Validating hidden Markov models for seabird behavioural inference

Rebecca A. Akeresola[1,2], Ruth King[1], Gail Robertson[2], Adam Butler[2], Víctor Elvira[1], Esther L. Jones[2], and Julie Black[3]

[1]School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh, UK

[2]Biomathematics & Statistics Scotland, Edinburgh, UK

[3]Joint Nature Conservation Committee, UK

September 28, 2025

CORRESPONDING AUTHOR

Rebecca A. Akeresola

School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh, UK

r.a.akeresola@sms.ed.ac.uk; orcid.org/0009-0005-0725-4630

Abstract

Understanding animal movement and behaviour can aid spatial planning and inform conservation management. However, it is difficult to directly observe behaviours in remote and hostile terrain such as the marine environment. Behaviours can be inferred from telemetry data using hidden Markov models (HMMs), but model predictions are not typically validated due to difficulty obtaining 'ground truth' behavioural information. We investigate the accuracy of HMM-inferred behaviours by considering a unique dataset provided by Joint Nature Conservation Committee. The data consist of simultaneous proxy movement tracks of the boat (defined as visual tracks as birds are followed by eye) and seabird behaviour obtained at the same time-frequency by observers on the boat. We use these data to assess whether (i) visual track is a good proxy for true bird locations in relation to HMM-inferred behaviours, and (ii) inferred behaviours from HMMs fitted to visual tracking data accurately represent true behaviours as identified by behavioural observations taken from the boat. We demonstrate that visual tracking data can be regarded as a good proxy for true movement data of birds in terms of similarity in inferred behaviours. Accuracy of HMMs ranging from 71% to 87% during chick-rearing and 54% to 70% during incubation was generally insensitive to model choice, even when AIC values varied substantially across different models. Finally, we show that for foraging, a state of primary interest for conservation purposes, identified missed foraging bouts lasted for only a few seconds. We conclude that HMMs fitted to tracking data can accurately identify important conservation-relevant behaviours, demonstrated using visual tracking data. Therefore, confidence in using HMMs for behavioural inference should increase even when validation data are unavailable. This has important implications for animal conservation, where the size and location of protected areas are often informed by behaviours identified using HMMs fitted to movement data.

1 INTRODUCTION

Seabirds are key indicators of marine environmental health (Parsons et al. 2008, Lascelles et al. 2012) but are the most threatened and anthropogenically pressured group of birds globally (Croxall et al. 2012). Threats, including invasive species at breeding colonies, climate change, over-fishing, and offshore renewable developments, have resulted in a global decline in seabird populations of 70% over the last five decades (Vulcano et al. 2021). In the UK, some species of seabirds (e.g. Northern fulmar (*Fulmarus glacialis*), little tern (*Sternula albifrons*), European shag (*Phalacrocorax aristotelis*), Arctic skua (*Stercorarius parasiticus*), and black-legged kittiwake (*Rissa tridactyla*)) have continued to decline (JNCC 2021). Of the 25 seabird species that regularly breed in the UK, 24 are listed as Red or Amber on the UK's Birds of Conservation Concern (Stanbury et al. 2021). Under the Habitats Directive (EC/92/43) and Birds Directive (EC/79/409), Special Protection Areas (SPA) are established to form the Natura 2000 network, which protects species and habitats (European-Commission et al. 2008). Although SPAs have historically been restricted to small areas focused on seabird breeding colonies, recent extensions and new classifications in the marine environment have expanded the SPA network across the UK (JNCC 2020). Seabirds are restricted to central-place foraging during the breeding season. Therefore, understanding at-sea behaviour, including characterising important foraging areas, is vital to ensure adequate protection measures are in place to prevent further population decline.

Seabird tracking studies, where individuals are tagged using biologging technology, are an effective way to understand space use and behaviour (Lascelles et al. 2012, Bennison et al. 2018, Wakefield et al. 2017). Technological advances have accelerated the availability of biologging information from devices such as Global Positioning System (GPS) transmitters, accelerometers, conductivity-temperature-depth (CTD) tags, and harmonic radar trackers (Cooke et al. 2004). Telemetry data provides information on animal locations at discrete intervals but does not provide direct information about the underlying behaviour of the tagged animals. To infer behavioural states such as foraging, flying, and resting from movement data, hidden Markov Models (HMMs) have been widely used (Morales et al. 2004, Patterson et al. 2009, Langrock et al. 2012, McKellar et al. 2015, McClintock 2021). HMMs are time series

3

models with observation and state processes where the latent (unobserved) states describe the underlying behaviour of the individual (Langrock et al. 2012). HMM-inferred behaviours can be used to inform conservation decision-making, for example, the size and location of protected areas.

One limitation of using HMM-inferred behaviours to inform conservation-relevant decision-making is the difficulty in validating models using ground truth data. Some studies have attempted to validate HMM-inferred behaviour from movement data, such as Joo et al. (2013), which validated the behaviour of fishing vessels using ground truth data recorded by onboard observers. Bennison et al. (2018) and Conners et al. (2021) also validated HMM-inferred behaviours of northern gannet (*Morus bassanus*) and albatross using behaviours from depth-recorder and sensors as ground truth data, respectively. However, depth recorders and sensors are also proxies for ground truth data with their own error structures. Overall, little research has focused on evaluating the performance of HMMs fitted to animal movement data through data validation because contemporaneous behavioural observations on tracked individuals can be challenging to collect, particularly in featureless environments, such as open ocean (Joo et al. 2013). To examine the performance of HMMs fitted to movement data, we consider a unique dataset provided by the Joint Nature Conservation Committee (JNCC) and obtained via the visual tracking of terns (*Sterna* spp.) using a rigid-hulled inflatable boat. A visual tracking method developed by Perrow et al. (2011) was conducted at several tern breeding colonies across the UK during chick-rearing and incubation in different years (Wilson et al. 2014). Proxy movement data, corresponding to the GPS location of the boat, and the observed behavioural data of the terns directly recorded by the observers on the boat were collected.

First-hand behavioural data of seabirds such as that collected by Wilson et al. (2014) is generally not feasible to collect directly alongside GPS tracking location data. We consider terns as a case study to examine the performance of HMMs for behavioural inference. To the best of our knowledge, this is the first study to validate HMM-inferred behaviour from movement data using observed behavioural data of seabirds. Our study aims to leverage the rare opportunity provided by the unique JNCC dataset to (i) examine whether boat locational

4

data are an adequate proxy of tern movement and (ii) validate inferred behaviours of seabirds from HMMs using observed seabird behavioural data.

# 2 MATERIALS AND METHODS

2.1 Study species and sites

This study investigates the movement behaviour of four tern (*Sterna* spp.) species: Arctic (*Sterna paradisaea*), common (*S.hirundo*), Sandwich (*S.sandvicensis*), and roseate (*S.dougallii*). Arctic terns tend to breed in coastal areas in the north and west of the UK, with 80% occurring in Shetland, Orkney, and the Outer Hebrides. Common terns have a widespread coastal distribution around the UK and also nest in small colonies inland along rivers and islets. Sandwich terns congregate in several large colonies, and most roseate terns breed on Rockabill, Ireland, with some pairs occasionally breeding in south-east Scotland, Norfolk, and Hampshire (Wilson et al. 2014). Study sites comprised of 9 breeding colonies across the UK (Figure 1): Blue Circle (54°49′N, 5°46′W) and Cockle Island (54°40′N, 5°37′W) in Northern Ireland; Cemlyn Bay (53°24′ N, 4°30′ W) in North Wales; Glas-Eileanan Island (56°49′N, 5°71′W), Forvie (57°18′N, 1°58′W), Isle of May (56°10′N, 2°32′W), Leith (55°96′N, 3°16′W) and South Shian (56°46′N, 5°36′W) in Scotland; and Coquet Island (55°20′N, 1°32′W) in England.

Terns are ground-nesting colonial breeders, raising one brood each breeding season (May-June) and laying a clutch of one to three eggs. While breeding adult terns are central-place foragers throughout the breeding season, they are particularly restricted during chick-rearing when they must return regularly to provision their chicks, and adults spend up to 80% of their time foraging (Thaxter et al. 2012). Sandwich terns are specialist predators that can exploit clupeids and sandeels from deeper water, potentially due to their wider foraging range. Likewise, roseate terns are specialists who also forage by plunge diving to depth, catching prey items of predominately sandeels, herring, and sprat. Common terns are generalist predators and prey items include invertebrates, clupeids, sandeels, and gadoids. Arctic terns forage using several techniques but are heavily dependent on sandeel and changes in prey availability can

¹²⁶ affect their breeding success (Eglington & Perrow 2014).
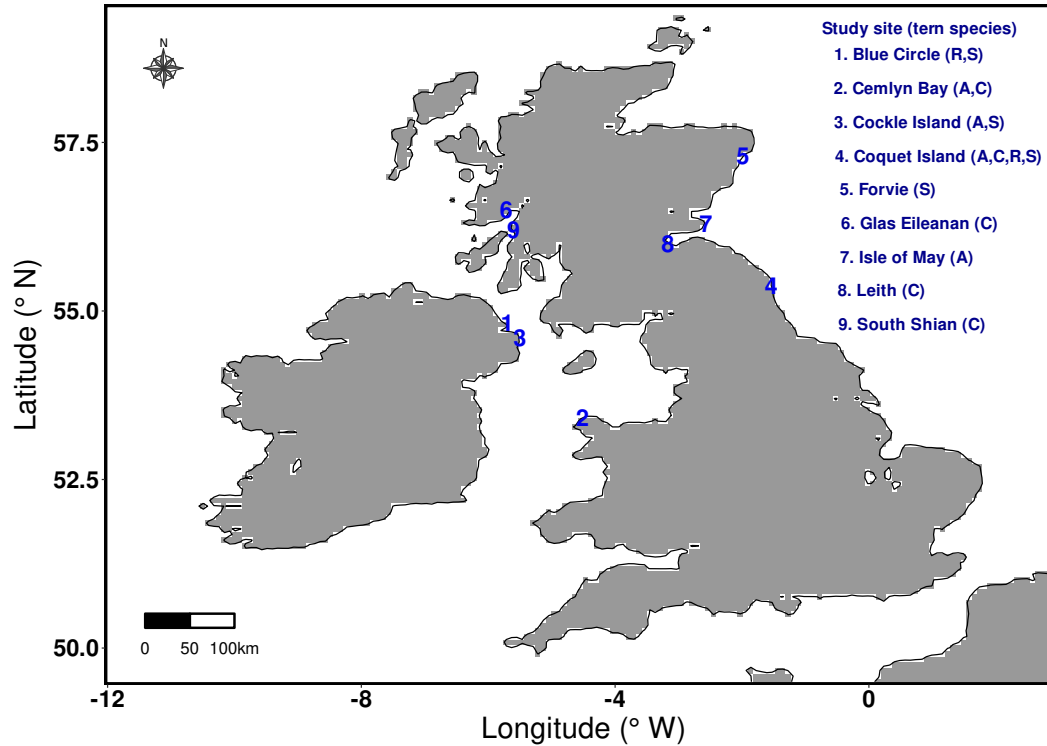
FIGURE 1 Study sites consisting of 9 tern breeding colonies in the United Kingdom. A-Arctic, C-Common, R-Roseate, and S-Sandwich tern.

2.2 Visual tracking data

¹²⁸ Visual tracking data were collected using a technique developed by Perrow et al. (2011) and
¹²⁹ detailed in Wilson et al. (2014). We summarise the protocol as follows: The visual tracking
¹³⁰ of terns was conducted during chick-rearing (June and July) and incubation (early May to
¹³¹ mid June) between 2009 and 2011. Rigid hull inflatable boats used for the visual tracking
¹³² were operated by different skippers across the study sites. The boats were kept c.50-200m
¹³³ from terns whilst an individual was tracked to avoid disturbing the birds and affecting their
¹³⁴ behaviour. Longitude and latitude of the boats were recorded using an onboard GPS device
¹³⁵ set to a 1-second sampling frequency. Individuals were tracked on return foraging trips from
¹³⁶ their breeding colony. One observer maintained constant sight of the tracked individual, while
¹³⁷ another recorded behavioural information.

¹³⁸ An ethogram of continuous flight behaviours and instantaneous foraging events was provided

to each observer, and the timing of each behaviour was recorded (Wilson et al. 2014). Flight behaviours were categorised as active search, transit search, and direct flight. Direct flight was defined as a clear and consistent direction with fast flight usually returning to the colony with food. An active search was defined as an erratic flight course actively searching for food, which may include instances of diving and surface feeding. It is hypothesised that for a direct flight, terns have a fixed location in view and fly in a clear and consistent direction, whereas for transit search, they may change direction but not erratically to search for food (Wilson et al. 2014). As a result, direct flight and transit search were defined as observed not-foraging behaviour while an active search was defined as observed foraging behaviour. These behavioural data are used as the validation data in the study.

The location of each observed behaviour was calculated from the boat's GPS track log. Unique IDs were assigned to the data of individual terns tracked in each colony. In 2009 and 2011, tracking only took place during chick-rearing. In 2010, tracking was conducted during chick-rearing and incubation periods. Figure 2 provides an example of visual tracks for the two breeding seasons. The data combined both complete and incomplete tracks of terns. The track of terns was considered complete if individual terns were tracked leaving and returning to the colony. Incomplete tracks were terns that could not be successfully followed back to the colony.

(a)   (b)
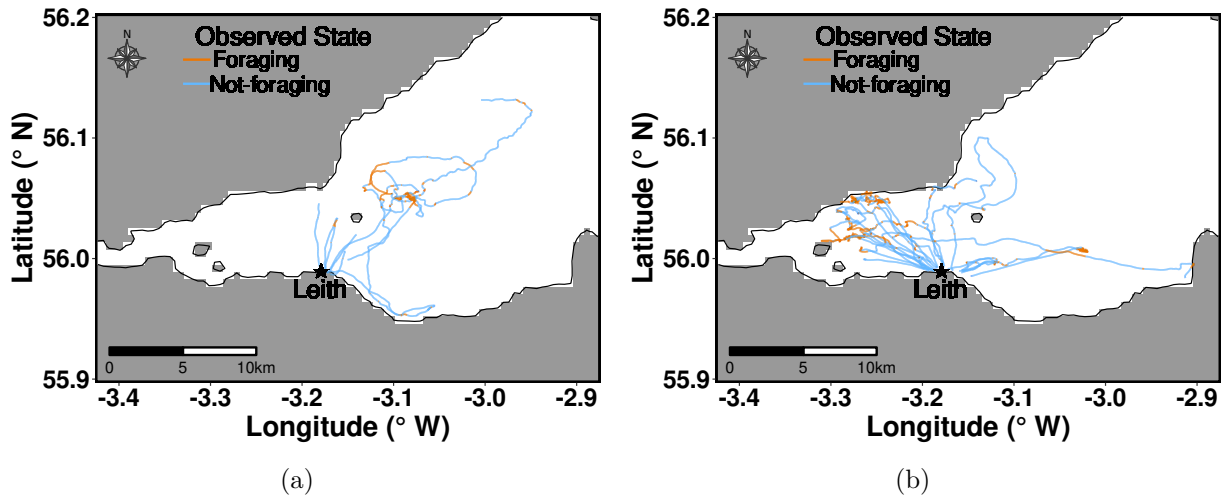
FIGURE 2 Visual tracks of common terns coloured with observed behavioural states during (a) incubation and (b) chick-rearing period from Leith dock, 2010.

157 Reasons for incomplete tracks could be individuals flying faster than the boat could follow,

158 flying over a physical obstruction that prevented the boat from following, observers confusing

159 the tracked individuals with other terns, or insufficient fuel in the boat (Perrow et al. 2011).

160 Visual tracks for which (i) a single observed behaviour was recorded throughout the tracking

161 trip and (ii) total tracking time that did not exceed 1 min are omitted (see Table S1 for a

162 summary of visual tracks used, Supporting Information). GPS coordinates of the boat are

163 subsequently converted into step length (km) and turning angle (radians). These calculated

164 metrics potentially provide information about tern behaviour. For example, foraging

165 behavioural activities are typically characterized by slow and tortuous flight, indicating

166 smaller step lengths and low directional persistence in turnings. In contrast, not-foraging

167 behavioural activities are generally characterized by longer step lengths and high directional

168 persistence in turnings (Morales et al. 2004).

169 2.3 Visual tracks as a proxy for tern tracks

170 Given that the boat followed at a distance c.50-200m from the tracked terns, we investigate

171 how well boat tracks replicate the movement of tracked individuals using additional

172 information on the animal's recorded position in relation to the boat. For a subset of tracks

173 recorded at the Coquet Island colony during the chick-rearing in 2009, additional data were

174 also collected corresponding to the distance and bearing of the tern from the boat, thus

175 permitting the reconstruction of the (approximate) longitude and latitude location of the

176 tern.

177 Mathematically, let Lon_{boat} and Lat_{boat} denote the boat's longitude and latitude position and

178 the bearing and distance of the boat to the tern be indicated by "bearing" and "distance",

179 respectively. Then the corresponding tern longitude and latitude (Lon_{tern}, Lat_{tern}) are given

180 by:

$$\text{Lat}_{\text{tern}} \approx \arcsin\Big(\sin\left(\text{Lat}_{\text{boat}}\right) \times \cos\left(\text{distance}/R\right) + \cos\left(\text{Lat}_{\text{boat}}\right) \times \sin\left(\text{distance}/R\right) \times \cos\left(\text{bearing}\right)\Big),$$

$$\text{Lon}_{\text{tern}} \approx \text{Lon}_{\text{boat}} + \text{atan2}(y, x),$$

where

$$R = 6371\text{km(radius of the earth)},$$

$$y = \sin\left(\text{bearing}\right) \times \sin\left(\text{distance}/R\right) \times \cos\left(\text{Lat}_{\text{boat}}\right), \quad \text{and}$$

$$x = \cos\left(\text{distance}/R\right) - \sin\left(\text{Lat}_{\text{boat}}\right) \times \sin\left(\text{Lat}_{\text{tern}}\right).$$

We compare (i) boat tracks and approximate tern tracks and (ii) the distribution of step length and angles corresponding to the boats and approximate tern tracks to determine whether the former can be used as an approximation for the movement of individual terns. We then model the boat and approximated tern tracking data using HMMs to account for the different movement patterns dependent on the (unknown) underlying behavioural states. We then extract the inferred behavioural states from models fitted to both datasets and create a confusion matrix to assess differences and similarities in inferred states.

2.4 Hidden Markov model (HMM)

A HMM (Figure 3) is a time series model with an observed component, X_t, driven by an underlying latent component known as the state process, S_t. The latter, S_t, takes a value on a finite set of N possible values and is assumed to be a first-order Markov chain with the state transition probability $\gamma_{ij} = P(S_t = j | S_{t-1} = i)$. The observed component, X_t, which can be univariate or multivariate, is assumed to be regularly spaced in time, t, with the associated observation process distribution $f(X_t | X_{t-1}, \ldots, X_1, S_t, \ldots, S_1) = f(X_t | S_t)$ at any given time t.
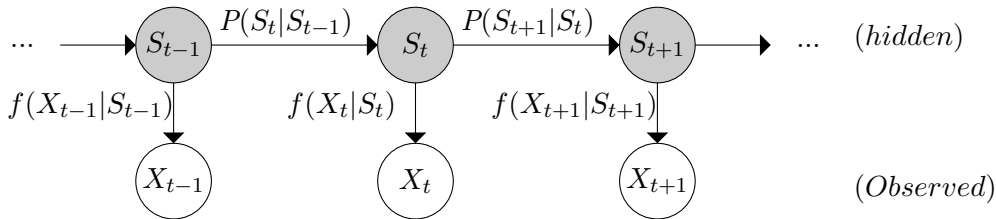
FIGURE 3 Graphical representation of a HMM where S_t and X_t denotes the state and observed process.

HMMs are suitable for fitting to the visual tracking data since observations are collected at a

regularly spaced interval, and for each time t, we specify $N = 2$ discrete states corresponding to foraging ($S_t = 1$) and not-foraging ($S_t = 2$).

Each observed data point, X_t, is bi-dimensional, consisting of the step length (km), r_t, and the turning angle (radians), ψ_t. At each time t, the distribution of X_t is conditional on the current hidden state, S_t, such that

$$f_j(X_t) = \Big(f(r_t \mid s_t = j), f(\psi_t \mid s_t = j) \Big) \quad \text{for} \quad j = 1, \cdots, N, \tag{1}$$

where r_t is modelled from a gamma distribution with parameters mean, μ, and standard deviation, σ i.e., $r_t | (S_t = j) \sim Gamma(\mu_j, \sigma_j)$, and ψ_t is modelled from a von Mises distribution with parameters mean, ρ, and concentration, κ i.e., $\psi_t | (S_t = j) \sim von-Mises(\rho_j, \kappa_j)$. We assume the distributions are independent for each time t, conditional on the underlying state S_t.

The corresponding likelihood of a HMM is a function of the following parameters: (i) $\boldsymbol{\delta}$: $(1 \times N)$ vector of initial state distribution given as $\boldsymbol{\delta} = (P(S_1 = 1), \ldots, P(S_1 = N))$ (ii) $\boldsymbol{\Gamma}$: $(N \times N)$ matrix of transition probabilities given as

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{pmatrix}, \tag{2}$$

(iii) $\mathbf{P}(X_t)$: $(N \times N)$ diagonal matrix corresponding to the observation process given as

$$\mathbf{P}(X_t = x_t) = \begin{pmatrix} f(X_t = x_t | S_t = 1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f(X_t = x_t | S_t = N) \end{pmatrix}. \tag{3}$$

The general likelihood function of a HMM is then given by

$$\mathcal{L} = P(X_t = x_t \mid \boldsymbol{\theta}) = \boldsymbol{\delta} \boldsymbol{\Gamma} \mathbf{P}(x_1) \ldots \boldsymbol{\Gamma} \mathbf{P}(x_T) \mathbf{1}, \tag{4}$$

where **1** is a column vector of length N with all entries equal to 1. We estimate these parameters, $\boldsymbol{\theta} = \{\boldsymbol{\delta}, \boldsymbol{\gamma}, \mathbf{P}(x_t)\}$, via maximum likelihood estimation and obtain the most likely state sequence using the Viterbi algorithm (Zucchini et al. 2016). We used the R package momentuHMM (McClintock & Michelot 2018) for fitting HMMs to the boat tracks since the package is widely used by scientists studying animal movement. We followed the guidance outlined in Michelot et al. (2016) to specify initial starting values for the model parameters.

2.5 HMMs specification and selection

Boat GPS locations were recorded at 1s intervals and do not have missing data. The completeness of the data means it is possible to use the recorded positions directly without the need to standardize the recording frequency by interpolating in time and space. Seabirds have been shown to vary their behaviour and area use at different breeding stages, travelling further from the colony to rich foraging grounds during incubation and remaining closer to the colony to feed chicks during chick-rearing (Robertson et al. 2014). As behaviour is expected to differ between the two periods, we expect model parameters to differ. We consider different models by varying model parameters for each tern species at each colony during the two periods, summarised in Table 1.

TABLE 1   HMMs (Models 0 - 6) fitted to the boat tracking data across study sites during incubation and chick-rearing. Covariate = Euclidean distance of the boat to the study site.

Models	Pooling effect		Covariate effect	
	State process	Observed process	State process	Observed process
0	✓	✓	✗	✗
1	✗	✓	✗	✗
2	✓	✗	✗	✗
3	✗	✗	✗	✗
4	✗	✓	✓	✗
5	✓	✗	✗	✓
6	✗	✗	✓	✓

Model 0, the base model, specifies that the state and observed processes are pooled across the individual visual tracks so that the model parameters are assumed to be the same for all individuals. Model 1 assumes a unique transition probability matrix parameter, Γ, for each

11

individual by removing the pooling effect on the state process. Model 2 assumes unique step

length parameters for each individual track by removing the pooling effect on the observed

process across tracks. The pooling effect on both state and observed process is not included

in Model 3.

The Euclidean distance of the boat to the colony was included as an environmental covariate

on the state process in Model 4, the observed process in Model 5, and both processes in Model

6. The parameters associated with the observed and state process are pooled across individual

tracks for Models 4 and 5, respectively. We include the covariate as a proxy for the energetic

cost of travelling to a particular location from the breeding colony since this cost constrains

the at-sea distribution of central-place foragers such as breeding seabirds (Wilson et al. 2014).

We assume a multinomial logistic regression for the models with a covariate included (Models

4, 5, and 6). Let c denote the covariate for the 2-state HMM, we set

$$\gamma_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{N} \exp(\eta_{ik})} \quad \text{for } i, j = 1, 2, \tag{5}$$

where

$$\begin{cases} \eta_{12} = \beta_0^{(12)} + \beta_1^{(12)} c_{12}, \\ \eta_{21} = \beta_0^{(21)} + \beta_1^{(21)} c_{21}, \\ \eta_{11} = \eta_{22} = 0, \end{cases} \tag{6}$$

and β_0, β_1 corresponds to the intercept and the regression parameter of the covariate, respectively.

Model selection was performed using the Akaike information criterion (AIC) (Burnham & Anderson 2002). The AIC value is expressed as

$$\text{AIC} = -2\ln(\widehat{\mathcal{L}}) + 2\text{p}, \tag{7}$$

where $\widehat{\mathcal{L}}$ is the likelihood evaluated at the MLE of the model parameters and p is the number of model parameters. We define ΔAIC_i as

$$\Delta \mathrm{AIC}_i = \mathrm{AIC}_i - \mathrm{AIC}_{min} \quad \text{for} \quad i = 0, \dots, 6, \tag{8}$$

such that $\Delta \mathrm{AIC}_i = 0$ for the model deemed optimal.

2.6 Model validation

The validation data consist of the observed behaviours of visually tracked terns. The inferred behavioural states from HMMs and validation data are assumed to be binary classifications: foraging and not-foraging. Common evaluation metrics for binary classification tasks include confusion matrix, F1-score, area under a ROC curve, and logarithmic loss (Hossin & Sulaiman 2015). We use the F1-score metric to validate behavioural states of visually tracked terns inferred from HMMs since the data are unbalanced; that is, observed behavioral state distribution is uneven. In particular, we identified an unbalanced classification for some breeding colonies such as Cemlyn, Isle of May, and Leith. The foraging behavioural state is of more interest as this helps to identify tern foraging areas. Therefore, a false negative, which fails to identify a foraging behaviour, is of higher importance in this context than a false positive. To calculate the F1-score metric, we obtain the (i) positive predictive value (PPV), which is the proportion of correct positives identified from all the predicted positives calculated as

$$\mathrm{PPV} = \frac{\text{number of true positive}}{\text{number of true positive} + \text{number of false positive}} \tag{9}$$

and (ii) true positive rate (TPR), which is the proportion of the positives that are predicted correctly and expressed as

$$\mathrm{TPR} = \frac{\text{number of true positive}}{\text{number of true positive} + \text{number of false negative}} \tag{10}$$

Using Equations (9) and (10), the F1-score is calculated as

$$\text{F1-score} = 2\left(\frac{\mathrm{PPV} * \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}}\right) \tag{11}$$

We also report the negative predictive value (NPV), which is the percentage of correct not-foraging behavioural states of all the decoded not-foraging states expressed as

$$\text{NPV} = \frac{\text{number of true negative}}{\text{number of true negative} + \text{number of false negative}} \tag{12}$$

Although the F1-score is a good validation metric, it does not account for how close the decoded behavioural state is to the observed behavioural state. However, the logarithmic loss metric, which is based on probability, does account for the uncertainty in the predicted classification (Hossin & Sulaiman 2015). Thus, we also consider the logarithmic loss for the fitted HMMs to account for the uncertainty of the decoded behavioural state. We use the observed behavioural states at each point, y_i, and the predicted probabilities of decoded behavioural state, q_i, to calculate the logarithmic loss metric as

$$\text{Log loss}(y, q) = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(q_i) + (1 - y_i) \log(1 - q_i) \right] \tag{13}$$

where n is the number of observations. The fitted model with the lowest log-loss value is deemed optimal for this criteria, and we report the F1-score, PPV, and TPR corresponding to optimal HMMs.

In addition to the validation metrics, we obtain the total number of foraging events identified within each observed behavioural data for visual tracking trips conducted across breeding colonies for each tern species. We define a foraging event as a bout within which only foraging behavioural states are recorded in the observed behavioural data of the individual tracked tern species. We then calculate the proportion of observed foraging events where optimal HMMs correctly infer (i) less than 25% (0% exclusive), (ii) 25 − 49%, (iii) 50 − 74%, and (iv) at least 75% of foraging behavioural states. Also, we obtain the proportion of observed foraging events completely missed from the foraging behavioural states inferred from optimal HMMs (i.e., observed foraging events where the model infers foraging at 0% of the time points).

14

3 RESULTS

## 3.1 Assessment of visual tracking data as a proxy for tern movement data

Reported results are based on visual tracking conducted at the Coquet Island colony during the chick-rearing period in 2009. We compare the boat locations to the associated inferred movement track of nine terns and distributions of the derived step lengths and turning angles. Typical foraging movement patterns generated by the boat and inferred tern tracks are provided in Figure 4 (and Figures S1-S4, Supporting Information). There are strong similarities between the locations (as would be expected given the boats were following the birds) and step length distributions. However, there appear to be more substantial differences with the turning angle distributions (lower panel of columns 2 and 3 in the figures). The latter difference can be explained by the bird making quicker turns compared to the boat, which has smoother turning movements.

FIGURE 4   Approximate roseate tern tracks and boat tracks from Coquet Island, 2009 (column 1). Histogram showing the distribution of step length (km) and turning angles (radians) from boat tracks (column 2) and from approximate roseate tern tracks (column 3).

302 We fitted HMMs to both boat and inferred tern location data. We observed little difference

303 in the inferred behavioural states when using boat location to approximate the location of

304 the tern. The confusion matrix metrics in Figure 5 indicate that the proportions of true

305 positives and true negatives when comparing behaviours derived from fitting HMMs to boat

306 and inferred tern locations against each other are higher than those of false negatives and

307 false positives.

16
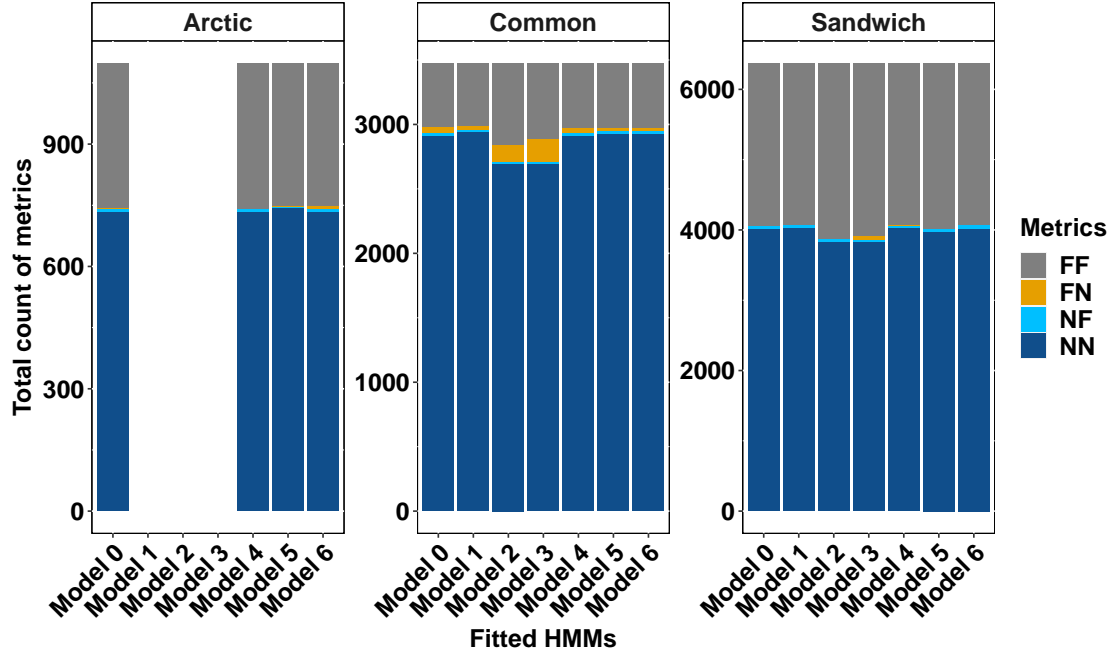
FIGURE 5 Confusion matrix metrics of behavioural states inferred from HMMs fitted to the boat and approximate location data of 1 Arctic, 2 common, and 5 Sandwich terns from Coquet colony during chick-rearing, 2009. F = Foraging, N = Not-Foraging. FF = true positive, FN = false negative, NF = false positive, and NN = true negative. Models 1, 2, and 3 require a minimum of 2 terns; hence, there is no bar for Arctic tern.

3.2 Validating HMM-inferred behavioural states

Reported results are based on HMMs deemed optimal (i.e., HMMs with the lowest log-loss value). Tables 2 and 3 present the summarised results of 2-state HMMs fitted to the visual tracking data during incubation and chick-rearing (see Tables S2-S4 for additional results, Supporting Information). Correctly decoded foraging states relative to total decoded foraging states ranged from 65% to 98% during chick-rearing.

17

TABLE 2 Validation results of 2-state HMMs fitted to visual tracking data of terns during incubation. Γ = transition probability matrix, covariate = Euclidean distance of boat to colony.

Incubation		HMM deemed optimal (i.e. based on lowest log-loss value)		Validation metrics (%)		
Colony	Species	Model	Model Description	PPV	TPR	F1-score
Leith	Common	0	complete pool	61.00	60.29	60.64
Blue Circle	Roseate	4	covariate on Γ	82.90	60.62	70.03
Cockle	Arctic	5	covariate on step	60.11	68.63	64.08
Cockle	Sandwich	6	covariate on Γ and step	59.54	49.65	54.15
Isle of May	Arctic			63.88	32.20	42.82

TABLE 3 Validation results of 2-state HMMs fitted to visual tracking data of terns during chick-rearing. Γ = transition probability matrix, covariate = Euclidean distance of boat to colony.

| Chick-rearing		HMM deemed optimal (i.e. based on lowest log-loss value)		Validation metrics (%)		
Colony	Species	Model	Model Description	PPV	TPR	F1-score
Coquet	Common	0	complete pool	88.44	79.19	83.56
Glas Eileanan				84.21	73.49	78.48
Coquet	Arctic	1	no pool on Γ	66.84	61.75	64.19
Isle of May				86.77	70.65	77.88
Blue Circle	Sandwich	2	no pool on step	80.84	78.36	79.58
Leith	Common			74.05	70.19	72.07
Cockle	Sandwich	3	no pool on Γ and step	84.85	91.11	87.87
Coquet				86.90	74.77	80.38
Forvie				65.15	79.90	71.77
Cemlyn	Arctic	4	covariate on Γ	98.91	58.92	73.85
Coquet	Roseate			68.66	86.06	76.38
South Shian	Common			70.51	90.64	79.32
Cemlyn	Common	6	covariate on Γ and step	71.93	81.21	76.29

314 We note that correct decoded foraging states relative to total observed foraging states ranged

315 from 70% to 91% except for Arctic terns from Cemlyn and Coquet study sites with 58% and

316 61%, respectively. Overall, the performance of HMMs in correctly inferring behavioural states

317 during chick-rearing is at least 71% across study sites except for Arctic terns in Coquet, with

a percentage of 64%. Validation of HMM results for incubation data shows a low performance compared to models fitted to chick-rearing data in inferring behavioural states. For example, we recorded at least 70% for only one roseate tern visually tracked at the Blue Circle colony during incubation. The overall low performance during this breeding season may be due to the small sample sizes.

Examining the corresponding observed behavioural data for each movement track of the boat, we identified and defined a foraging bout within each track where observed foraging behaviours were recorded as a foraging event. Optimal models correctly identify at least 50% of foraging behaviour within each observed foraging event, most times during chick-rearing (Figure 6). The reverse is, however, the case during incubation (Figure 7). The number of observed foraging events completely missed across study sites (i.e., observed foraging events where the model infers foraging at 0% of the time points) sums to 65, with an average time of 21 secs (Figure 8).
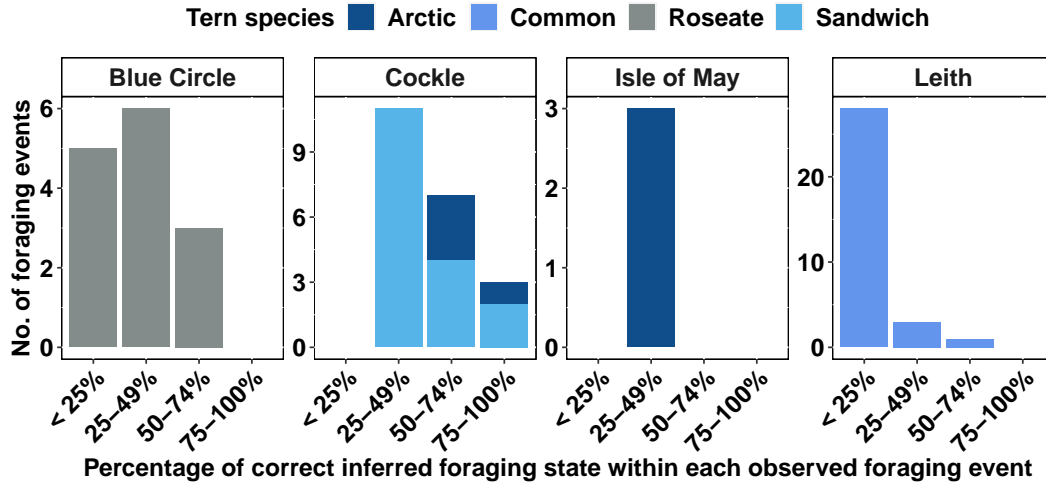
FIGURE 6 Proportion of correctly inferred foraging states within each observed foraging event across study sites during incubation.

19
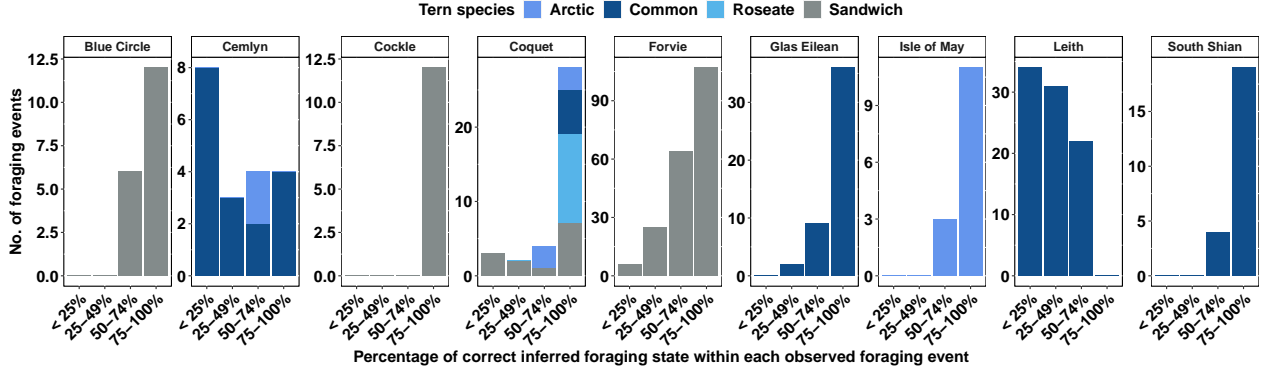
FIGURE 7 Proportion of correctly inferred foraging states within each observed foraging event across the study sites during chick-rearing.
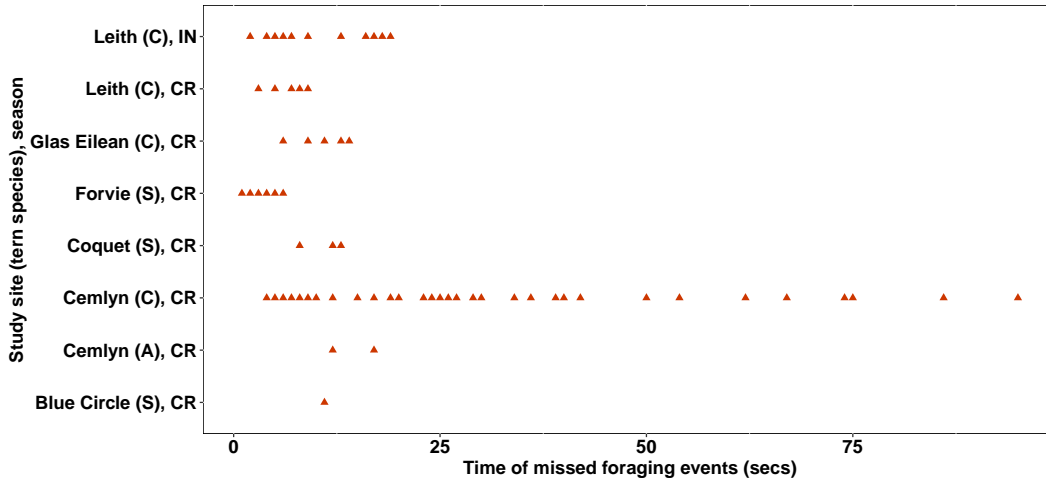
FIGURE 8 Observed foraging events completely missed from inferred foraging events across the study sites during chick-rearing (CR) and incubation (IN). A-Arctic, C-Common, and S-Sandwich tern.

The visual tracks coloured with behavioural states (see, for example, Figure 9) reveals similarity in the inferred and observed behavioural states across time points within visual tracking trips conducted across breeding colonies. Figure 10 provides histograms of the step length and turning angle overlaid with the density curves of the inferred behavioural states for a given track (see Figures S5-S7 for additional tracks, Supporting Information). The inferred states assigned to foraging show shorter step lengths and lower directional persistence in turnings than the not-foraging states, which exhibit larger step lengths and high directional persistence in turnings.

20
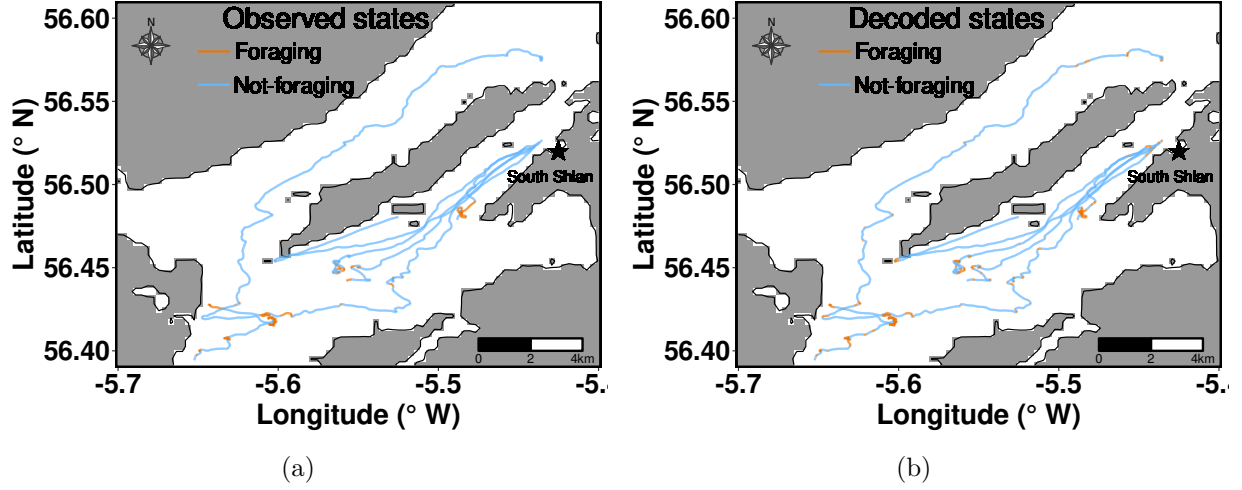
FIGURE 9 Visual tracks of 6 common terns coloured with (a) observed and (b) decoded behavioural states from South Shian colony during the chick-rearing period, 2011.
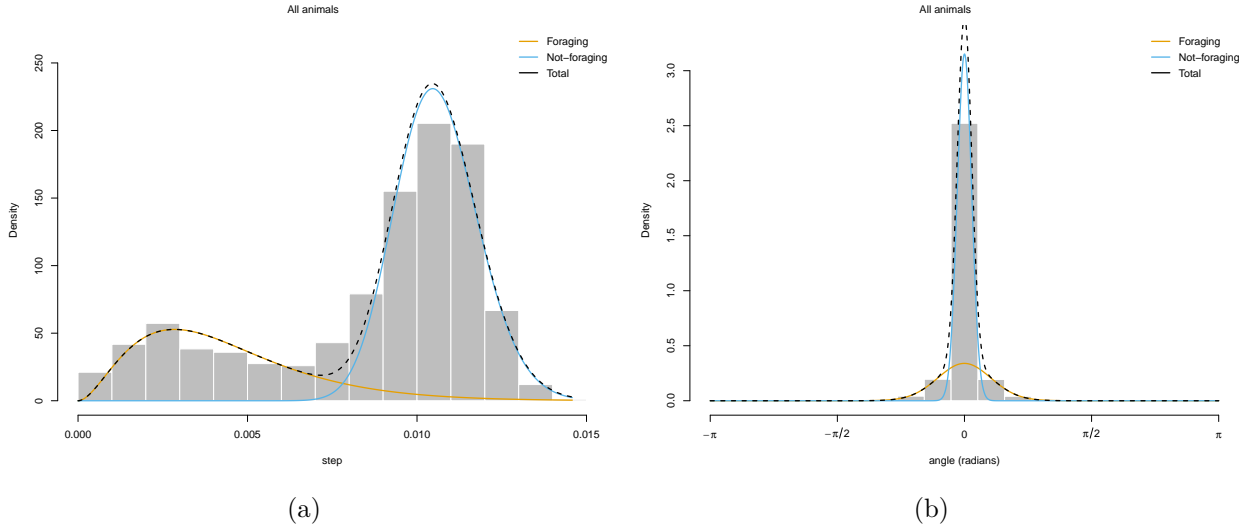
FIGURE 10 Histograms showing the distribution of (a) step length (km) and (b) turning angle of 6 visually (boat) tracked common terns from South Shian colony during chick-rearing period, 2011. Lines represent HMM-fitted state-dependent distributions coloured according to the decoded behavioural states.

All models fitted appeared to have similar inferred states so that the inferred states were largely insensitive to the set of models considered. However, AIC identified the same, relatively complex model (e.g., an HMM with a relatively large number of model parameters) across many species and breeding colonies, while the validation metrics identified much simpler models. During incubation, we observe that the HMM accounting for the Euclidean distance

21

of the boat to the colony as a covariate effect is mostly considered optimal compared to the chick-rearing period. Furthermore, since there are no young terns to look after at the colony during incubation, terns are likely to forage further from the colony during this period compared to chick-rearing. Thus, accounting for the distance of the terns to the colony in HMMs may provide better behavioural inference.

4 DISCUSSION

4.1 Visual tracking as a tool for validating HMMs

HMM-inferred behavioural states from telemetry data have not been validated in many previous studies due to the difficulty in obtaining concurrent observed behavioural data. However, these inferred behaviours are used in ecology to delineate important areas, such as those used for foraging, and effective conservation planning and management decisions are taken based on the location of these behaviours. Given the current climate and appetite for increasing the number of protected areas on land and sea globally (e.g., protecting 30% of the earth by 2030 target from the UN Biodiversity COP 15), it is crucial to assess the validity of behaviors inferred from HMMs used in identifying the size and location of essential areas to be protected.

In practice, behavioural states of seabirds are mostly inferred from HMMs fitted to telemetry data (Langrock et al. 2012), and our study is the first to infer behavioural states of seabirds from visual tracking data using HMMs. We acknowledge that there may be potential effects of the boat following the seabirds on their behaviour, the inferred states, and the validation process itself. However, previous studies have shown that the visual tracking method does not unduly affect bird behaviour due to a reasonable distance maintained between the individuals and the boat; moreover, most birds appear to ignore the boat (Robertson et al. 2014, Wilson et al. 2014, Perrow et al. 2011). The distance between the boat and the bird was, however, increased when there was a noticeable change in behaviour, such as evasive flight, observed for a few birds (Robertson et al. 2014, Wilson et al. 2014, Perrow et al. 2011). These previous studies did not investigate the extent to which boat-based tracks replicate the path taken by

the birds. Our study shows that movement data from the boat being used to visually track terns closely replicated those from the estimated location data of the terns being tracked, particularly for movement tracks corresponding to the foraging behavioural states of terns. Additionally, similar behavioural states of terns were inferred from HMMs fitted to the boat tracks and the corresponding actual (estimated) tern location data. We acknowledge that the boat and approximate tern position were compared for a small number of tern species and restricted to a single colony (Coquet) and breeding period (chick-rearing), which may impose limitations on the how representative the data and how generalised the interpretation of the results can be. However, there are previous studies where individual terns were tracked visually using a boat with tracks obtained from the onboard GPS as proxies for foraging tracks have been used to successfully identify foraging behaviours and areas of tern species (Wilson et al. 2014, Perrow et al. 2011).

The unique approach of the visual tracking method provides telemetry data for the boat, a proxy for the tracks of the terns they are following, and additional behavioural observation data, which are difficult to access in terrains such as the marine environment. Consequently, it allows HMM-inferred behaviours of seabirds to be validated using behavioural observations. From our findings, we can conclude that visual tracking is a suitable method to identify foraging movement and at-sea behaviour of terns, consistent with Perrow et al. (2011). Furthermore, we show that visual tracking provides an effective alternative to telemetry in contexts where attaching biologging devices may not be feasible or appropriate (e.g., in species particularly susceptible to behavioural impacts from attachment process or devices (Gillies et al. 2020).

4.2 Validating HMM-inferred behavioural states

Our study investigated the accuracy of HMMs fitted to visual tracking data from different tern species across breeding colonies in the UK during the breeding season, using behavioural observation data recorded by observers on the boats. Results suggest that HMMs can correctly infer behavioural states from tracking data. A similar observation has been shown for inferred behavioural states from HMMs using additional accelerometer and magnetometer

data from four species of albatross Conners et al. (2021) and fishermen's movement data with frequency differing from the observed behaviours (Joo et al. 2013). These methods used to infer behaviours are subject to the accuracy of the measurement devices. Our study is the first to validate HMMs using observed behaviours taken concurrently as the tracking data in the same spatial and temporal context. Generally, HMMs performed reasonably well at decoding behavioural states. However, the performance during incubation was poor compared to chick-rearing, particularly for Arctic terns at the Isle of May (42% see Table 2). Terns on the Isle of May had reduced breeding success in 2010. Therefore, terns that were tracked may have included failed or non-breeders which are not required to return to the colony regularly to attend to eggs or chicks, and so the data for this colony and year may be potentially unrepresentative of breeding adults (Wilson et al. 2014).

The capacity of HMMs in identifying and capturing most foraging behavioural activities within a foraging bout was low for roseate terns at Blue Circle and common terns at Leith during incubation in 2010 (Figure 6) and common terns at Cemlyn and Leith during chick-rearing (Figure 7). The visual tracking method was aimed at chick-rearing (2009-2011) but was extended to incubation in 2010, resulting in a reduction in the frequency of data collection (through survey effort being split between time periods) (Wilson et al. 2014), which may be a potential reason for the poor performance of fitted HMMs during incubation and chick-rearing in 2010. Observed behavioural data showed that common terns at Leith colony foraged closer to the colony during chick-rearing, 2010 (Figure 2b). The Leith common tern colony is in a port, so there may have been speed restrictions on the boat and limitations to how well the boat could closely replicate the movement of the terns. It is unclear from our study the exact reason why fitted HMMs did not identify most foraging behavioural states of common terns within foraging events at Leith and Cemlyn. However, overall, 70% (Leith) and 81% (Cemlyn) of the foraging behavioural states were decoded correctly from HMMs.

HMMs inferred foraging behavioural states 0% of the time for some observed foraging events that lasted for an average of 21 seconds. These missed foraging events were most common in chick-rearing. Terns forage close to the colony during chick-rearing and do not travel for long distances (as they do in incubation) (Eglington & Perrow 2014). Also, observers noted

24

short sessions of foraging behavioural activities of some tracked terns in some colonies (JNCC personal communication). As a result, the track of the boat may not capture tern movement corresponding to these short observed foraging events. Consequently, boat tracks may not have represented the tern's track correctly within those short phases of foraging events. As such, the HMMs fitted to boat tracks from such a scenario could not have decoded foraging states within the foraging bout from the boat tracks.

The choice of the number of behavioural states to fit in HMMs is a major challenge in animal movement modelling particularly when the goal is to infer behavioural states from telemetry data. AIC tends to select HMMs with more states but may not correspond to or have a meaningful biological interpretation of the studied animal. Pohle et al. (2017) provides practical guides in selecting the number of states to fit HMMs. Given a fixed number of states, an additional model selection process may include covariates or consider pooling across individual tracks. However, our study showed that these different models did not lead to any substantial differences between the inferred behavioural states, as identified by McClintock (2021). Therefore, fitting less complex HMMs may likely outperform complex models in inferring hidden behavioural states from movement data. As such, when behavioural inference is the study's goal, it may be preferable to consider simpler models (i.e., including a smaller number of model parameters) when choosing an appropriate HMM to fit after selecting the desired number of states.

Our findings are informative for conservation management and planning. Seabird colonies are more likely to be included as part of protected area networks due to their aggregated nature and relative ease of delineation than areas used by seabirds at sea, especially for species with large foraging ranges from the colony. Foraging areas are considered important habitats to include within seabird-protected area networks (Lascelles et al. 2016). Thus, foraging behavioural activities can be a focus for future studies looking at using behavioural states to inform conservation and management, such as identifying the optimal size and location of foraging areas around seabird colonies. In addition, our study could be extended to assess how temporal validation translates to spatial validation. The visual tracking data could be used to compare the spatial distribution of behaviours inferred from HMMs with the spatial

distribution of observed behaviours to determine the accuracy of foraging areas detected using HMMs with real-world implications for conservation and management.

Our study shows that using HMMs to infer foraging behavioural states can help identify most foraging events correctly as HMMs decoded foraging activities within observed foraging events. Furthermore, missed foraging events or bouts may be less frequent from HMMs fitted to telemetry data of seabirds as GPS devices attached to seabirds are more likely to capture movement patterns influenced by short foraging behavioural activities that last a short time than HMMs fitted to visual tracking data. Therefore, using HMMs for behavioural inference, particularly the foraging behaviour of seabirds, can aid spatial planning and inform conservation decisions, hence providing a tool for the effective management of the impact of human activities on seabirds and other species.

In summary, using HMMs to infer important conservation-relevant behaviours from telemetry data appears defensible based on our results and can inform the design of designated protected areas. Furthermore, the visual tracking method may also be a useful data collection method for ecological researchers in situations where some factors, such as small species relative to device weights, inaccessibility of colony, and species known to be particularly susceptible to disturbance, may not facilitate the attachment of GPS tags to animals. Lastly, there is evidence from our validation study that given the same number of behavioural states, there may be no substantial differences in the performance of simpler and complex HMMs in inferring behavioural states even in situations where standard model selection approaches, such as AIC, strongly suggest the use of more complex models.

DATA AVAILABILITY STATEMENT

Visual tracking data used in this work is provided by the Joint Nature Conservation Committee (JNCC).

The data and code are contained in the repository: `https://github.com/aotara/Validati ng-HMMs-project`

The visual tracks for tern species in Coquet Island Colony are already published in `https:`

26

`//hub.jncc.gov.uk/assets/0de5aa81-6aa1-4d33-a239-4484c5b13573#` while the visual tracks for other colonies considered in our paper will be published.

CONFLICT OF INTEREST STATEMENT

All authors have no conflict of interest to declare.

AUTHOR CONTRIBUTIONS

RK, VE, AB, GR, and ELJ supervised and conceived the idea. JB provided the visual tracking data. RAA conducted the analysis and led the writing of the manuscript. All authors contributed to the drafts and gave final approval for submission.

ACKNOWLEDGEMENTS

508 Science and Analytical Services Division.

References

Bennison, A., Bearhop, S., Bodey, T., Votier, S., Grecian, W., Wakefield, E., Hamer, K. & Jessopp, M. (2018), 'Search and foraging behaviors from movement data: A comparison of methods', *Ecology and Evolution* **8**, 13–24. `https://doi.org/10.1002/ece3.3593`.

Burnham, K. P. & Anderson, D. R. (2002), *Model selection and multimodel inference: A practical information- Theoretic approach (2nd ed.)*, Springer.

Conners, M. G., Michelot, T., Heywood, E. I., Orben, R. A., Phillips, R. A., Vyssotski, A. L., Shaffer, S. A. & Thorne, L. H. (2021), 'Hidden markov models identify major movement modes in accelerometer and magnetometer data from four albatross species', *Movement Ecology* **9**(7). `https://doi.org/10.1186/s40462-021-00243-z`.

Cooke, S., Hinch, S., Wikelski, M., Andrews, R., Kuchel, L., Wolcott, T. & Butler, P. (2004), 'Biotelemetry: a mechanistic approach to ecology', *Trends in Evolution and Ecology* **19**, 334–343. `https://doi.org/10.1016/j.tree.2004.04.003`.

Croxall, J., Butchart, S., Lascelles, B., Stattersfield, A., Sullivan, B., Symes, A. & Taylor, P. (2012), 'Seabird conservation status, threats and priority actions: a global assessment', *Bird Conservation International* **22**, 1–34. `https://doi.org/10.1017/S0959270912000020`.

Eglington, S. M. & Perrow, M. R. (2014), *Literature review of tern (Sterna & Sternula spp.)*, Report to JNCC, under Contract ref. C13-0204-0686. `https://data.jncc.gov.uk/data/926cdbbd-c384-42a9-b9e5-81abd778bbd0/JNCC-Report-500-Annex8-Eglington-Perrow2014.pdf`.

European-Commission, for Environment, D.-G., Mézard, N., Sundseth, K. & Wegefelt, S. (2008), *Natura 2000: protecting Europe's biodiversity*, European Commission. `https://data.europa.eu/doi/10.2779/45963`.

Gillies, N., Fayet, A. L., Padget, O., Syposz, M., Wynn, J., Bond, S., Evry, J., Kirk, H., Shoji, A., Dean, B., Freeman, R. & Guilford, T. (2020), 'Short-term behavioural impact contrasts with long-term fitness consequences of biologging in a long-lived seabird', *Scientific Reports* **10**(15064). `https://doi.org/10.1038/s41598-020-72199-w`.

Hossin, M. & Sulaiman, M. (2015), 'A review on evaluation metrics for data classification evaluations', *International Journal of Data Mining Knowledge Management Process* **5**, 01–11. `https://aircconline.com/ijdkp/V5N2/5215ijdkp01.pdf`.

JNCC (2020), 'Spas with marine components'. `https://jncc.gov.uk/our-work/spas-with-marine-components/`.

JNCC (2021), 'Seabird Population Trends and Causes of Change: 1986–2019 Report. Joint Nature Conservation Committee, Peterborough. Updated 20 May 2021'. `https://jncc.gov.uk/our-work/smp-report-1986-2019`.

Joo, R., Bertrand, S., Tam, J. & Fablet, R. (2013), 'Hidden markov models: The best models for forager movements?', *PLoSONE* **8**(8). `https://doi.org/10.1371/journal.pone.0071246`.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. & Morales, J. (2012), 'Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions.', *Ecology* **93**, 2336–2342. `https://doi.org/10.1890/11-2241.1`.

Lascelles, B., Langham, G., Ronconi, R. & Reid, J. (2012), 'From hotspots to site protection: Identifying marine protected areas for seabirds around the globe', *Biological Conservation* **156**, 5–14. `https://doi.org/10.1016/j.biocon.2011.12.008`.

Lascelles, B., Taylor, P., Miller, M., Dias, M., Oppel, S.and Torres, L., Hedd, A., Le Corre, M., Phillips, R., Shaffer, S. & Weimerskirch, H. (2016), 'Applying global criteria to tracking data to define important areas for marine conservation', *Diversity and Distributions* **22**(4), 422–431. `https://doi.org/10.1111/ddi.12411`.

McClintock, B. (2021), 'Worth the effort? A practical examination of random effects in hidden Markov models for animal telemetry data', *Methods in Ecology and Evolution* pp. 1–23. `https://doi.org/10.1111/2041-210X.13619`.

McClintock, B. T. & Michelot, T. (2018), 'momentuhmm: R package for generalized hidden markov models of animal movement', *Methods in Ecology and Evolution* **9**(6), 1518–1530. `https://doi.org/10.1111/2041-210X.12995`.

30

McKellar, A., Langrock, R., Walters, J. & Kesler, D. (2015), 'Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird', *Behavioral Ecology* **26**, 148—157. `https://doi.org/10.1093/beheco/aru171`.

Michelot, T., Langrock, R. & Patterson, T. A. (2016), 'movehmm: R package for the statistical modelling of animal movement data using hidden markov models', *Methods in Ecology and Evolution* **7**(11), 1308—-1315. `https://doi.org/10.1111/2041-210X.12578`.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. & Fryxell, J. M. (2004), 'Extracting more out of relocation data: Building movement models as mixtures of random walks', *Ecology* **85**(9), 2436–2445. `https://doi.org/10.1890/03-0269`.

Parsons, M., Mitchell, I., Butler, A., Ratcliffe, N., Frederiksen, M., Foster, S. & Reid, J. (2008), 'Seabirds as indicators of the marine environment.', *ICES Journal of Marine Science* **65**, 1520–1526. `https://doi.org/10.1093/icesjms/fsn155`.

Patterson, T., Basson, M., Bravington, M. & Gunn, J. (2009), 'Classifying movement behaviour in relation to environmental conditions using hidden markov models', *Journal of Animal Ecology* **78**(6), 1113—-1123. `https://doi.org/10.1111/j.1365-2656.2009.015 83.x`.

Perrow, M. R., Skeate, E. R. & Gilroy, J. J. (2011), 'Visual tracking from a rigid-hulled inflatable boat to determine foraging movements of breeding terns', *Journal of Field Ornithology* **82**(1), 68–79. `https://doi.org/10.1111/j.1557-9263.2010.00309.x`.

Pohle, J., Langrock, R., Beest, F. M. & Schmidt, N. M. (2017), 'Selecting the number of states in hidden markov models: Pragmatic solutions illustrated using animal movement', *Journal of Agricultural, Biological and Environmental Statistics* **22**, 270–293. `https://do i.org/10.1007/s13253-017-0283-8`.

Robertson, G., Bolton, M., Grecian, W., Wilson, L., Davies, W. & Monaghan, P. (2014), 'Resource partitioning in three congeneric sympatrically breeding seabirds: Foraging areas and prey utilization', *The Auk* **131**, 434–446. `https://doi.org/10.1642/AUK-13-243.1`.

Stanbury, A., Eaton, M., Aebischer, N., Balmer, D., Brown, A., Douse, A., Lindley, P.,

McCulloch, N., Noble, D. & Win, I. (2021), 'The status of our bird populations: the fifth birds of conservation concern in the united kingdom, channel islands and isle of man and second iucn red list assessment of extinction risk for great britain', *British Birds* **114**, 723—747. `https://britishbirds.co.uk/sites/default/files/BB_Dec21-B oCC5-IUCN2.pdf`.

Thaxter, C. B., Lascelles, B., Sugar, K., Cook, A. S., Roos, S., Bolton, M., Langston, R. H. & Burton, N. H. (2012), 'Seabird foraging ranges as a preliminary tool for identifying candidate marine protected areas', *Biological Conservation* **156**, 53–61. `https://doi.or g/10.1016/j.biocon.2011.12.009`.

Vulcano, A., Rutherford, C., Staneva, A. & Mitchell, D. (2021), 'How are seabirds doing in the EU and UK?'. `https://www.birdlife.org/news/2021/04/02/how-are-seabirds-d oing-eu-and-uk/`.

Wakefield, E. D., Owen, E., Baer, J., Carroll, M. J., Daunt, F., Dodd, S. G., Green, J. A., Guilford, T., Mavor, R. A., Miller, P. I., Newell, M. A., Newton, S. F., Robertson, G. S., Shoji, A., Soanes, L. M., Votier, S. C., Wanless, S. & Bolton, M. (2017), 'Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species', *Ecological Applications* **27**(7), 2074–2091. `https://doi.org/10.1002/eap.1591`.

Wilson, L. J., Black, J., Brewer, M. J., Potts, J. M., Kuepfer, A., Win, I., Kober, K., Bingham, C., Mavor, R. & Webb, A. (2014), 'Quantifying usage of the marine environment by terns sterna sp. around their breeding colony spas'. `https://data.jncc.gov.uk/data/926cdb bd-c384-42a9-b9e5-81abd778bbd0/JNCC-Report-500-FINAL-WEB.pdf`.

Zucchini, W., MacDonald, I. L. & Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R (2nd ed.)*, Chapman and Hall/CRC. `https://doi.org/ 10.1201/b20790`.