RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds

Ella Browning[1,2] (ID) | Mark Bolton[3] | Ellie Owen[4] | Akiko Shoji[5] | Tim Guilford[5] | Robin Freeman[2]

[1]Centre for Biodiversity and Environment Research, University College London, London, UK

[2]Institute of Zoology, Zoological Society of London, London, UK

[3]RSPB Centre for Conservation Science, Sandy, Bedfordshire, UK

[4]RSPB Centre for Conservation Science, Inverness, UK

[5]Department of Zoology, Oxford University, Oxford, UK

Correspondence
Ella Browning
Email: ella.browning.14@ucl.ac.uk
and
Robin Freeman
Email: robin.freeman@ioz.ac.uk

Abstract

1. To prevent further global declines in biodiversity, identifying and understanding key habitats is crucial for successful conservation strategies. For example, globally, seabird populations are under threat and animal movement data can identify key at-sea areas and provide valuable information on the state of marine ecosystems. To date, in order to locate these areas, studies have used global positioning system (GPS) to record position and are sometimes combined with time–depth recorder (TDR) devices to identify diving activity associated with foraging, a crucial aspect of at-sea behaviour. However, the use of additional devices such as TDRs can be expensive, logistically difficult and may adversely affect the animal. Alternatively, behaviours may be resolved from measurements derived from the movement data alone. However, this behavioural analysis frequently lacks validation data for locations predicted as foraging (or other behaviours).

2. Here, we address these issues using a combined GPS and TDR dataset from 108 individuals by training deep learning models to predict diving in European shags, common guillemots and razorbills. We validate our predictions using withheld data, producing quantitative assessment of predictive accuracy. The variables used to train these models are those recorded solely by the GPS device: variation in longitude and latitude, altitude and coverage ratio (proportion of possible fixes acquired within a set window of time).

3. Different combinations of these variables were used to explore the qualities of different models, with the optimum models for all species predicting non-diving and diving behaviour correctly over 94% and 80% of the time, respectively. We also demonstrate the superior predictive ability of these supervised deep learning models over other commonly used behavioural prediction methods such as hidden Markov models.

4. Mapping these predictions provides useful insights into the foraging activity of a range of seabird species, highlighting important at sea locations. These models have the potential to be used to analyse historic GPS datasets and further our understanding of how environmental changes have affected these seabirds over time.

KEYWORDS

common guillemot, European shag, foraging, machine learning, prediction, razorbill, time–depth recorder

1 | INTRODUCTION

Marine ecosystems are under threat from anthropogenic pressures such as climate change, ocean acidification and overfishing (Frederiksen, Edwards, Richardson, Halliday, & Wanless, 2006; Furness & Camphuysen, 1997; Halpern et al., 2008). Seabirds are the most threatened bird taxa globally (Croxall et al., 2012), with population sizes declining by 69.7% between 1950 and 2010 in response to these threats (Paleczny, Hammill, Karpouzi, & Pauly, 2015). As seabirds have a low reproductive output, high survival rate and deferred maturity (Gaston, 2004), they can be slow to recover from population crashes.

Gaining the information required for successful conservation strategies can be difficult in highly mobile and broad-ranging species, such as pelagic seabirds. Compared with the extent of protected area designation of colonies on land, seabirds are less well protected at sea, and marine nature reserves are most frequently designated directly around colonies (Guilford et al., 2008). However, foraging areas are known to be vital for breeding success in seabirds (Thaxter et al., 2012) and potential marine-protected areas (MPAs) can be identified (Guilford et al., 2009) by monitoring seabird movements. BirdLife International's Important Bird Area (IBA) Programme, a stepping stone to designating MPAs, provides a list of criteria and protocols for identifying areas critical for seabirds, which includes assessment through tracking (BirdLife International, 2010). Additionally, seabirds are valuable biological indicators for the marine environment, providing information on ecosystem health (Einoder, 2009; Furness & Camphuysen, 1997). The movements of these wide-ranging birds can inform us about the condition of large parts of the often inaccessible ocean (Mallory, Robinson, Hebert, & Forbes, 2010), and seabirds are easy to monitor as during the breeding season, they return to the same colony (Einoder, 2009).

Traditional methods of monitoring movements and populations of elusive seabirds include ringing recoveries and at-sea surveys. However, these methods do not provide detailed information on behaviour or movements at important locations (Guilford et al., 2009; Maclean, Rehfisch, Skov, & Thaxter, 2013). Advances in telemetry and biologging systems have made it possible to monitor pelagic seabirds in more detail. Initially, only larger species over 1,000*g*, such as albatrosses, were tracked (Weimerskirch et al., 2002). The recent reduction in size of devices has enabled the tracking of smaller species, such as the Manx shearwater (*Puffinus puffinus*) (*c*. 400*g*) and Black-legged Kittiwake, *Rissa tridactyla*, (*c*. 400*g*) using GPS loggers (Guilford et al., 2008, 2009; Kotzerka, Garthe, & Hatch, 2009). Early methods of determining depth use by seabirds only allowed the maximum depths reached to be recorded. More recently, however, time–depth recorder (TDR) devices have been used to record dive profiles continuously (Dean et al., 2012; Shoji et al., 2015; Wanless, Harris, Burger, & Buckland, 1997).

Much of the information about foraging behaviour has been gained through the combined use of GPS and TDR devices; the latter cost up to ten times as much as GPS devices. The quantity of data a tracking device is able to collect has risen sharply in recent years, leading to challenges in how to analyse big data (Urbano et al., 2010). The use of several devices further exacerbates the problem, increasing the amount of data a single study produces. While we have learnt a great deal about seabird distribution, much of the potential information that may be gleaned from tracking studies about animal movements remains unutilised and there is a substantial amount of historical GPS tracking data. Additionally, despite the reductions in size of transmitters and loggers, adverse effects may still be observed in study animals. Animal ethics are an important consideration in tracking studies (Kays, Crofoot, Jetz, & Wikelski, 2015), using more than one logger increases total mass, as well as significantly increasing handling time. Therefore, developing accurate methods to identify foraging locations from a single device remains important.

There are several ways in which previous studies have attempted to identify foraging behaviour in seabirds, such as, multiscale straightness index (Postlethwaite, Brown, & Dennis, 2013), first passage time (Fauchald & Tveraa, 2003), positional entropy (Roberts, Guilford, Rezek, & Biro, 2004), tortuosity (Benhamou, 2004; Dicke & Burrough, 1988), speed (Meier et al., 2015) or tortuosity and speed (Dean et al., 2015; Freeman et al., 2013; Guilford et al., 2008). Periods of low speed are generally associated with resting and high-speed directed movement with travel between foraging or resting locations and the colony. Tortuous movements at high speed are usually considered to be associated with foraging behaviour (Freeman et al., 2013; Guilford et al., 2008). Additionally, modelling methods have been used to predict when particular behaviours occur in space and time such as Gaussian mixtures (Guilford et al., 2008) or hidden Markov models (HMMs) (Dean et al., 2012; Roberts et al., 2004). The latter are state-space models, which can be used to predict the sequence of behavioural states and account for the non-independent nature of tracking data (Jonsen, Myers, & Flemming, 2003; Patterson, Basson, Bravington, & Gunn, 2009) and have been widely used to classify animal behaviours from tracking data (e.g. Block et al., 2011; Breed, Costa, Jonsen, Robinson, & Mills-Flemming, 2012; Breed, Jonsen, Myers, Bowen, & Leonard, 2009; Breed et al., 2017; Forester et al., 2007; Maxwell et al., 2011; Royer, Fromentin, & Gaspar, 2005). However, few studies are able to validate their predictions with true dive locations or with data withheld from predictive models. Dean et al. (2012) withheld a subsection of their data on which to validate their predictions for foraging behaviour in Manx shearwaters using known dive locations recorded by TDRs, revealing that their predictions were accurate.

Several studies have successfully used supervised machine learning (ML) methods (where a labelled dataset is used to learn to identify known classes) to predict animal behaviours from accelerometry data such as in cows (Martiskainen et al., 2009), cheetahs (Grünewälder et al., 2012) and penguins (Carroll, Slip, Jonsen, & Harcourt, 2014). Nathan et al. (2012) compare five supervised learning algorithms—SVMs, linear discriminant analysis, random forest (RF), classification and regression trees and artificial neural networks (ANN)—to predict behavioural modes in vultures from GPS and accelerometer data. Unsupervised approaches (where self-similar patterns are identified within a dataset) such as hidden Markov models or Gaussian mixture models have also been used extensively in identifying and modelling biologging and telemetry data (e.g. Breed et al., 2012; Gibb et al., 2017; Langrock et al., 2012; Michelot, Langrock, & Patterson, 2016;

Patterson et al., 2009). Across many of these cases, particular features (reduced metrics derived from the raw data) were extracted from the data to simplify the predictive task. This can be a laborious process and is often overlooked in the complexity of implementing such models.

Deep neural networks (a more complex, recent form of ANNs) can be used with very large input feature sets (e.g. complete pixel arrays for images or, as here, x/y values), often reducing the need for complex feature identification. This more automated form of feature extraction has been successfully applied in speech, audio and image recognition where they have outperformed other machine learning techniques (see LeCun, Bengio, & Hinton, 2015 for a review). Deep learning is a relatively new ML technique that to our knowledge has not been applied to animal tracking data or animal behavioural studies to date, but that has been suggested to be a potentially useful tool (Valletta, Torney, Kings, Thornton, & Madden, 2017).

Since 2010, the Royal Society for the Protection of Birds (RSPB) has been carrying out an extensive seabird-monitoring project around the UK and Ireland as part of the Future of the Atlantic Marine Environment (FAME, www.fameproject.eu) and Seabird Tracking and Research (STAR) projects. Included in these studies are three diving species: razorbills *Alca torda*, and common guillemots *Uria aalge* (hereafter guillemots), members of the auk family (*Alcidae*), and European shags *Phalacocorax aristotelis* (hereafter shags), a member of the cormorant family (*Phalacrocoraidae*). The foraging strategies of these three species all differ slightly as they utilise different portions of the water column to find prey: razorbills feed within the water column (Thaxter et al., 2010), whereas guillemots and shags are benthic feeders (Thaxter et al., 2010; Wanless et al., 1997), although the latter has been suggested to have a flexible foraging strategy (Wanless et al., 1997).

Here, we aim to explore whether the location of diving behaviour can be predicted purely from GPS data across three seabird species. We accomplish this using supervised deep learning models. Deep learning is a newly developed method of supervised learning, by which the relationships within data may be found without prior manipulation (LeCun et al., 2015). Our predictions are then validated using
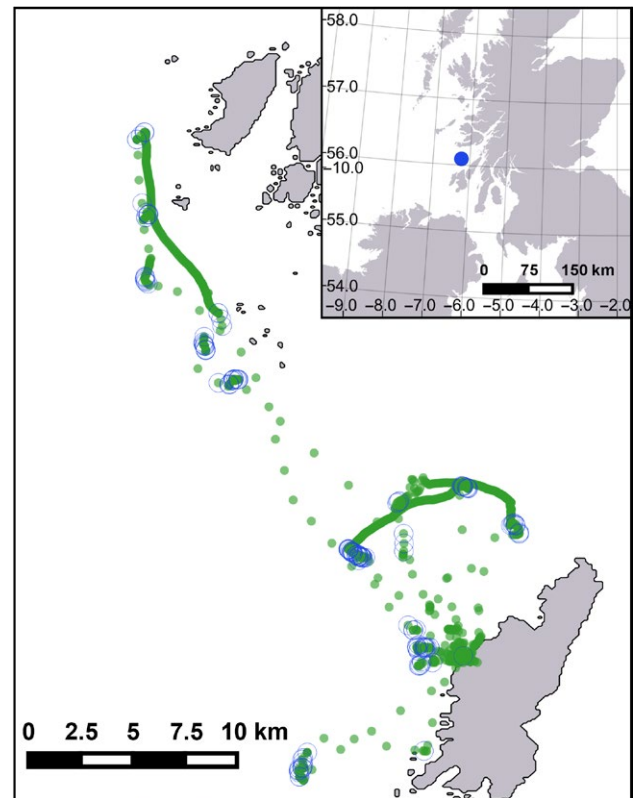

FIGURE 2 Example of an interpolated GPS track collected from a single razorbill nesting on Colonsay Island off the east coast of Scotland (RAZO0668). The location of the colony in relation to the UK is shown. The green dots are interpolated GPS locations, and the blue open circles are locations where dives were recorded by the TDR device
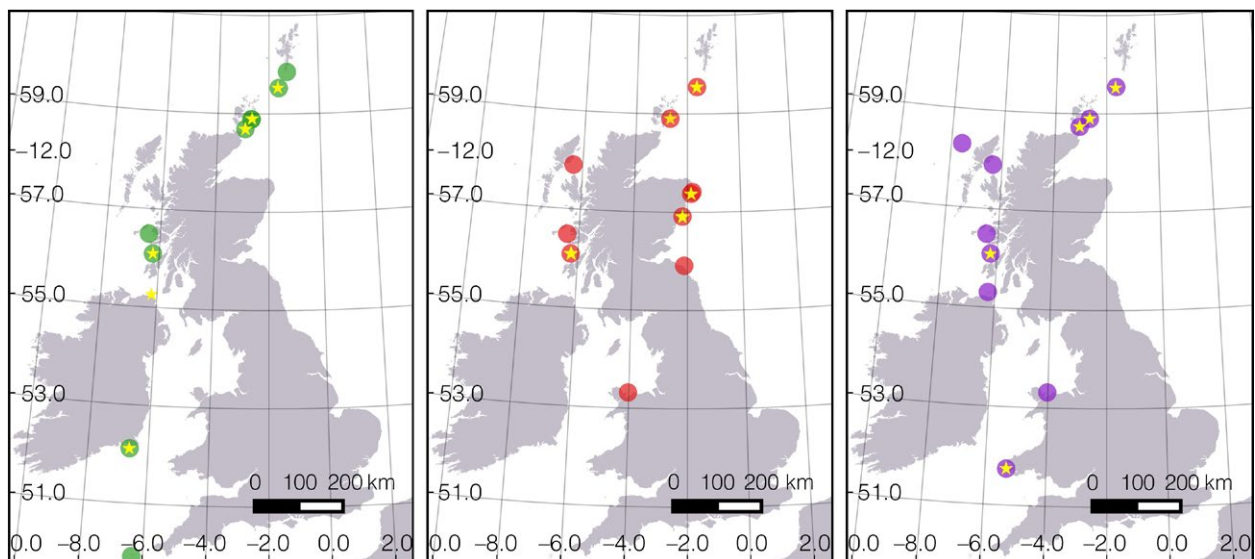

FIGURE 1 The locations of colonies where European shags (a), common guillemots (b) and razorbills (c) were tracked using either, GPS and TDR devices (yellow stars) or just GPS devices

information collected by the concurrently deployed TDR devices. Additionally, we compare the performance of our models with that of other commonly used predictive methods such as HMMs and a naïve Bayes classifier.

2 | MATERIALS AND METHODS

2.1 | The dataset

Data used were collected between 2010 and 2014 from 108 individuals (15 shags, 31 guillemots and 62 razorbills; see Table 1) tracked at eight locations (Fair Isle, Colonsay, Fowlsheugh, Orkney, Skomer, Rathlin Island, Great Saltee Island and Whinnyfold) fitted with both GPS and TDR devices (Figure 1). Additionally, data were retrieved from 291 individuals tracked with only GPS devices (80 shags, 81 guillemots and 130 razorbills).

TDR and GPS devices were attached to birds using the methods described in Dean et al. (2012) or Shoji et al. (2016). GPS devices were attached to individuals' backs using cloth-backed tape. The TDRs were attached either to a plastic ring on the leg, tail-mounted, taped directly to the GPS tags or attached to the tail feathers (Shoji et al., 2016). CEFAS TDR loggers (Cefas G5, Cefas Technology Ltd) recorded pressure data every second and temperature data either every second or every 15 s, whereas GPS devices collected data roughly every 100 s. Due to gaps in the GPS data acquisition, the tracks were interpolated using a linear method to generate points every 100 s (see Figure 2

for an example interpolated GPS track). While previously the curvilinear method has been shown to accurately interpolate tracking data (Tremblay et al., 2006), here we found it often produced erroneous interpolations where GPS data were lost for a number of minutes. We discuss the impact of using a linear method in Appendix S2, supporting information, as well as the impacts more commonly used interpolation methods had on the data and model.

2.2 | Model data

The maximum depth in each interpolated 100-s window (note this is different to the window referred to below) was calculated from the TDR data. As the TDR device recorded depth in decibar, the recorded pressures were multiplied by 1.01974 to convert to metres (Cefas Technology Ltd, 2012) (see Appendix S1). Windows containing dives where the maximum depth was greater than 3 m were classified as containing diving behaviour (although we explored the impact of this threshold, see Appendix S3). This minimised the inclusion of non-foraging dives, where birds may be bathing or carrying out other activities (Thaxter et al., 2010). The percentage of interpolated fixes that were recorded as dives (>3 m) were 18.7%, 14.1% and 14.3% for guillemots, shags and razorbills, respectively (Table 1). As such, around 81%–86% of each track contains non-diving behaviour. Two birds (a shag and razorbill) were removed from the dataset due to erroneous data, where the TDR device apparently malfunctioned and recorded depths of over 200 m. We also calculated the coverage ratio (hereafter
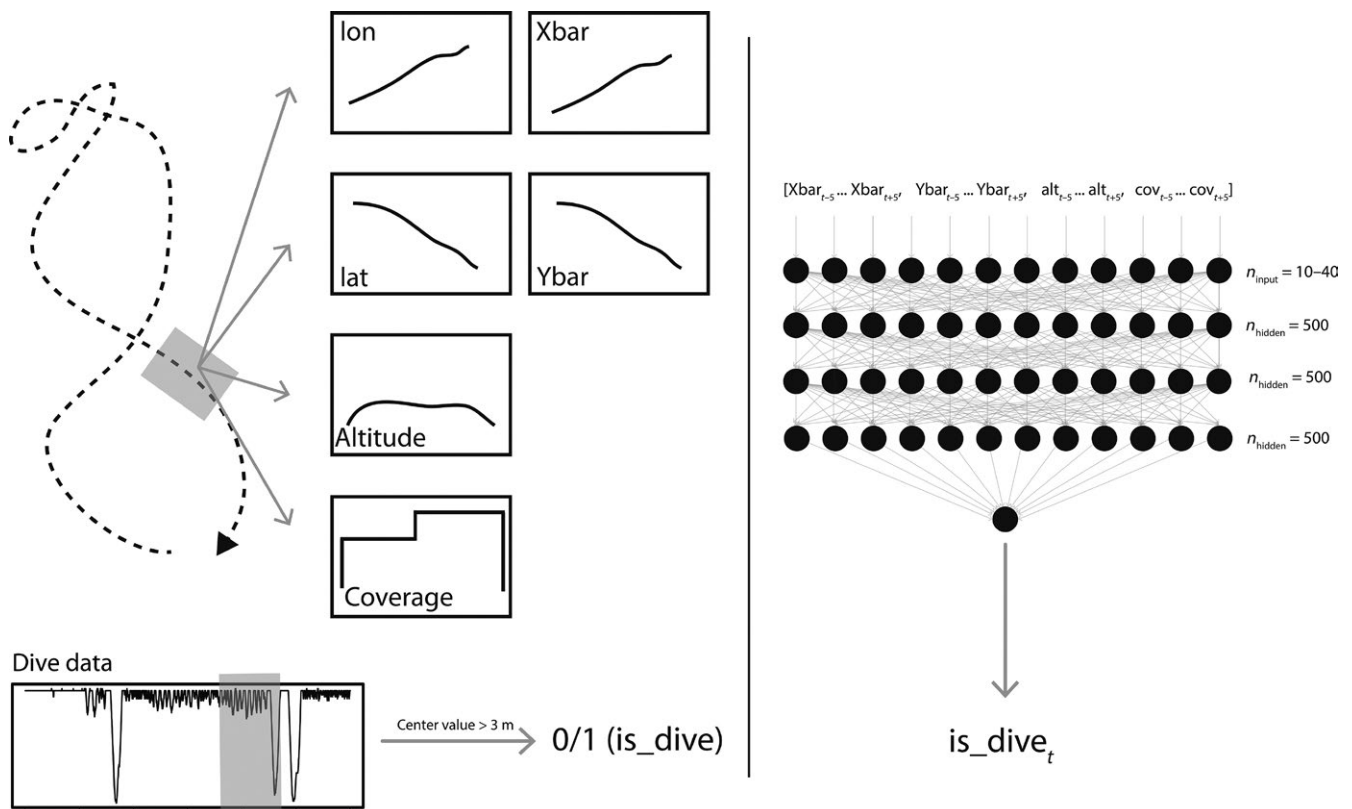

FIGURE 3 Schematic of deep learning model, showing the window of 50 points and the structure of the neural network with three hidden layers of 500 nodes

TABLE 1 The total number of birds tracked with both GPS and TDR devices, interpolated fixes, the percentage of fixes where dives deeper than 3 m were recorded and the percentage of dives recorded during the day for European shags, common guillemots and razorbills, and maximum mean maximum dive depth

Species	Number of birds	Number of interpolated fixes	Dives (%)	Dives during the day (%)	Maximum dive depth (m)	Mean maximum dive depth (m)
European shag	15	37,379	14.08	75.77	64.88	38.46
Common guillemot	31	63,925	18.66	78.29	132.43	65.56
Razorbill	62	1,62,413	14.34	76.76	88.59	20.07

TABLE 2 The mean results of the 10-fold cross-validated models that produced the optimal predictions for common guillemots, European shags and razorbills. PPV is the positive predictive value; NPV is the negative predictive value

Species	Model inputs	Training AUC	Validation AUC	Sensitivity	Specificity	PPV	NPV
Common guillemot	Coverage, *Xbar, Ybar*	0.99	0.96	80.26%	95.22%	0.74	0.97
	Altitude, *Xbar, Ybar*	0.98	0.94	74.61%	93.56%	0.66	0.96
	Coverage	0.95	0.91	72.67%	92.91%	0.70	0.94
	Altitude, coverage	0.97	0.93	71.67%	94.93%	0.70	0.95
European shag	Altitude, coverage, *Xbar, Ybar*	0.99	0.97	86.87%	96.76%	0.81	0.98
	Coverage, *Xbar, Ybar*	0.99	0.97	84.21%	95.90%	0.76	0.98
	Xbar, Ybar	0.98	0.95	77.16%	94.78%	0.70	0.96
	Coverage	0.97	0.93	74.14%	93.74%	0.65	0.96
Razorbill	Altitude, coverage, *Xbar, Ybar*	0.98	0.95	80.74%	94.14%	0.76	0.96
	Coverage	0.98	0.95	82.97%	92.82%	0.73	0.96
	Coverage, *Xbar, Ybar*	0.97	0.94	78.67%	92.75%	0.71	0.95
	Altitude, coverage	0.86	0.84	72.78%	78.11%	0.44	0.93

referred to as coverage), defined as the number of fixes acquired by the GPS over the number that could have been recorded for a given time period. A value of 1 would indicate no fixes had been missed and 0 would indicate all possible fixes were missed. All analyses were carried out using the computer programme R versions 3.1.2–3.4.2 (R Core Team, 2017).

The projection of the longitude and latitude recorded by the GPS devices was projected using the Universal Transverse Mercator (UTM) projection, giving an X and Y coordinate for each point and allowing distances to be calculated in metres.

2.3 | Model training

Deep learning models were used to predict behavioural states. H2O (Aiello, Kraljevic, & Maj, 2015), an open-source platform, was used in R to construct an artificial neural network and perform predictive modelling. A random hyperparameter search was conducted to determine the optimum model structure. The number of layers and hidden nodes per layer were varied, from one to four layers and from 20 hidden nodes to 1,000. The hyperparameter search was allowed to run for 24 hr, and automatically stopped when the top ten models had a

log loss of at most 0.001. This resulted in an optimum neural network structure of on input layer, three layers of 500 hidden nodes (rectifier nodes), followed by a softmax binary output layer.

The size of the input layer depended on the number of input variables being used (we explored the impact of withholding certain variables on the predictive accuracy). In total, we consider four variables. These variables were defined as: variation in X and Y (hereafter referred to as *Xbar and Ybar*, respectively), over a window of 5,000 s (50 points in interpolated data—giving a total of 100 values here), raw altitude over 50 points and the coverage over 50 points (Figure 3). Therefore, models including all variables used 200 input neurons, those with three variables used 150 input neurons and so on. $\bar{x}$ and $\bar{y}$ were calculated as x_i minus the mean of x_i to x_{i+w} (from i to i plus the window length, w) of interpolated data, where i is the value of X or Y at a given point.

$$\bar{x}_i = x_i - \mu\left(x_i : x_{(i+w)}\right)$$

Importantly, the variables used in the input layer were only those collected by the GPS logger—TDR data were subsequently used to validate predictions. Models were trained on species individually, with all colonies grouped together, using different input combinations of these variables to determine which might best predict non-diving and diving
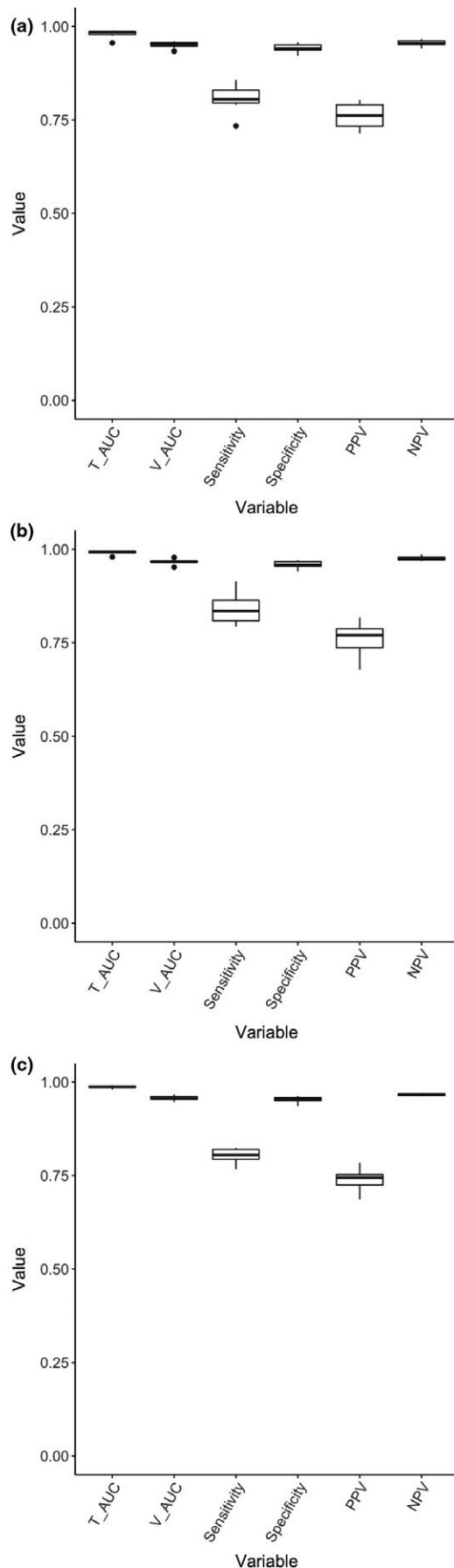
events. The impact of window size was explored; increasing it consistently increases model performance (see Appendix S4).

The data for each model were randomly split into 10 equal parts for *k*-fold cross-validation. Each model was then trained on 90% of the data and validated on the remaining 10%; this was performed for each tenth of the data. Additionally, to determine how well these models might perform on data collected on birds from different colonies or studies, leave-one-out cross-validation was also carried out. This involved removing a single bird from the dataset for each species, training the model on the other birds, and then validating the model on the single bird.

The area under the receiving operating characteristic curve (AUC) for training and validation data and sensitivity and specificity for each model was calculated, as were the positive predicted value (PPV) and negative predicted value (NPV). Sensitivity and specificity are the proportion of positives and negatives correctly identified, respectively, in the withheld data. PPV is the number of true predicted positives divided by all predicted positives and NPV is the number of true negatives divided by all predicted negatives. A perfect model, therefore, would have high sensitivity, specificity, PPV and NPV. These, along with validation AUC, were used to determine the optimal model for each species. The models were then used to predict the diving locations of birds monitored with only GPS devices.

2.4 | Alternative prediction methods

To compare predictions obtained from deep learning models with methods used in previous studies classifying foraging behaviour in seabirds HMMs, a naïve Bayes classifier and speed and tortuosity predictions were implemented on the data (see Appendix S8, supporting information for details of the latter). The Naïve Bayes classification, a supervised learning method, was implemented in R using the e1071 package (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2015). The inputs were the same as used in the H2O model, and the dependent variable (variable to be predicted) was the binary dive/not dive column.

HMMs are an unsupervised learning method that identifies discrete states within time-series data and have been used extensively to model animal movement data (Franke, Caelli, & Hudson, 2004; Patterson et al., 2009). Here, we wished to explore how well an HMM would capture our independently recorded dive bouts from the TDR data (note: these are unsupervised models, where we wish to assess how well the independently predicted states capture diving). We trained HMMs on the GPS data (longitude, latitude) of each bird for each species using the *moveHMM* package in R (Michelot et al., 2016)

(a) Orkney colony
(b) All GPS points
(c) TDR recorded dives
(d) TDR recorded non-dives
(e) Correctly predicted dives
(f) Correctly predicted non-dives
(g) Incorrectly predicted dives
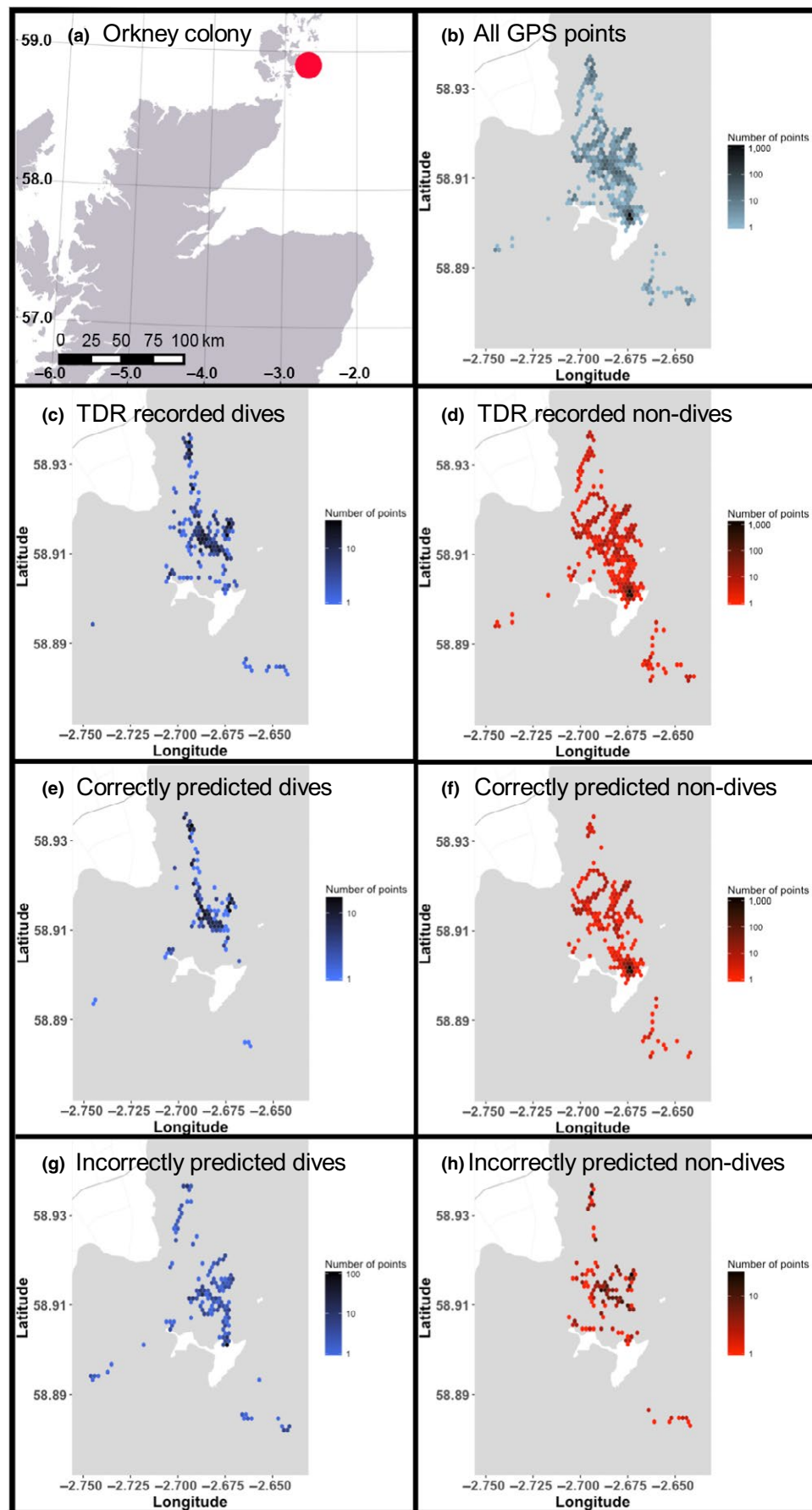**(h)** Incorrectly predicted non-dives

FIGURE 5 Maps showing the distribution of common guillemot points around the colony at Orkney in Scotland (a) from the GPS and model predictions, darker shades indicate a greater number of locations. All recorded locations are shown in (b), dive locations are shown in blue (c, e, g) and non-dives in red (d, f, h). The true locations, recorded by the TDRs, of dives and non-dives(c and d), the true predicted locations (true positives and negatives) (e and f) and the false predicted locations (false positives and negatives, e.g. where no dive occurred but the model predicted one) (g and h) are shown. Note the different scales of the number of locations

and explored how well each state predicted diving. In this case, step length and turning angle are derived from location information and used to construct the models. We initially tested between two and eight states, generally finding that three-state models produced the lowest AICs. We then constructed two- and three-state models, and, in each case, recorded the scores of the most accurately predicting state (as states are unlabelled and may be disordered across birds), therefore choosing the most generous prediction of diving from the two- and three-state models.

3 | RESULTS

3.1 | Model results

3.1.1 | 10-fold cross-validation

Nearly all models had high mean specificity values above 90%, meaning they predicted non-diving events with high accuracy, and most had mean sensitivity values above 70%, indicating they also predicted diving events well (Table 2; see Appendix S5, supporting information for the full results). The optimal models were used to predict diving locations from birds tracked only with GPS devices (Figure S7.1). In shags and razorbills, the models trained with all four variables (altitude, coverage, *Xbar* and *Ybar*) produced the most accurate predictions with the highest validation AUC values of 0.97 and 0.95, respectively (Figure 4). In shags, 86.9% and 96.7% of diving and non-diving events were correctly predicted, respectively. In razorbills, 80.7% and 94.1% of diving and non-diving events were correctly

predicted, respectively. However, coverage alone in razorbills was a more accurate predictor of dives (83.0% correct), but non-dives were predicted 92.8% correctly. Coverage, *Xbar* and *Ybar* were the most accurate predictors for dives and non-dives in guillemots, with 80.3% and 95.2%, respectively (Figure 4). Altitude alone was a poor predictor for all species, although for razorbills, *Xbar* and *Ybar* were the model inputs that produced the poorest results.

Figure 5 shows an example of guillemot data at Orkney in Scotland and the correct predictions produced from the model trained with all variables. The locations with the highest number of correctly predicted dives match the locations with the greatest number of true dives. The false positives and negatives are also shown, demonstrating the accuracy of the model.

3.1.2 | Leave-one-out cross-validation

The results of models trained with leave-one-out cross-validation showed increased variation and slightly reduced accuracy compared to 10-fold cross-validated models (Table 3; see Appendix S6, supporting information for full results). In razorbills, the variation was greatest, although the mean model output values were higher, particularly for dive prediction accuracy (Figure 6). The models trained with altitude, *Xbar* and *Ybar* performed the best for shags (mean validation AUCs of 0.85 and PPV 0.51) (Table 3); in guillemots, the models trained with coverage, *Xbar* and *Ybar* had the highest mean validation AUC (0.87) and coverage alone was the variable which produced the optimal model for razorbills with a mean validation AUC of 0.88 and a PPV of 0.62 (Figure 6c and Table 3). Like with 10-fold cross-validation, there

TABLE 3 The mean results of the leave-one-out cross-validated models that produced the optimal predictions for common guillemots, European shags and razorbills PPV is the positive predictive value; NPV is the negative predictive value

| Species	Model inputs	Training AUC	Validation AUC	Sensitivity	Specificity	PPV	NPV
Common guillemot	Coverage, *Xbar*, *Ybar*	0.90	0.87	73.15%	86.94%	0.59	0.93
	Coverage	0.82	0.80	63.99%	80.23%	0.46	0.93
	Altitude, coverage	0.81	0.81	67.89%	78.24%	0.45	0.93
	Xbar, *Ybar*	0.78	0.78	67.17%	77.88%	0.42	0.95
European shag	Altitude, coverage	0.88	0.85	66.69%	88.12%	0.51	0.95
	Altitude, coverage, *Xbar*, *Ybar*	0.91	0.84	65.15%	88.72%	0.50	0.95
	Xbar, Ybar	0.86	0.82	68.80%	82.46%	0.39	0.94
	Coverage	0.89	0.78	65.21%	81.80%	0.44	0.94
Razorbill	Coverage	0.91	0.88	72.63%	87.58%	0.62	0.93
	Coverage, *Xbar*, *Ybar*	0.91	0.88	71.25%	88.59%	0.61	0.93
	Altitude, coverage, *Xbar*, *Ybar*	0.89	0.87	73.95%	85.70%	0.61	0.93
	Altitude, coverage	0.77	0.77	82.49%	63.91%	0.35	0.95
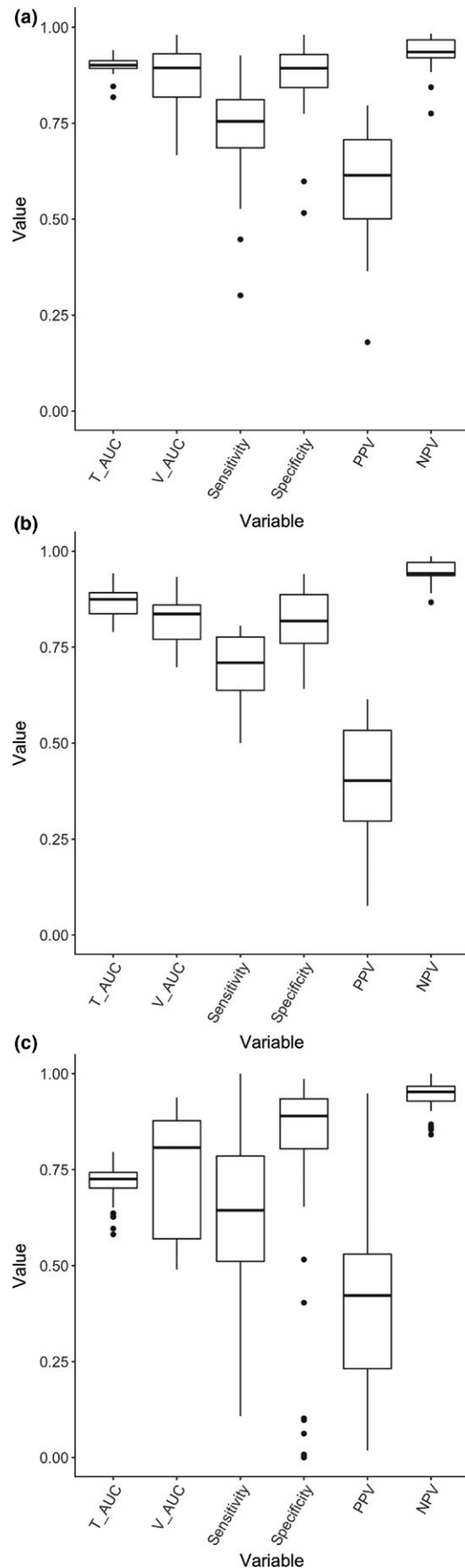
FIGURE 6 Box plots of the optimum models for (a) guillemots trained with coverage, *Xbar* and *Ybar* (b) shags trained with altitude and coverage and (c) razorbills trained with coverage, using leave-one-out cross-validation. T_AUC is the training AUC, V_AUC is the validation AUC, PPV is the positive predicted value and NPV the negative predicted value. The solid line in the middle of the boxes represents the mean

was variation in the optimum models between species; however, altitude and coverage together were also good predictors of diving behaviour in all species (Table 3). The weakest models remained the same as those using 10-fold cross-validation (Appendix S6).

3.2 | Alternative prediction methods

For several individuals of each species, the HMMs collapsed for both three- and two-state models. This meant that predictions were made for 31 and 26 guillemots, 11 and 9 shags and 41 and 37 razorbills, for two- and three-states, respectively. Two-state models were assumed to represent rest and foraging—or diving and non-diving-, and three-state models included an intermediate state. The HMMs did not predict diving behaviour as accurately as the deep learning models. Specificity, sensitivity, PPV and NPV were consistently lower for all species using both two- and three-state models (Table 4). The highest sensitivity was 81.41% for shags using a three-state HMM and specificity was 77.50% using two states. There was a trade-off where increasing the number of states increased sensitivity, so increasing the accuracy of dive behaviour prediction, but a decrease in non-diving or resting behaviour predictions, the specificity.

The predictions made using Naïve Bayes were poor for dives across all three species and variable combination, although non-dives were predicted well, with most specificity values between 0.85 and 0.96. The sensitivity was highest for guillemots and shags using altitude and coverage at 0.40. However, for razorbills, all variable combinations resulted in sensitivity values of 0.70 and 0.90.

4 | DISCUSSION

Using the combined information from GPS and TDR devices, we were able to train deep neural networks to predict the diving behaviour of shags, guillemots and razorbills. The predictions are strong and well validated with known dive locations collected by TDR loggers. Our results show that we can correctly automatically predict non-diving events over 92% of the time and diving events over 80% of the time in shags, guillemots and razorbills. The use of *Xbar*, *Ybar*, altitude and coverage to predict diving shows how GPS data can be used to monitor foraging successfully. Furthermore, both using 10-fold and leave-one-out cross-validation we demonstrate that these predictions are robust across the dataset. We also show, for the species considered, that our method produces more accurate predictions than commonly used behavioural classification methods such as HMMs.

Species	Number of states	Sensitivity	Specificity	PPV	NPV
Common guillemot	2	67.13%	53.02%	0.87	0.28
	3	84.39%	42.49%	0.87	0.32
European shag	2	56.97%	77.50%	0.95	0.32
	3	81.41%	59.84%	0.93	0.34
Razorbill	2	56.72%	61.63%	0.87	0.22
	3	75.11%	46.45%	0.88	0.23

TABLE 4 The results of two- and three-state hidden Markov models. PPV is the positive predicted value, and NPV is the negative predicted value

Under both cross-validation methods, the strongest models in all species used various combinations of the variables, from coverage alone to all four. This flexibility allows for the use of these models in a wider range of tracking studies as the data obtained can vary. For instance, the sampling rate of the logger may be unknown making the use of coverage not possible. Models using coverage, altitude or only *Xbar/Ybar* have predictive utility (see Appendices S5,S6). Altitude was included in the models as it may contain useful dive information, but inaccurate altitudes (often found with GPS devices) may make predictions here harder. However, it seems clear that combining altitude with other variables produces robust models that can predict more accurately. We hypothesised that coverage would be a powerful predictor as missed fixes may be indicative of diving and the models trained with coverage alone were in the top four for all three species. Indeed for razorbills, all the top models included coverage as an input. It should be noted that the coverage was calculated over a window of 10 possible points (1,000 s), whereas the window the model predicted dives in was 50 points (5,000 s). This implies that there is some interaction between the coverage over a 10-point window within a 50-point window the model is able to discover; however, due to the 'black box' nature of the method, it is not clear what this is.

Such 'black box' limitations are common to many supervised machine learning models, including some of the alternative methods presented here. While the complexity and volume of animal tracking data can hinder the use of more traditional statistical methods, it lends itself well to these more data-intensive machine learning approaches. However, structured models, such as HMMs, do present excellent opportunities for understanding some of the processes and mechanisms underlying the recorded data (e.g. extraction of behaviours). Here, however, we focus on the predictive power using a 'black box' method, demonstrating that diving is highly predictable, and thus highlighting the need for further understanding the processes that may underpin this relationship. In performing other supervised and unsupervised learning methods to predict, we further demonstrate the value of deep learning for predicting analysis of animal behaviour. Previous studies have used methods such as HMMs and speed and tortuosity to locate foraging patches in other seabird species and other animals, but these predictions frequently lack validation (Breed et al., 2009; Dean et al., 2012; Freeman et al., 2013; Guilford et al., 2008). However, by validating predictions made using HMMs, here we show that our model is superior in predictive power for both diving and non-diving behaviour in the species considered. The alternative methods considered also produced poorer behavioural predictions, further demonstrating the

promising application of deep learning methods for analysing animal movement datasets.

The computational power requirements and perceived complexity of constructing deep learning networks may have hindered their use in previous studies; Valletta et al. (2017) in their review highlight the potential for using deep learning methods in predictive animal behaviour studies, but conclude that they must be packaged more accessibly before wider uptake. We argue that this is no longer a barrier and deep learning models may be relatively easily implemented using existing R packages and hope there will be further studies using this method.

As the cost and size of GPS devices decreases, there is an ever-growing archive of GPS tracking data (see Kays et al., 2015 for a review) that remains largely unutilised, not only for seabirds but for other taxa as well. The robust deep learning models presented here may be used on much of this historical GPS data in order to determine foraging locations, providing valuable insight into patch use variation over time and indicating seabird responses to environmental change. Accurately mapping these sites is key for identifying candidate MPAs and informs marine developments such as the positioning of offshore wind farms (Thaxter et al., 2012).

AUTHORS' CONTRIBUTIONS

E.B., R.F. and M.B. developed the project idea. E.B. and R.F. constructed the deep learning models in R using H2O and conducted the

comparative modelling. E.B. and R.F wrote the manuscript. All authors contributed to editing of the manuscript. R.F. and M.B. supervised this work.

DATA ACCESSIBILITY

GPS track data used in the analysis are available at (http://seabirdtracking.org/mapper/contributor.php?contributor_id=950). Other data used within the analysis (GPS & dive data, and associated matrices used within the models) are available at https://doi.org/10.5061/dryad.t7ck5 (Browning et al., 2017), the GPS dive data are under a one-year embargo. Please contact the corresponding authors for more information.

ORCID

Ella Browning 🔗 http://orcid.org/0000-0002-7959-9292

REFERENCES

Aiello, S., Kraljevic, T., & Maj, P. (2015). R Interface for H2O. R package version 3.0.0.22. http:www.h2o.ai

Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, *229*, 209–220. https://doi.org/10.1016/j.jtbi.2004.03.016

BirdLife International. (2010). *Marine important bird areas toolkit: Standardised techniques for identifying priority sites for the conservation of seabirds at-sea*. Cambridge, UK: BirdLife International.

Block, B. A., Jonsen, I. D., Jorgensen, S. J., Winship, A. J., Shaffer, S. A., Bograd, S. J., ... Costa, D. P. (2011). Tracking apex marine predator movements in a dynamic ocean. *Nature*, *475*, 86–90. https://doi.org/10.1038/nature10082

Breed, G. A., Costa, D. P., Jonsen, I. D., Robinson, P. W., & Mills-Flemming, J. (2012). State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecological Modelling*, *235–236*, 49–58. https://doi.org/10.1016/j.ecolmodel.2012.03.021

Breed, G. A., Jonsen, I. D., Myers, R. A., Bowen, W. D., & Leonard, M. L. (2009). Sex-specific, seasonal foraging tactics of adult grey seals (*Halichoerus grypus*) revealed by state–space analysis. *Ecology*, *90*, 3209–3221. https://doi.org/10.1890/07-1483.1

Breed, G. A., Matthews, C. J. D., Marcoux, M., Higdon, J. W., LeBlanc, B., Petersen, S. D., ... Ferguson, S. H. (2017). Sustained disruption of narwhal habitat use and behavior in the presence of Arctic killer whales. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 2628–2633. https://doi.org/10.1073/pnas.1611707114

Browning, E., Bolton, M., Owen, E., Shoji, A., Guilford, T., & Freeman, R. (2017). Data from: Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.t7ck5

Carroll, G., Slip, D., Jonsen, I., & Harcourt, R. (2014). Supervised accelerometry analysis can identify prey capture by penguins at sea. *The Journal of Experimental Biology*, *217*, 4295–4302. https://doi.org/10.1242/jeb.113076

Cefas Technology Ltd. (2012). Cefas Host User Manual.

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., 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

Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R. A., Perrins, C. M., & Guilford, T. (2012). Behavioural mapping of a pelagic seabird: Combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *Journal of The Royal Society Interface*, *10*, 20120570. https://doi.org/10.1098/rsif.2012.0570

Dean, B., Kirk, H., Fayet, A., Shoji, A., Freeman, R., Leonard, K., ... Guilford, T. (2015). Simultaneous multi-colony tracking of a pelagic seabird reveals cross-colony utilization of a shared foraging area. *Marine Ecology Progress Series*, *538*, 239–248. https://doi.org/10.3354/meps11443

Dicke, M., & Burrough, P. A. (1988). Using fractal dimensions for characterizing tortuosity of animal trails. *Physiological Entomology*, *13*, 393–398. https://doi.org/10.1111/j.1365-3032.1988.tb01122.x

Einoder, L. D. (2009). A review of the use of seabirds as indicators in fisheries and ecosystem management. *Fisheries Research*, *95*, 6–13. https://doi.org/10.1016/j.fishres.2008.09.024

Fauchald, P., & Tveraa, T. (2003). Using first-passage time in the analysis of area-restricted search and habitat selection on JSTOR. *Ecology*, *84*, 282–288. https://doi.org/10.1890/0012-9658(2003)084[0282:UFPTIT]2.0.CO;2

Forester, J. D., Ives, A. R., Turner, M. G., Anderson, D. P., Fortin, D., Beyer, H. L., ... Boyce, M. S. (2007). State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park. *Ecological Monographs*, *77*, 285–299. https://doi.org/10.1890/06-0534

Franke, A., Caelli, T., & Hudson, R. J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, *173*, 259–270. https://doi.org/10.1016/j.ecolmodel.2003.06.004

Frederiksen, M., Edwards, M., Richardson, A. J., Halliday, N. C., & Wanless, S. (2006). From plankton to top predators: Bottom-up control of a marine food web across four trophic levels. *Journal of Animal Ecology*, *75*, 1259–1268. https://doi.org/10.1111/jae.2006.75.issue-6

Freeman, R., Dean, B., Kirk, H., Leonard, K., Phillips, R. A., Perrins, C. M., & Guilford, T. (2013). Predictive ethoinformatics reveals the complex migratory behaviour of a pelagic seabird, the Manx Shearwater. *Journal of the Royal Society, Interface/The Royal Society*, *10*, 20130279. https://doi.org/10.1098/rsif.2013.0279

Furness, R., & Camphuysen, C. J. (1997). Seabirds as monitors of the marine environment. *ICES Journal of Marine Science*, *54*, 726–737. https://doi.org/10.1006/jmsc.1997.0243

Gaston, A. J. (2004). *Seabirds: A natural history*. London: A & C Black Publishers Ltd.

Gibb, R., Shoji, A., Fayet, A. L., Perrins, C. M., Guilford, T., & Freeman, R. (2017). Remotely sensed wind speed predicts soaring behaviour in a wide-ranging pelagic seabird. *Journal of the Royal Society, Interface*, *14*, 20170262. https://doi.org/10.1098/rsif.2017.0262

Grünewälder, S., Broekhuis, F., Macdonald, D. W., Wilson, A. M., McNutt, J. W., Shawe-Taylor, J., & Hailes, S. (2012). Movement activity based classification of animal behaviour with an application to data from Cheetah (*Acinonyx jubatus*). *PLoS ONE*, *7*, e49120. https://doi.org/10.1371/journal.pone.0049120

Guilford, T. C., Meade, J., Freeman, R., Biro, D., Evans, T., Bonadonna, F., ... Perrins, C. M. (2008). GPS tracking of the foraging movements of Manx Shearwaters *Puffinus puffinus* breeding on Skomer Island, Wales. *Ibis*, *150*, 462–473. https://doi.org/10.1111/j.1474-919X.2008.00805.x

Guilford, T., Meade, J., Willis, J., Phillips, R., Boyle, D., Roberts, S., ... Perrins, C. (2009). Migration and stopover in a small pelagic seabird, the Manx shearwater *Puffinus puffinus*: Insights from machine learning. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1215–1223. https://doi.org/10.1098/rspb.2008.1577

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., ... Watson, R. (2008). A global map of human impact on marine ecosystems. *Science (New York, N.Y.)*, *319*, 948–952. https://doi.org/10.1126/science.1149345

Jonsen, I. D., Myers, R. A., & Flemming, J. M. (2003). Meta-analysis of animal movement using state-space models. *Ecology*, *84*, 3055–3063. https://doi.org/10.1890/02-0670

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*, aaa2478. https://doi.org/10.1126/science.aaa2478

Kotzerka, J., Garthe, S., & Hatch, S. A. (2009). GPS tracking devices reveal foraging strategies of Black-legged Kittiwakes. *Journal of Ornithology*, *151*, 459–467. https://doi.org/10.0.3.239/s10336-009-0479-y

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. https://doi.org/10.1890/11-2241.1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539

Maclean, I. M. D., Rehfisch, M. M., Skov, H., & Thaxter, C. B. (2013). Evaluating the statistical power of detecting changes in the abundance of seabirds at sea. *Ibis*, *155*, 113–126. https://doi.org/10.1111/j.1474-919X.2012.01272.x

Mallory, M. L., Robinson, S. A., Hebert, C. E., & Forbes, M. R. (2010). Seabirds as indicators of aquatic ecosystem conditions: A case for gathering multiple proxies of seabird health. *Marine Pollution Bulletin*, *60*, 7–12. https://doi.org/10.1016/j.marpolbul.2009.08.024

Martiskainen, P., Järvinen, M., Skön, J.-P., 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. https://doi.org/10.1016/j.applanim.2009.03.005

Maxwell, S. M., Breed, G. A., Nickel, B. A., Makanga-Bahouna, J., Pemo-Makaya, E., Parnell, R. J., ... Coyne, M. S. (2011). Using satellite tracking to optimize protection of long-lived marine species: Olive ridley sea turtle conservation in Central Africa. *PLoS ONE*, *6*, e19905. https://doi.org/10.1371/journal.pone.0019905

Meier, R. E., Wynn, R. B., Votier, S. C., McMinn Grivé, M., Rodríguez, A., Maurice, L., ... Guilford, T. (2015). Consistent foraging areas and commuting corridors of the critically endangered Balearic shearwater *Puffinus mauretanicus* in the northwestern Mediterranean. *Biological Conservation*, *190*, 87–97. https://doi.org/10.1016/j.biocon.2015.05.012

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. *R package version 1.6-7*. Retrieved from: https://CRAN.R-project.org/package=e1071.

Michelot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, *7*, 1308–1315. https://doi.org/10.1111/2041-210X.12578

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., & Getz, W. M. (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. https://doi.org/10.1242/jeb.058602

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population trend of the world's monitored seabirds, 1950–2010. *PLoS ONE*, *10*, e0129342. https://doi.org/10.1371/journal.pone.0129342

Patterson, T. A., Basson, M., Bravington, M. V., & Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *The Journal of Animal Ecology*, *78*, 1113–1123. https://doi.org/10.1111/jae.2009.78.issue-6

Postlethwaite, C. M., Brown, P., & Dennis, T. E. (2013). A new multi-scale measure for analysing animal movement data. *Journal of Theoretical Biology*, *317*, 175–185. https://doi.org/10.1016/j.jtbi.2012.10.007

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/ [accessed 7 October 2017].

Roberts, S., Guilford, T., Rezek, I., & Biro, D. (2004). Positional entropy during pigeon homing I: Application of Bayesian latent state modelling. *Journal of Theoretical Biology*, *227*, 39–50. https://doi.org/10.1016/j.jtbi.2003.07.002

Royer, F., Fromentin, J.-M., & Gaspar, P. (2005). A state-space model to derive bluefin tuna movement and habitat from archival tags. *Oikos*, *109*, 473–484. https://doi.org/10.1111/oik.2005.109.issue-3

Shoji, A., Aris-Brosou, S., Owen, E., Bolton, M., Boyle, D., Fayet, A., ... Guilford, T. (2016). Foraging flexibility and search patterns are unlinked during breeding in a free-ranging seabird. *Marine Biology*, *163*, 72. https://doi.org/10.1007/s00227-016-2826-x

Shoji, A., Elliott, K., Fayet, A., Boyle, D., Perrins, C., & Guilford, T. (2015). Foraging behaviour of sympatric razorbills and puffins. *Marine Ecology Progress Series*, *520*, 257–267. https://doi.org/10.3354/meps11080

Thaxter, C. B., Lascelles, B., Sugar, K., Cook, A. S. C. P., Roos, S., Bolton, M., ... Burton, N. H. K. (2012). Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. *Biological Conservation*, *156*, 53–61. https://doi.org/10.1016/j.biocon.2011.12.009

Thaxter, C. B., Wanless, S., Daunt, F., Harris, M. P., Benvenuti, S., Watanuki, Y., ... Hamer, K. C. (2010). Influence of wing loading on the trade-off between pursuit-diving and flight in common guillemots and razorbills. *The Journal of Experimental Biology*, *213*, 1018–1025. https://doi.org/10.1242/jeb.037390

Tremblay, Y., Shaffer, S. A., Fowler, S. L., Kuhn, C. E., McDonald, B. I., Weise, M. J., ... Costa, D. P. (2006). Interpolation of animal tracking data in a fluid environment. *Journal of Experimental Biology*, *209*, 128–140. https://doi.org/10.1242/jeb.01970

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., & Neteler, M. (2010). Wildlife tracking data management: A new vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *365*, 2177–2185. https://doi.org/10.1098/rstb.2010.0081

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Wanless, S., Harris, M. P., Burger, A. E., & Buckland, S. T. (1997). Use of time-at-depth recorders for estimating depth and diving performance of European shags. *Journal of Field Ornithology*, *68*, 547–561.

Weimerskirch, H., Bonadonna, F., Bailleul, F., Mabille, G., Dell'Omo, G., & Lipp, H.-P. (2002). GPS tracking of foraging albatrosses. *Science (New York, N.Y.)*, *295*, 1259. https://doi.org/10.1126/science.1068034

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.