 we were able to learn about the effects that atmospheric variables have on activity levels of a soaring raptor, while for the blacktip reef shark we examined temporal and tidal inputs effects on its activity levels. The HMM produced similar temporal patterns of activity to a previous analysis of the blacktip reef shark data set using GAMMs (Papastamatiou *et al.*, 2015). Both analytical methods revealed crepuscular and/or nocturnal increases in activity with a tidal component, with the shark most active at the high tide or as tide was about to ebb. By incorporating swimming depth and temperature, it was also revealed that highest activity was seen when the shark was at the surface in waters of 28-29 °C. More importantly, the analysis showed that the shark was inactive when in very warm (>29 °C) shallow water or deeper water. These results agree with a previous hypothesis that sharks are 'hunting warm, and resting warmer' and use warmer water ($>$ 29 °C) to increase the rate of some physiological function such as digestion, and not for foraging (see Papastamatiou *et al.*, 2015). The HMM in this case allows us to explain the drivers of activity in the shark and move beyond just describing its movements, but rather explain 'why' it may be moving or selecting certain habitats. The HMM also provided a measure of the change in probability of the individual being in active states. Although there was a clear temporal pattern of activity, the HMM identified the shark as 30% more likely to be in an active state during the late evening hours. For the adult black eagle, the HMM provided a direct modelling approach to examine the effect of wind speed and temperature on its activity level. The results suggests that the black eagle spent more time in the relatively active state overall, and was more likely active in windier conditions. These results are in line with theoretical (Pennycuick, 1972) and empirical (Lanzone *et al.*, 2012) studies.

We have covered the basic HMM framework here, but the popularity of the HMM framework is due in part to its many extensions. In particular, there are two HMM extensions that have been proven useful in classification of human activities: the hidden semi-Markov model (HSMM) (Langrock & Zucchini, 2012) and the hierarchical hidden Markov model (HHMM) (Fine *et al.*, 1998). The HSMM models the time spent within a state by some probability distribution with support on the positive real integers, thereby allowing for more complex state dwell time distributions than can be provided by an HMM (namely only geometric distributions). For instance, an HMM may not model the time

spent in a resting behaviour adequately if the animal is known to rest for long periods of time. The HHMM provides the framework necessary to identify composite behaviours. For instance, lunge feeding in baleen whales is a composite behaviour made up of (1) initial increase in acceleration with (2) a positive pitch angle, as animals commonly approach prey schools from below, followed by (3) a rapid deceleration after the whale opens its mouth increasing its drag (Owen *et al.*, 2015). The HHMM models each composite behaviour as its own HMM, and models the transitions between composite behaviours, i.e. switches between HMMs.

6 Acknowledgements

YYW was funded by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (grant reference 25850138). TP was supported by a South African National Research Foundation Scarce Skills postdoctoral research fellowship, and TP and YPP received funding from the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) and their support is gratefully acknowledged. MASTS is funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions. TP and MM gratefully acknowledge the hardware, software, support and expertise contributed by Prof. Willem Bouten and his research group UvA-BiTS (University of Amsterdam Bird Tracking System). The authors are grateful to Orr Spiegel, the Associate Editor and an anonymous referee for their useful comments that substantially improved the manuscript.

7 Data Accessibility

Data deposited in the Dryad repository: http://datadryad.org/resource/doi:10.5061/dryad.6bm2c

References

Ákos, Z., Nagy, M., Leven, S. & Vicsek, T. (2010) Thermal soaring flight of birds and unmanned aerial vehicles. *Bioinspiration & Biomimetics*, **5**, 045003.

Altun, K., Barshan, B. & Tunçel, O. (2010) Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, **43**, 3605–3620.

Bao, L. & Intille, S.S. (2004) Activity recognition from user annotated acceleration data. *Pervasive computing*, pp 1–17. Springer, Berlin Heidelberg.

Bouten, W., Baaij, E.W., Shamoun-Baranes, J., Camphuysen, K.C.J. (2013) A flexible GPS tracking system for studying bird behaviour at multiple scales. *Journal of Ornithology*, **154**, 571-–580.

Broekhuis, F., Grünewälder, S., McNutt, J.W. & Macdonald, D.W. (2014) Optimal hunting conditions drive circalunar behavior of a diurnal carnivore. *Behavioral Ecology*, **25**, 1285–1275.

Brown, D., Kays, R., Wikelski, M., Wilson, R. & Klimley, A.P. (2013) Observing the unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry*, **1**, 1–16.

Campbell, H.A., Gao, L., Bidder, O.R., Hunter, J. & Franklin, C.E. (2014) Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species. *Journal of Experimental Biology*, **216**, 4501–4506.

Carroll, G., Slip, D., Jonsen, I. & Harcout, R. (2014) Supervised accelerometry analysis can identify prey capture by penguins at sea. *Journal of Experimental Biology*, **217**, 4295–4302.

Elliott, K.H., Le Vaillant, M., Kato, A., Speakman, J.R. & Ropert-Coudert, Y. (2013) Accelerometry predicts daily energy expenditure in a bird with high activity levels. *Biology letters*, **9**, 20120919.

Fine, S., Singer, Y. & Tishby N. (1998) The hierarchical hidden Markov model: Analysis and applications. *Machine Learning*, **32**, 41–62.

Gleiss, A.C., Wilson, R.P. & Shepard, E.L.C. (2011) Making overall dynamic body acceleration work: on the theory of acceleration as a proxy for energy expenditure. *Methods in Ecology and Evolution*, **2**, 23–33.

Gleiss, A.C., Wright, S., Liebsch, N., Wilson, R.P. & Norman, B. (2013) Contrasting diel patterns in vertical movement and locomotor activity of whale sharks at Ningaloo Reef. *Marine Biology*, **160**, 2981–2992.

Graf, P.M., Wilson, R.P., Qasem, L., Hackländer, K. & Rosell, F. (2015) The use of acceleration to code for animal behaviours; A case study in free-ranging Eurasian beavers *Castor fiber*. *PLoS ONE*, **10**, e0136751.

Hart, T., Mann, R., Coulson, T., Pettorelli, N. & Trathan, P.N. (2010) Behavioural switching in a central place forager: patterns of diving behaviour in the macaroni penguin (*Eudyptes chrysolophus*). *Marine Biology*, **157**, 1543–1553.

Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning*, New York, NY: Springer.

He, J., Li, H. & Tan, J. (2007) Real-time daily activity classification with wireless sensor networks using hidden Markov model. *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 3192–3795. IEEE.

Heerah, K., Hindell, M., Guinet, C. & Charrassin, J.B. (2014) A new method to quantify within dive foraging behaviour in marine predators. *PLoS ONE*, **9**, e99329.

Iwata, T., Sakamoto, K.Q., Edwards, E.W.J., Staniland, I.J., Trathan, P.N., Goto, Y., Sato, K., Naito, Y. & Takahashi, A. (2015) The influence of preceding dive cycles on the foraging decisions of Antarctic fur seals. *Biology Letters*, **11**, 20150227.

Katzner, T., Johnson, J.A., Evans, D.M., Garner, T.W.J., Gompper, M.E., Altwegg, R., Branch, T.A., Gordon, I.J. & Pettorelli, N. (2013) Challenges and opportunities for animal conservation from renewable energy development. *Animal Conservation*, **16**, 367–369.

Konishi, S. & Kitagawa, G. (2008) *Information Criteria and Statistical Modeling*, New York, NY: Springer.

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

Langrock, R., Kneib, T., Sohn, A. & DeRuiter, S.L. (2015) Nonparametric inference in hidden Markov models using P-splines. *Biometrics*, **71**, 520–528.

Langrock, R. & Zucchini, W. (2012) Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics and Data Analysis*, **55**, 715–724.

Lanzone, M.J., Miller, T.A., Turk, P., Brandes, D., Halverson, C., Maisonneuve, C., Tremblay, J., Cooper, J., O'Malley, K., Brooks, R.P. & Katzner, T. (2012), Flight responses by a migratory soaring raptor to changing meteorological conditions *Biology Letters*, **8**, 710–713.

Leenders, N.Y.J.M., Sherman, W.M. & Nagaraja, H.N. (2000) Comparisons of four methods of estimating physical activity in adult women. *Medicine & Science in Sports & Exercise*, **32**, 1320–1326.

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T.A., Watanabe Y.Y., Murtroyd, M., Papastamatiou, Y. (2016) Data from: Analysis of animal accelerometer data using hidden Markov models. *Methods in Ecology and Evolution* doi:10.5061/dryad.6bm2c

MacDonald, I.L. (2014) Numerical maximisation of likelihood: A neglected alternative to EM? *International Statistical Review*, **82**, 296–308.

Mannini, A. & Sabatini A.M. (2010) Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, **10**, 1154–1175.

Mannini, A. & Sabatini A.M. (2011) Accelerometry-based classification of human activities using Markov modeling. *Computational Intelligence and Neuroscience*, **2011**, 647858.

Martiskainen, P., Järvinen, M., Skön, J., Tiirikainen, J., Kolehmainen, M. & Mononen, J. (2009) Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, **119**, 32–38.

McKellar, A.E., Langrock, R., Walters, J.R. & Kesler, D.C. (2015) Using mixed hidden Markov models to examine behavioural states in a cooperatively breeding bird. *Behavioral Ecology*, **26**, 148–157.

Naito, Y. Bornemann, H., Takahashi, A, McIntyre, T. & Plötz, J. (2010) Fine-scale feeding behavior of Weddell seals revealed by a mandible accelerometer. *Polar Science*, **4**, 309–316.

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M. & Getz, W. (2012) Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *The Journal of Experimental Biology*, **215**, 986–996.

Owen, K., Dunlop, R.A., Monty, J.P., Chung, D, Noad, M.J., Donnelly, D., Golditzen, A.W. & Mackenzie, T. (2015) Detecting surface-feeding behavior by rorqual whales in accelerometer data. *Marine Mammal Science*, **32**, 327–348.

Papastamatiou, Y.P., Watanabe, Y.Y., Bradley, D., Dee, L.E., Lowe, C.G. & Caselle, J. (2015) Drivers of daily routines in an ectothermic marine predator: hunt warm, rest warmer? *PLoS ONE*, **10**, e0127807.

Patterson, T.A., Basson, M., Bravington, M.V. & Gunn, J.S. (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78**, 1113–1123.

Pennycuick, C. J. (2008) *Modelling the Flying Bird* London: Academic Press, Elsevier.

Pennycuick, C. J. (1972) Soaring behaviour and performance of some east african birds, observed from a motor-glider *Ibis*, **114**, 178-–218.

Phillips, J.S., Patterson, T.A., Leroy, B., Pilling, G.M. & Nicol, S.J. (2015), Objective classification of latent behavioral states in bio-logging data using multivariate-normal hidden Markov models. *Ecological Applications*, 25(5), 1244–1258.

Qasem, L., Cardew, A., Wilson, A., Griffiths, I., Halsey, L.G., Shepard, E.L.C., Gleiss, A.C. & Wilson, R. (2012) Tri-axial dynamic acceleration as a proxy for animal energy expenditure; Should we be summing values or calculating the vector? *PLoS ONE*, **7**, e31187.

Rabiner, L.R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE Proceedings*, **77**, 257–286.

Ravi, N., Dandekar, N., Mysore, P. & Littman, M.L. (2005) Activity recognition from accelerometer data. *American Association for Artifical Intelligence*, **5**, 1541–1546.

Sakamoto, K.Q., Sato K., Ishizuka, M., Watanuki, Y., Takahashi, A., Daunt, F. & Wanless, S. (2009) Can ethograms be automatically generated using body acceleration data from free-ranging birds? *PLoS One*, **4**, e5379.

Sato, K., Daunt, F., Watanuki, Y., Takahashi, A. & Wanless, S. (2008) A new method to quantify prey acquisition in diving seabirds using wing stroke frequency. *The Journal of Experimental Biology*, **211**, 58–65.

Sato, K., Mitani, Y., Cameron, M.F., Siniff, D.B. & Naito, Y. (2003) Factors affecting stroking patterns and body angle in diving Weddell seals under natural conditions. *The Journal of Experimental Biology*, **206**, 1461–1470.

Shepard, E.L.C., Wilson, R., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A., Morgan, D., Myers, A.E., Newman, C. & Macdonald, D. (2008) Identification of animal movement patterns using tri-axial accelerometry. *Endangered Species Research*, **10**, 47–60.

Simon, M., Johnson, M. & Madsen, P.T. (2012) Keeping momentum with a mouthful of water: behavior and kinematics of humpback whale lunge feeding. *The Journal of Experimental Biology*, **215**, 3786–3798.

Suzuki, I., Naito, Y., Folkow, L.P., Nobuyuki, M. & Blix, A.B. (2009) Validation of a device for accurate timing of feeding events in marine animals. *Polar Biology*, **32**, 667–671.

Watanabe, Y.Y., Lydersen, C., Fisk, A.T. & Kovacs, K.M. (2012), The slowest fish: swim speed and tail-beat frequency of Greenland sharks. *Journal of Experimental Marine Biology and Ecology*, **426**, 5–11.

Watanuki, Y., Niizuma, Y., Gabrielsen, G.W., Sato, K. & Naito, Y. (2003), Stroke and glide of wing-propelled divers: deep diving seabirds adjust surge frequency to buoyancy change with depth. *Proceedings of the Royal Society of London B*, **270**, 483–488.

Ward, J.A., Lukowicz, P., Troster, G. & Starner, T.E. (2006) Activity recognition of assembly tasks using body-worn microphones and accelerometers. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, **28**, 1553–1567.

Williams, H.J., Shepard, E.L.C., Duriez, O. & Lambertucci, S.A. (2015) Can accelerometry be used to distinguish between flight types in soaring birds? *Animal Biotelemetry*, **3**, 1–11.

Wilson, R.P., Shepard, E.L.C. & Liebsch, N. (2008) Prying into the intimate details of animal lives: use of a daily diary on animals. *Endangered Species Research*, **4**, 123–137.

Wilson, R.P., White, C.R., Quintana, F., Halsey, L.G., Liebsch, N, Martin, G.R. & Butler, P.J. (2006) Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant. *Journal of Animal Ecology*, **75**, 1081–1090.

Ydesen, K.S., Wisniewska, D.M., Hansen, J.D., Beedholm, K., Johnson, M. & Madsen, P.T. (2014) What a jerk: prey engulfment revealed by high-rate, super-cranial accelerometry on a harbour seal (*Phoca vitulina*). *The Journal of Experimental Biology*, **217**, 2239–2243.

Zucchini, W., MacDonald, I.L. & Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction using R*, 2nd Edition, Chapman & Hall/CRC, FL, Boca Raton.

8 Supporting Information

Supporting Information	Description
Appendix	Further mathematical details for HMMs. (Pseudo-)residual plots and model checking for both HMM applications presented in manuscript.
Comparing Supervised Learning Approaches	A comparison of four supervised learning approaches when there is varying levels of autocorrelation present in the data.
R code for HMMs	Documented R code presented for applications of HMMs in both a supervised and unsupervised learning approach.

9 Tables & Figures

Model	Log-likelihood	AIC	Δ AIC	BIC	Δ BIC
No covariates	2000.2	-3980.4	21.6	-3929.4	21.6
Temperature	2001.9	-3979.9	22.1	-3918.6	17.0
Wind speed	2010.4	-3996.9	5.1	-3935.6	**0**
Wind speed, Temperature	2011.6	-3995.2	6.8	-3923.7	11.9
Wind speed, Temperature, Wind speed * Temperature	2017.0	-4002.0	**0**	-3920.3	15.3

Table 1: Model fitting results for the Verreaux's eagle. Based on the AIC the model selected is the full model including wind speed, temperature and their interaction. Based on the BIC, the model selected includes only wind speed.

Model	Log-likelihood	AIC	Δ AIC	BIC	Δ BIC
No covariates	639299.2	-1278370	779	-1277178	692
Time	639558.1	-1278872	277.2	-1277645	225
Time, High	639657.6	-1279063	86.2	-1277819	51
Time, High, Flood	639695.2	-1279130	19	-1277869	1
Time, High, Flood, Ebb	639708.7	-1279149	**0**	-1277870	**0**

Table 2: Model fitting results for a blacktip reef shark. Based on the AIC and BIC, the model selected includes time of day and includes differences in activity levels based on all categories of tide levels.

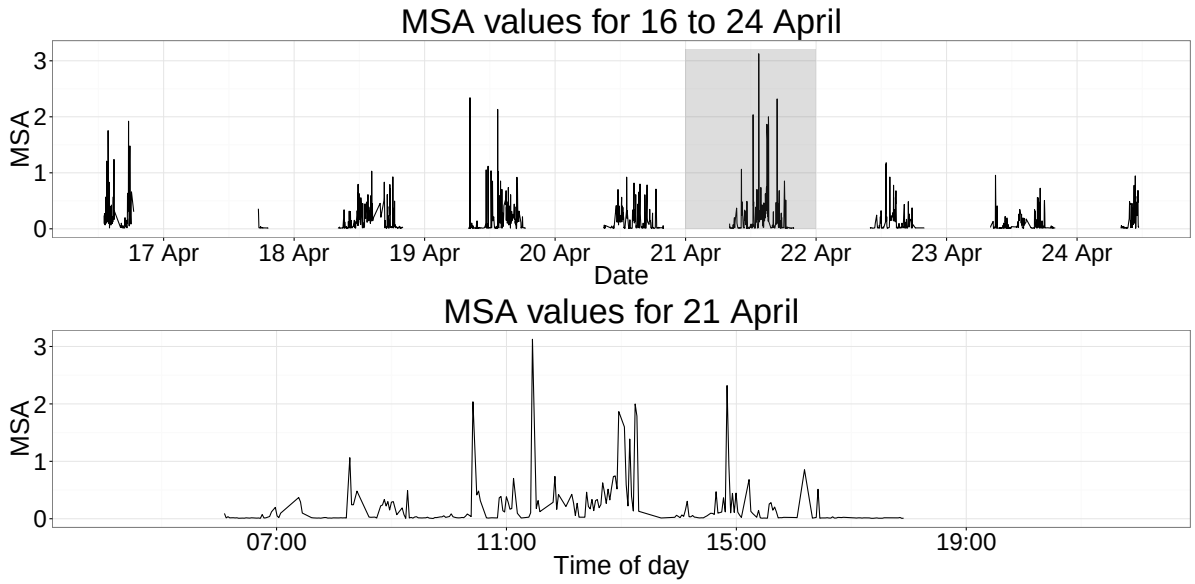

Figure 1: Minimum specific acceleration values derived from three-axis acceleration data from a Verreaux's eagle collected over 9 days, 16-24 July 2013 (top). Minimum specific acceleration values from the 21st of April, corresponding to the shaded area in the upper plot (bottom).

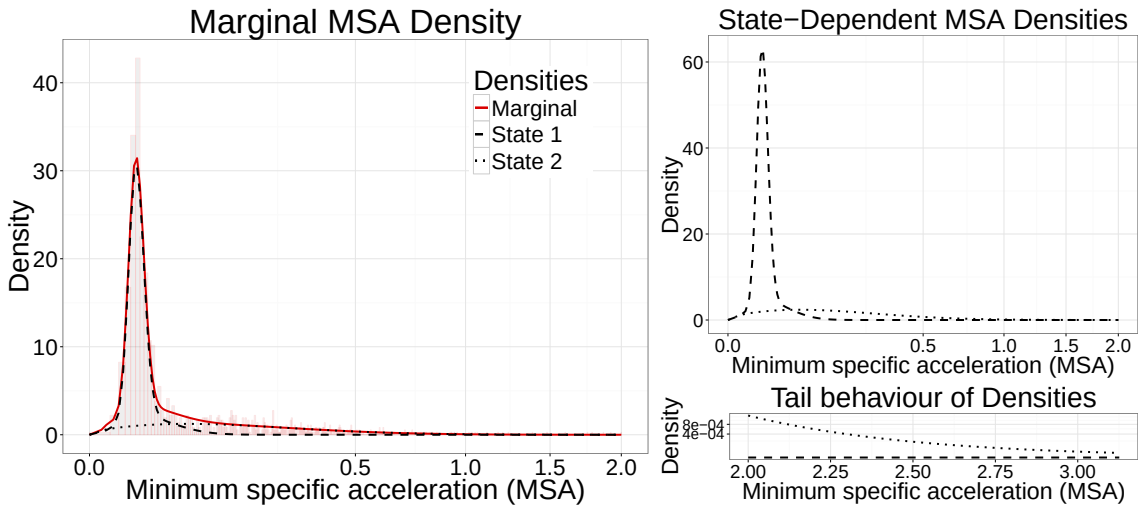

Figure 2: Histogram of minimum specific acceleration (MSA) from a Verreaux's eagle, truncated at MSA=2, with marginal density (the distribution of observations not conditional on process history) and state-dependent densities weighted according to the proportion of observations assigned to each state (left). Unweighted state-dependent densities (top right) and close-up of the tail behaviour of the densities (bottom right). A square root coordinate transformation for the x-axis was used in all plots and for the y-axis only for the tail behaviour plot.

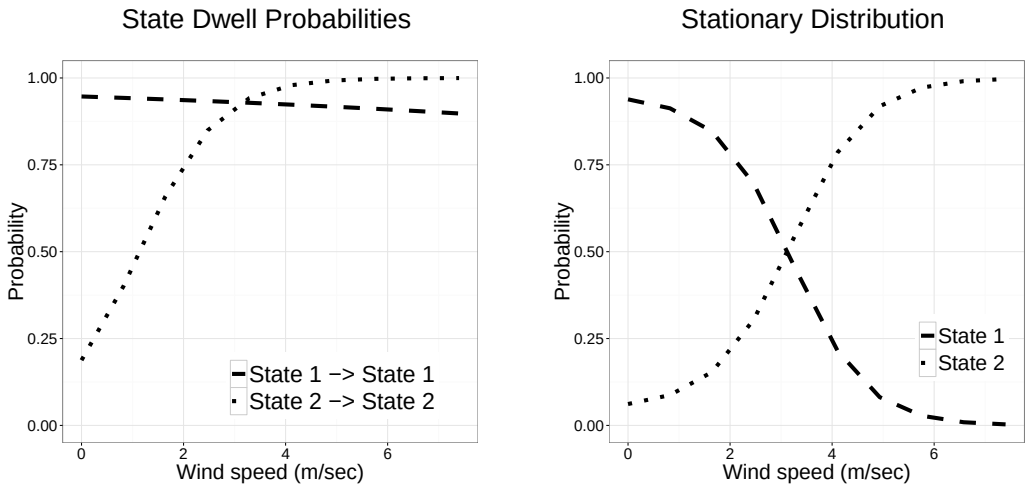

Figure 3: For the Verreaux's eagle example, estimated state-dwell probabilities (probability of remaining in a state) as a function of wind speed (left), and estimated equilibrium state probabilities (marginal probability of a state at a fixed value of the covariate) as a function of wind speed (right).

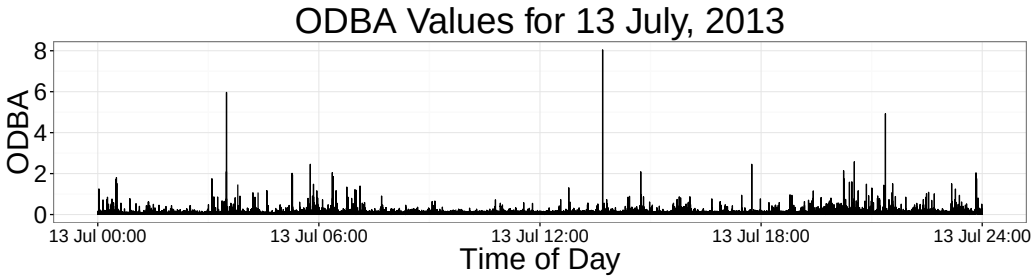

Figure 4: Overall dynamic body acceleration values from a blacktip reef shark, averaged over 1 second intervals, for 13 July, 2013.

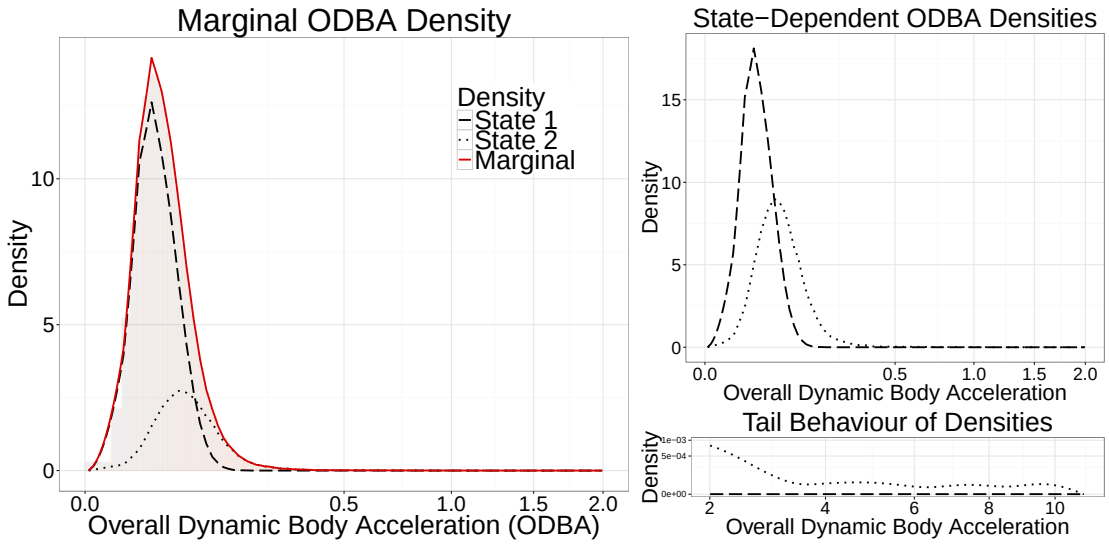

Figure 5: Histogram of ODBA from a blacktip reef shark, truncated at ODBA=2, with marginal density and state-dependent densities weighted according to the proportion of observations assigned to each state. (left). Unweighted state-dependent densities (top right) and close-up of the tail behaviour of the densities (bottom right). A square root coordinate transformation for the x-axis was used in all plots and for the y-axis only for the tail behaviour plot.

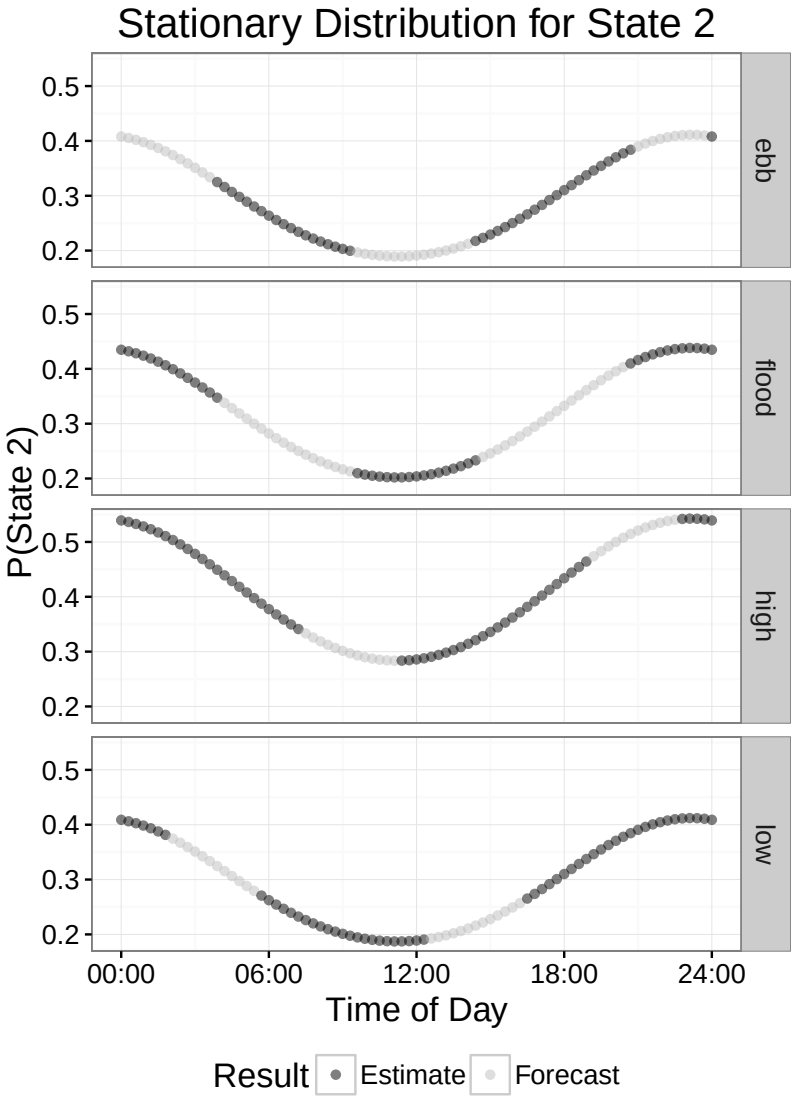

Figure 6: Implied stationary distribution for state 2, the more active state, by time of day and tide level for the blacktip reef shark example. For tide levels, we distinguish between model estimates, such that the corresponding tide level was observed at that time of day, and forecasts, where we did not observe the tide level at that time of day.

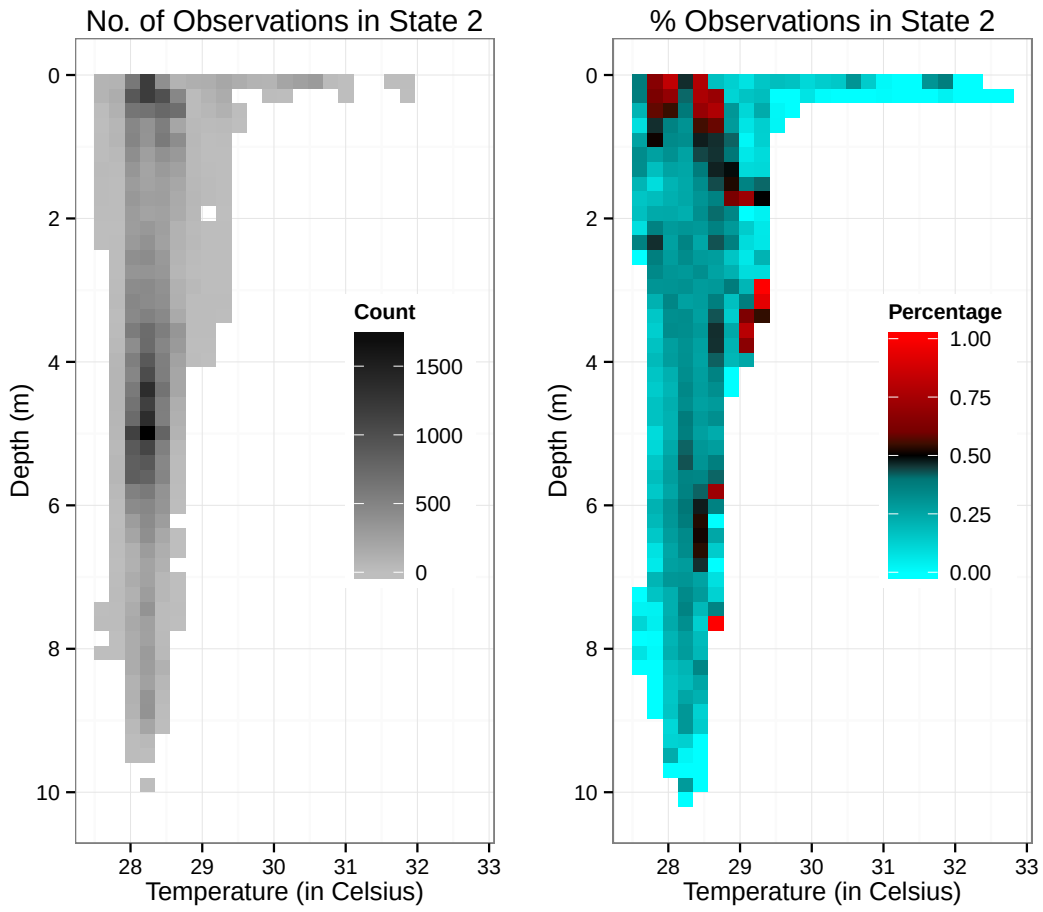

Figure 7: For the blacktip reef shark example, the number of observations in each grid cell that correspond to state 2. Zero counts appear in white. (left) Percentage of observations in each cell that correspond to state 2. (right)