INTERFACE

royalsocietypublishing.org/rsif

Research

Cite this article: Swaby L, Carr P, Dunn R, Nicoll MAC, Trevail AM, Votier SC, Wood H, Freeman R. 2026 Deep neural networks to predict foraging behaviour: saltwater immersion data can accurately predict diving in seabirds. *J. R. Soc. Interface* **23**: 20250172.
https://doi.org/10.1098/rsif.2025.0172

Received: 27 February 2025
Accepted: 2 April 2026

Subject Category:
Life Sciences–Engineering interface

Subject Areas:
environmental science

Keywords:
foraging, behaviour, biologging, geolocation, deep learning, tracking, seabirds, prediction

Author for correspondence:
Robin Freeman
e-mail: robin.freeman@ioz.ac.uk

Supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8479352.

THE ROYAL SOCIETY
PUBLISHING

Deep neural networks to predict foraging behaviour: saltwater immersion data can accurately predict diving in seabirds

Luke Swaby[1,2], Peter Carr[2], Ruth Dunn[3,4], Malcolm A. C. Nicoll[2], Alice M. Trevail[5], Stephen C. Votier[4], Hannah Wood[2,4,5] and Robin Freeman[2,6]

[1]Imperial College London, London, UK
[2]Institute of Zoology, Zoological Society of London, London, UK
[3]Lancaster Environment Centre, Lancaster University, Lancaster, UK
[4]The Lyell Centre, Heriot-Watt University, Edinburgh, UK
[5]Environment and Sustainability Institute, University of Exeter, Penryn, UK
[6]Centre for Biodiversity and Environment Research, University College London, London, UK

RD, 0000-0003-0927-2734; AMT, 0000-0002-6459-5213; RF, 0000-0002-0560-8942

The accurate identification of foraging locations is critical for wildlife conservation. While remote sensing and biologging devices provide much of the necessary data, their deployment is often complicated by factors such as weight, battery life and sensor capacity, limiting their effectiveness for long-term tracking of wide-ranging species. In this study, we evaluate the effectiveness of saltwater immersion data from light-level geolocation loggers (global location sensor; GLS) as a predictor of foraging behaviour in a pursuit-diving seabird, the red-footed booby (*Sula sula rubripes*). Using co-deployed tri-axial acceleration data as a high-resolution benchmark, we compare the performance of deep learning models for classifying dive and non-dive states. Predictions are cross-validated on withheld individuals for generalizability. Using a small pilot dataset, we find that models trained solely on GLS data only slightly underperform those trained on acceleration data despite the resolution discrepancy, classifying the diving behaviours of unseen birds with 93.65% accuracy. These findings suggest that GLS data alone may be sufficient to reliably infer dive events and, by extension, foraging locations, for pursuit-diving seabirds, providing a minimally invasive, scalable method to enrich year-round GLS migratory tracking studies using models derived from co-deployment of GLS and global positioning system devices.

1. Introduction

In the marine environment, the foraging behaviours of many predators are shaped by the distribution and predictability of resources across large spatial scales and are accordingly considered to encode valuable information about global ecosystem health [1–3]. Pelagic seabirds, in particular, are often proposed as valuable ecological indicators and have been used extensively to inform ecosystem management and conservation strategies [4–8]. However, with direct observation largely precluded by their elusive and wide-ranging nature, fine-scale behavioural data on these species have only become available in recent years thanks to innovations in telemetry and biologging technology, which have liberated researchers from the constraints of sparse presence/absence datasets generated by traditional collection methods such as at-sea surveys and ringing recoveries [9–15].

Perhaps most notably, the advent and declining size and cost of global positioning systems (GPS) and time-depth recorders (TDR) have enabled the collection of vast quantities of high-resolution movement data that have been especially valuable in identifying the behaviours associated with foraging for pursuit-diving seabirds [5,16]. Studies investigating such behaviours typically use a combination of the two, mapping continuous TDR-derived dive profiles to GPS coordinates to pinpoint dive events that can serve as proxies for identifying foraging locations [17–20]. However, this technique has several limitations. Multi-sensor deployments increase net device load, can cause significant behavioural confounds [21–24], and when deployed with a harness can have acute impacts [25,26]. For seabirds specifically, Gillies *et al*. [27] found that Manx shearwaters (*Puffinus puffinus*) fitted with 17 g GPS loggers (4.2% of bird body mass) more than doubled the length of their foraging trips and gained 25% less body mass from those same trips compared with birds carrying no device or small 2.5 g leg-mounted global location sensor (GLS) devices (approx. 0.62% of bird body mass). Furthermore, the expense and relatively short battery lives of high-sampling-rate GPS loggers often restrict the duration of studies that use them to just several weeks or months.

One alternative to GPS for long-term animal tracking is light-level geolocation (GLS). These lightweight devices (greater than 0.3 g) are equipped with wet/dry sensors as well as ambient light sensors to determine location and are often used to infer bird migration routes [28,29]. The spatio-temporal resolution of the location data they produce is coarse compared with GPS data (approx. 12 h and up to 304 ± 413 km; [30,31]). However, their exceptional battery lives permit deployments lasting several years, resulting in vast GLS datasets covering numerous species and individuals for which no other data are available. This widespread availability of existing data, in combination with the device's ultra-light weight and long operational life, offers the opportunity to develop methods to increase the value of data from these devices. In particular, while the location data they produce is coarse, they often collect coincident saltwater immersion data at a much higher temporal resolution. By leveraging modern analytical techniques such as deep learning, researchers may be able to extract much finer-scale behavioural insights from such data than previously thought possible. Indeed, if it were possible, for instance, to confidently infer foraging behaviour from GLS/immersion data alone, it would significantly extend the reach of movement ecology studies at a fraction of the cost while minimizing potential impacts on study animals.

Traditionally, the challenges of analysing large, complex biologging datasets have been tackled in behavioural studies with unsupervised approaches such as state-space models (e.g. hidden Markov models and Gaussian mixture models [19,32–34]. More recently, some have undertaken the laborious task of assembling labelled datasets to leverage the power of traditional supervised machine learning (ML) techniques that can achieve similar predictive feats with greater accuracy [9,35–38]. However, many of these techniques require substantial user input, such as manual feature extraction, which in turn requires substantial domain expertise and can dramatically increase implementation complexity.

Deep neural networks (DNNs) are a family of highly flexible, general-purpose ML models that learn hierarchical representations directly from raw inputs, thereby reducing the need for manual feature engineering. This capability has led to state-of-the-art performance across various domains, including computer vision, natural language processing and bioinformatics. ([39,40], see [41–43] for reviews). In the context of behavioural ecology, Browning *et al*. [44] recently demonstrated that DNNs outperform both hidden Markov models and a naive Bayes classifier in predicting the diving behaviour of three pelagic seabird species (*Gulosus aristotelis, Uria aalge* and *Alca torda*) from GPS data alone. These findings not only attest to the untapped potential of DNNs as behavioural classification models, but also importantly highlight a potential shift away from relying on costly and cumbersome TDR devices for identifying seabird foraging. However, with GPS loggers still limited by short battery life and potential challenges associated with long-term device attachment (e.g. using a harness), there remains a strong incentive to explore whether similar insights can be derived from a less obstructive and longer-lasting alternative such as GLS.

Here, we use GPS devices with integrated accelerometers and time-depth/pressure recorders that enable us to independently identify dives. We then use DNNs to evaluate how well co-deployed, GLS-derived saltwater immersion data (hereafter, *immersion*) are able to predict these dives in a pelagic seabird. As a performance benchmark, we also assess the predictive power of the high-resolution tri-axial acceleration data (hereafter, *acceleration*) for the same task. Building such models on the co-deployments in our study here would enable their application to individuals tracked solely with GLS-derived immersion data. Acceleration data is a useful benchmark as it has previously been demonstrated as a useful and accurate method for identifying behaviour and energy expenditure (e.g. [45,46]). We construct our datasets using a sliding window approach, assigning depth-validated binary labels from pressure sensor data to indicate the occurrence of diving events. For both predictors (*immersion* and *acceleration*), we also examine the impact of input window width on model performance and map dive predictions to GPS coordinates under a custom georeferencing method to examine their spatial distributions. We apply these methods to investigate the foraging behaviour of the red-footed booby (*Sula sula*; hereafter 'RFB'), a pantropical, diurnal, central-place foraging seabird widely understood to be the most pelagic of its genus [13,47,48]. We have recently explored the spatial foraging patterns and environmental drivers of diving behaviour in this species [13,20]. By combining data from multi-sensor deployments, we here investigate if models trained on immersion data alone can accurately predict diving behaviour and if they perform as well as those trained on tri-axial acceleration data, exploring the potential of immersion sensors as lightweight, long-term alternatives for monitoring diving behaviour in a pelagic species.

2. Methods and material

2.1. Problem formulation

We evaluate the ability of DNNs to identify dive events in fixed-length rolling windows of sensor time-series data. Specifically, we compare two different predictors: tri-axial acceleration data and GLS immersion data, evaluating their relative effectiveness in detecting dive behaviour. We formally define our problem as a binary classification task in which the objective is to learn a function

$$f_\theta : R^{d \times T} \rightarrow \{0,1\}$$

where

- d represents the number of predictor features (three for acceleration, one for immersion).
- T is the rolling window length.
- f_θ is the DNN, parametrized by θ, tasked with predicting the presence ($\widehat{y_t} = 1$) or absence ($\widehat{y_t} = 0$) of a dive event within each window $x_t \in \mathbb{R}^{d \times T}$.

2.2. Data collection

Fieldwork was conducted at Barton Point, Diego Garcia in the Chagos Archipelago, Indian Ocean (07°14′ S, 72°26′E). Fifteen adult RFBs were caught by hand over a three to six days period in February 2019 during chick-rearing. All individuals were fitted with high-sampling-rate tail-mounted multi-sensor GPS loggers (AxyTrek Marine, Technosmart, 18 g, 1.7–2.3% bird body mass, range = 790–1020 g; devices were tail mounted oriented along the body axis, Trevail *et al.* [13], Dunn *et al.* [20] for further details) and leg-mounted light-level GLS/saltwater immersion loggers (Intigeo C330, Migrate Technology, 3 g, 0.3–0.4% bird body mass, attached to PVC leg rings from Interrex-rings, Poland). Geolocator devices were configured to use mode '9', light range 4, max light regime (sampling light every 60 s, recording the maximum value in each 5 min window), and sampling a binary wet/dry value every 6 s (recording the duration of time spent before switching state). GPS loggers were configured to record geographical location every 30 s, pressure/depth every 1 s, and tri-axial acceleration at 25 Hz (±2G, 10-bit). Figure 1 illustrates an example of how these data are distributed over the study period for a single bird. Of the 15 sets deployed, failure to recover hardware or subsequent data corruption meant that data could only be successfully extracted from nine GPS and six GLS loggers.

2.3. Data pre-processing

All pre-processing was carried out using the programming languages R (v. 4.1.0) [49] and Python (v. 3.9.6). First, raw pressure readings from the GPS logger were converted to depth data using XManager software (https://www.technosmart.eu/), and subsequently 'de-noised' to control for device idiosyncrasies. This was achieved by offsetting all values within a rolling window of 30 records by the median of that window, effectively smoothing the low-level background fluctuations in each time series while preserving the shape and location of each spike, thus enabling the determination of a fixed threshold for identifying dive events (see electronic supplementary material 'Depth de-noising & window size'). For each bird, all data preceding the first aerial departure and following the final return to the island were clipped to ensure that none of the data reviewed here were influenced by device attachment/detachment processes (as multiple devices were logging data before and after their time fixed to a bird). GPS coordinates were also linearly interpolated in gaps of 60 s or more (i.e. where one or more readings had been missed, to a consistent 30 s resolution). GLS data (immersion) were then matched by ID to their corresponding GPS and depth recordings (dropping excess data according to the same time frames). Ultimately, this resulted in a preliminary dataset comprising 87 787 041 multivariate samples of 25 Hz tri-axial acceleration data ($d = 3$) across nine birds for acceleration, and 214 838 univariate samples of GLS immersion data ($d = 1$) across five birds for GLS. To visualize these data, we mapped at-sea GPS tracks for all nine birds (trimmed to first departure and last arrival to colony) and calculated for each bird the total time tracked, total distance travelled (km), furthest distance from the nest (assumed to be the mode lat/lon coordinates), maximum depth recorded and the total number of observed dive and non-dive events (see figure 2; method for determining these described below).

From the processed data, several labelled datasets were then constructed for each predictor (immersion and acceleration) using a rolling window of raw data points as inputs and binary values indicating whether a dive had occurred within each given window as the corresponding targets. Dives were classified in windows where at least one depth value exceeded a predetermined threshold of 0.1 m, which was chosen through visual inspection of depth time-series plots to distinguish periods of residual background noise from the periods of sharp fluctuation that are characteristic of genuine diving events. A range of window widths was tested for each predictor to determine the optimum. These were chosen to span time intervals that could reasonably be expected to capture dive behaviour based on the resolution and distribution of the data (2, 4, 6, 8 and 10 s for acceleration, and 1, 3, 5, 7 and 9 min for immersion). To investigate the effects of using a much simpler 'diluted' metric from the acceleration data, an additional dataset was also created using only the mean and sum of absolute differences of 4 min windows of data points along the z-axis as inputs.

The method described above produced a number of large datasets that were highly imbalanced (up to 1 TB for larger window widths, in which only approx. 0.18% of rows contained dive behaviour), so to sidestep consequent performance and memory-related issues, a smaller, more balanced subset was extracted from each by randomly undersampling the majority class (i.e. non-dives) to roughly match the number of dive rows, reducing the maximum file size to approximately 2 GB.

2.4. Model training

Model training and evaluation were conducted using Keras (v. 2.5.0), implementing a basic feed-forward architecture with two hidden layers, each consisting of 200 ReLU-activated nodes and a dropout rate of 0.2 to mitigate overfitting. The final output layer was a binary softmax classifier, producing probabilistic predictions for each behavioural state $P(y_t = 1 \mid x_t)$. The

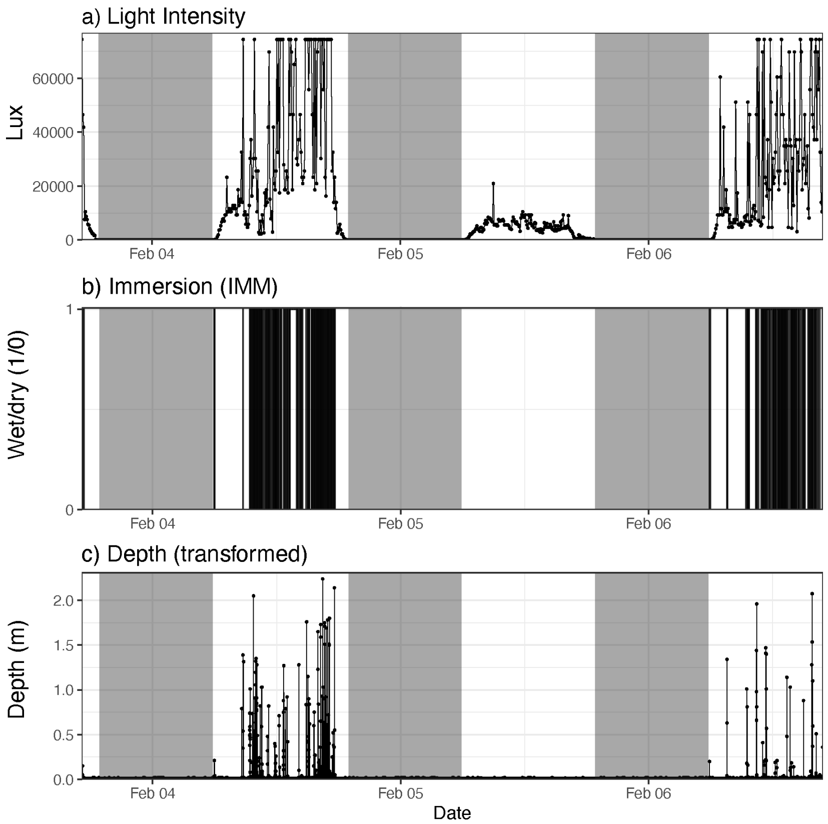

Figure 1. Example distribution showing a 3 days section of (a) ambient light level Lux; (b) binary saltwater immersion; and (c) de-noised depth data for a single bird (ch_gps16) over the period it was tracked. Dark (night-time) periods are shown shaded grey. The plot shows how the bird interacts with the water exclusively during daylight hours and spends the whole of 5 February off the water, probably on the nest when LUX levels are consistently low. Spikes in depth (i.e. diving events) correlate strongly with periods immersed.

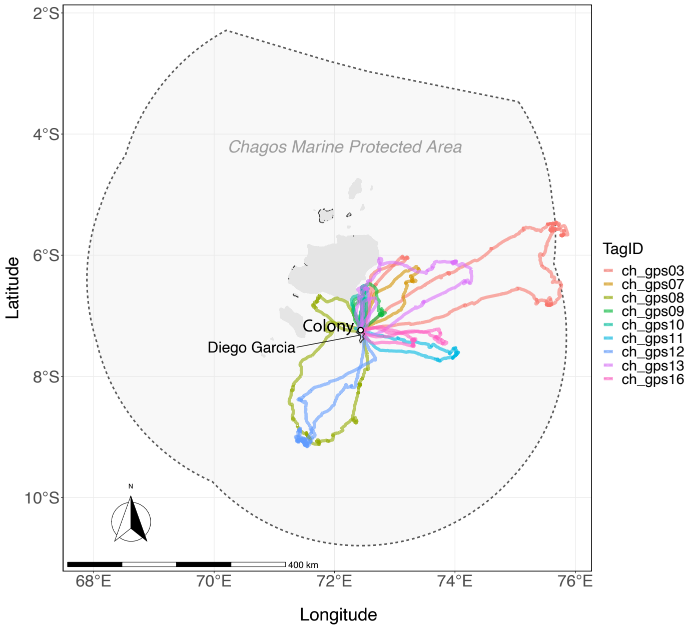

Figure 2. Flight paths for all nine GPS-tagged red-footed boobies in the Chagos Archipelago Marine Protected Area (outlined).

shape of the input layer varied based on the selected rolling window width T, the number of predictor features d and the predictor sampling frequency s (Hz), following the equation: $input = T \times d \times s$. Input layer dimensionality therefore ranged from 10 nodes where a window width of 1 min ($T = 60$) was used, to 90 where a width of 9 min ($T = 540$) was used for the univariate immersion data ($d = 1, s = \frac{1}{6}$), and two nodes where summary statistics were used to 750 where a width of 10 s was used for the 25 Hz tri-axial acceleration data ($d = 3, s = 25$).

2.5. Evaluation

A customized 'leave-one-out' cross-validation method was adopted, whereby each model was validated with data from a withheld bird, ensuring generalizability of results and reflecting how such models would probably be used in practice. All models were saved for downstream analysis, resulting in a single saved model for each bird for each window width. As an additional benchmark for the immersion classifier, a 'naive' classifier was constructed that simply predicts dive events wherever windows contained a mix of wet and dry samples and non-dive events otherwise (i.e. where windows were all dry or all wet). We applied this model to the same, downsampled 5 min data as the DNNs to ensure a valid performance comparison. The classification accuracy, area under the receiver operating characteristic (ROC) curve (AUC), precision, sensitivity and specificity were calculated for all models, as well as scaled confusion matrices. These metrics were then individually averaged to produce single, cross-validated scores for each window width, which were used to determine the optimum window size for each predictor. The model with the highest accuracy was selected as the best model (highlighted in table 1).

2.6. Georeferencing predictions

The predictions of the best models were georeferenced to gain a deeper understanding of how these can be used to pinpoint foraging locations in practice. To this end, the timestamp at the midpoint of each input window was saved along with its corresponding class during the construction of the rolling window data to enable mapping to the nearest GPS coordinates. GPS data were at the original 30 s resolution for mapping high-resolution acceleration predictions, but for the immersion predictions, to account for the relatively large window widths used to compensate for the data's low temporal resolution (4 m in the best model), GPS data were first collapsed to 5 min resolution to ensure that the window surrounding each GPS coordinate (±2.5 min) contained the majority of several immersion windows rather than just a minority of a handful. Keeping GPS data at 30 s resolution would result in the midpoint of six immersion windows falling within 15 s of—and consequently being mapped to—each GPS coordinate, despite the fact that each of the 4 min immersion windows could also contain up to 18 other GPS coordinates. Using this approach, the class of a specific GPS location was then determined from the proportion of dive predictions in its immediate vicinity. If this proportion exceeded a threshold of 0.5 (50% of predictions within the GPS windows being dives), then the GPS coordinate would qualify as a dive location. This georeferencing method was applied for both predictors, with the threshold set to a default value of 50% of predictions within the corresponding GPS window in each case (the effect of changing this value was not investigated here but could be subject to further inquiry—see §4).

3. Results

Of the 15 sets of GLS/GPS deployed, failure to recover devices or data corruption meant we only successfully extracted data from nine GPS and six GLS loggers. Full data (GPS and GLS) was therefore available for five birds and GPS data for an additional four birds. Figure 2 shows the mapped GPS tracks for all nine GPS-tagged red-footed boobies, with accompanying summary statistics displayed in table 2.

3.1. Acceleration data

Using raw tri-axial acceleration data to predict dive events, we found all leave-one-out cross-validated models scored mean accuracy values exceeding 97%, demonstrating strong predictive power for this modality that generalizes to unseen individuals (table 1). Additionally, sensitivity and specificity remained high across all window widths tested, indicating a general proficiency at detecting both dive events and non-dive events (table 1). The optimal model used a 4 s time window to generate training data (see figure 3 for the resulting confusion matrix, and figure 4 for the distribution of metrics across cross-validation folds for this model), scoring a mean accuracy of 98.22% that varied negligibly across cross-validation folds (figure 4). Mean sensitivity and precision scores for this model were each second highest of those tested (table 1), representing the best combination of minimizing the number of genuine dive sites missed (false negatives) while preventing over-classification by penalizing bogus dive sites (false positives). Models trained only with z-axis acceleration summary statistics performed substantially worse, scoring a mean classification accuracy of 87.15% (sensitivity = 95.6%, specificity = 78.7%), suggesting that diluting the data in this way does preserve some measure of what constitutes a dive, but also drives misclassification of non-dive events.

3.2. Immersion data

The results of models using immersion data to predict dive events showed increased variation and slightly reduced accuracy, ranging from 81.47% to 93.65%, where window widths of 1 and 5 min were used, respectively (table 1). Precision saw the most significant reductions of all metrics compared with acceleration data, dropping by a mean average of nearly 8% (figure 4). While sensitivity scaled roughly in proportion to window width, a negative correlation was observed between window width and precision (and window width and specificity), suggesting that larger window widths capture higher proportions of genuine dive events but also lead to increased misclassification of non-dive events. The optimal window width was the median of those tested (see figure 3 for the resulting confusion matrix, and figure 4 for the distribution of metrics across cross-validation folds for this model). When cross-validating the predictions of this model, the model evaluated using data from unseen bird ch_gps12 scored significantly lower precision (87.01%) than the others (93.51%), suggesting that this bird often exhibited 'dive-looking'

Table 1. The ability of acceleration and immersion to predict seabird dive behaviour. Cross-validated mean classification metrics for each window width tested for each predictor. Using accuracy as a guide, optimal predictions were produced when a 4 s window width was used for acceleration data and a 5 min window width was used for immersion data. Italic values indicate best performance for that predictor/metric.

predictor	window width	accuracy (%)	AUC (%)	precision (%)	sensitivity (%)	specificity (%)
acceleration	2s	98.17	*98.97*	98.31	*97.97*	98.39
	4s	*98.22*	98.70	*98.61*	97.73	**98.69**
	6s	97.95	98.56	98.41	97.36	98.50
	8s	97.98	98.60	98.22	97.63	98.31
	10 s	98.11	98.60	97.98	98.14	98.09
immersion	1 min	81.47	85.12	*92.08*	67.63	*94.53*
	3 min	88	95.45	86.08	93.55	82.60
	5 min	*93.65*	97.16	91.66	95.85	91.51
	7 min	93.41	*97.35*	90.32	*97.01*	89.96
	9 min	93.56	97.23	90.53	*97.01*	90.31

Table 2. Summary statistics for each GPS-tracked bird, with birds from whom GLS data were also successfully recovered highlighted with an asterisk. (One outlier sample was removed from ch_gps03 where the TDR recorded a single depth value of over 10 m.) Dive (depths recorded greater than 0.1 m) and non-dive counts are presented over the whole dataset as recorded by TDR pressure data. Mean dive durations are the period over which the dive was recorded as exceeding 0.1 m—this can result in some dives appearing to have instantaneous duration, as they are only briefly recorded as exceeding the threshold—these dives are recorded as having duration 1.

tag ID	days tracked	total distance travelled (km)	max. distance travelled from the colony (km)	maximum diving depth (m)	mean dive duration (s, min*–max)	dives	non-dives
ch_gps03	5.78	1969	420	3.89	5.96 (1–90)	162	16 555
ch_gps07*	3.91	619	156	2.15	5.76 (1–31)	25	11 300
ch_gps08	5.24	1268	226	2.58	7.99 (1–90)	154	14 979
ch_gps09	4.32	790	86	1.91	8.71 (1–61)	71	12 397
ch_gps10*	4.31	597	81	1.90	10.07 (1–60)	34	12 428
ch_gps11*	3.19	731	182	2.59	8.58 (1–60)	67	9149
ch_gps12*	4.10	1027	234	2.39	2.45 (1–31)	52	11 807
ch_gps13	5.78	1308	231	2.60	13.33 (1–179)	109	16 582
ch_gps16*	4.02	902	157	2.53	9.33 (1–182)	135	11 491

behaviour while not actually foraging (figure 3). A larger number of individuals may highlight if this variation in behaviour is expected or may allow models to better predict dives given a better distribution of training data.

The 5 min naive immersion classifier correctly classified 92.62% of the same data with high sensitivity (92.71%), but lower specificity (88.01%) and precision (88.64%). Electronic supplementary material, figure S1 compares the performance of the DNN immersion classifier and the naive immersion classifier. We note that there are a few 'true' dives that were recorded as fully wet.

3.3. Georeferencing predictions

When mapped to nearest GPS coordinates, the distribution of dive and non-dive predictions showed an imbalance comparable to that of the data prior to downsampling (figure 5). This was due to most dives produced in the rolling window data subsets being generated around single, highly localized events, meaning that any positive predictions corresponding to these samples collapsed back into a small handful of points when georeferenced. Both predictors remained highly proficient at detecting genuine dive and non-dive sites (sensitivity = 96.43% and specificity = 99.16% for acceleration; sensitivity = 96.55% and specificity = 95.23% for immersion—calculated from the confusion matrix statistics displayed in figure 3). Naturally, however, the increase in non-dive sites pushed up the relative incidence rate of true negatives and false positives. Therefore, the negative predictive value of this georeferencing method was near-perfect (99.96% for acceleration, 99.81% for immersion), while the precision was significantly lower (56.25% for acceleration, 51.85% for immersion). When we explore the spatial distribution of errors, false positives seem to generally cluster around true positives and are spread much more sporadically elsewhere (figure 5).

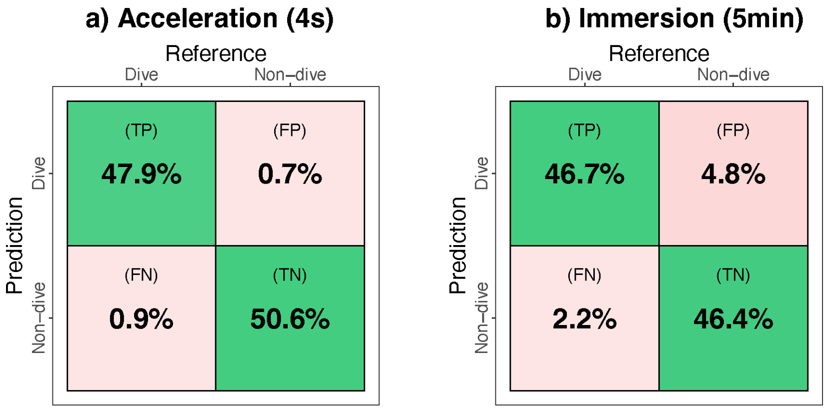

Figure 3. Scaled confusion matrices for the optimal models of both (a) acceleration and (b) immersion data, showing the proportion of total data points in each classification bucket. TP, true positives (correctly predicted dives); FP, false positives (predicted dives that were not dives); FN, false negatives (dives that were not predicted); TN, true negatives (correctly predicted non-dives).

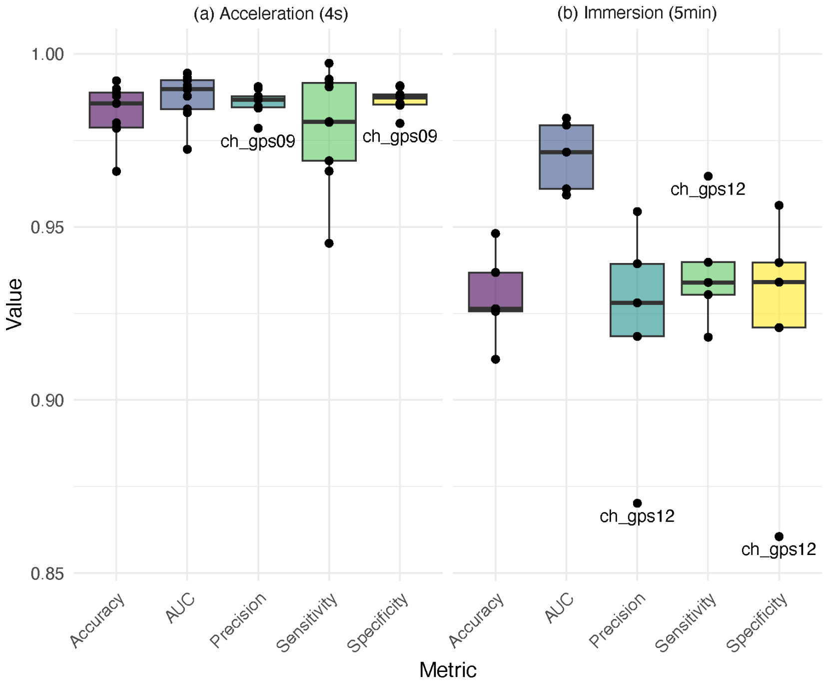

Figure 4. The performance of models using acceleration (a) and immersion (b) data to predict seabird dive events. Plots show classification metrics scored by optimal models for both predictors across folds of the leave-one-out cross-validation process. Each observation corresponds to a bird withheld for model testing. Outliers are labelled with tag ID.

4. Discussion

Combining data from GLS and integrated GPS/TDR/acceleration loggers, we trained a feed-forward DNN to compare the performances of 25 Hz tri-axial acceleration data and 6 s saltwater immersion data as predictors of seabird diving behaviour. Predictions were validated with diving events identified using pressure data. We found that DNNs can predict the dive events within a 4 s window of 25 Hz tri-axial acceleration data with 98.22% accuracy, and within a 5 min window of 6 s immersion data with 93.65% accuracy. Nevertheless, while immersion data could correctly classify similar proportions of genuine diving events to acceleration data, it identified significantly fewer non-dives. It therefore seems plausible that immersion data from GLS devices can be used to reveal diving behaviour for wide-ranging animals throughout their annual cycles; a promising result in a field where large datasets exist and such insights have hitherto been available only for larger birds over short time frames.

Reconciling these dives spatially would be harder with positions inferred from GLS (as the spatial errors are so large), but the temporal resolution (e.g. 5 min windows) would still be high—this would enable exploration of timings and duration of foraging bouts (and approximate locations of these) across much longer-term deployments. It is also important to note that the method of attachment of GLS devices (often leg-mounted) versus GPS/TDR (back or tail-mounted) allows the former device to be deployed for much longer periods, resulting in large datasets for several species.

The specific foraging behaviours of the target species also have important implications for data labelling under our method. For example, we here use a fixed depth threshold of 0.1 m to identify dive events for all birds. However, it is possible that idiosyncrasies in both individual bird behaviour and device sampling could make a single threshold inappropriate for

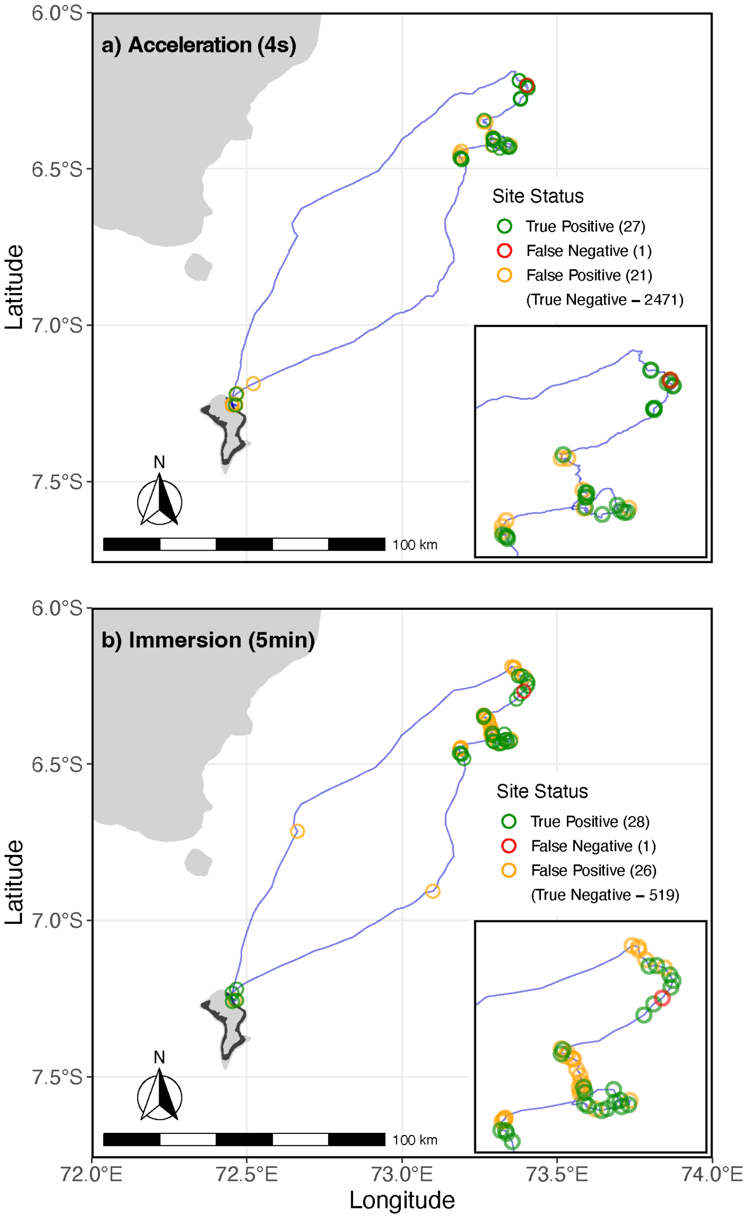

Figure 5. Example of a GPS track with georeferenced model predictions overlaid for (a) 4 s windows of acceleration data; and (b) 5 min windows of immersion data from the reduced data for bird ch_gps07 clusters of overlapping points in the same region. The mismatch of the total number of predictions between the two maps is due to the differing resolutions of the GPS data and the predictors between them. Tracks show an increase in the proportion of false positives compared with the distribution of errors in the predictions prior to mapping, yet largely clustered around true positives and sparse elsewhere.

multiple birds. While our median filtering presents a method to establish a consistent baseline approach to establishing this threshold, differing values for different birds (e.g. using change-point analysis) may improve the labelling accuracy of the data and consequently reduce the errors made by models trained on it. Furthermore, RFBs engage in a variety of diving behaviours when foraging, including surface dives, plunge dives from low altitudes and aerial pursuit to catch flying fish or squid above the surface [50–52]. Their mix of shallow diving behaviours and near-surface aerial pursuit could complicate inferring any distinctions from depth data alone, as, for example, the depth value separating a shallow surface dive from a period merely sitting on the surface with legs submerged may be obscure. Indeed, in the case of immersion data for RFBs, aerial foraging activity will be missed altogether. This may be compounded by unknown shifts in foraging modality outside of the breeding season. Without ground-truth data through the annual cycle, it will be important to make the caveats clear. For species whose interactions with the water during foraging are more distinctive, such as deep-divers or those who engage in an active swim phase (e.g. gannets; [17]), determining labels in the training data should be less problematic, leading to improved model performance. However, the existing models are probably very species-specific, and multi-species studies would help demonstrate and probably improve generalizability.

Model performance could also be improved by using more sophisticated machine learning techniques. For example, as all models evaluated here were only exposed to a small proportion of the available ‘non-dive’ data during training, it is possible that important insights as to what constitutes a ‘non-dive’ sample were missed. Model precision (and, therefore, overall performance) could therefore possibly be enhanced by adopting alternative data balancing methods such as oversampling the minority class (e.g. by generating synthetic data samples using SMOTE; [53]), or adjusting class weights to assign higher misclassification costs to the minority class, thereby accounting for the class imbalance without altering the original data distribution. Additionally, specialized model architectures better suited to capturing temporal dependencies in structured time-series data, such as Long Short-Term Memory networks (LSTMs) and transformers, could extract more meaningful features for identifying nuanced diving behaviours. A comparison of different deep architectures for predicting dives from GPS data highlighted that CNN and U-Net architectures outperform Hidden Markov Models (HMMs), can do so using less data, and transfer well from other species [54].

In line with previous studies [55], our results confirm that tri-axial acceleration data are excellent predictors of diving behaviour in pelagic seabirds. The relatively poor performance of models trained here using only 'diluted' z-axis summary statistics emphasizes the importance of incorporating full tri-axial data. While the devices used here (approx. 18 g) can only currently be deployed on birds of a similar size (greater than 1 kg) for a relatively short period (days to weeks), our findings indicate that smaller, lightweight accelerometers could enable long-term data collection, further enhancing the utility of predictive models for studying seabird foraging ecology. We also note that the much higher temporal frequency of the accelerometers may be additional utility in identifying higher frequency behaviours (e.g. descent/ascent/foraging phases) than the immersion data.

Our results demonstrate the potential of immersion loggers to accurately quantify diving behaviour (here with accuracies of up to 93.65%), only marginally underperforming those trained on tri-axial acceleration data (−4.57%). We also note that the naive classifier performed surprisingly well (92.62%). Its strong performance, as well as its increased over-classification of dives, is probably linked to the specific foraging behaviours of the RFB. Namely, RFBs engage in frequent sea-surface landings during foraging bouts [50], which generate windows with mixed wet and dry readings. These 'mixed' windows, therefore, often either overlap with true dive events, driving true positives, or lie near them, driving false positives. Electronic supplementary material, figure S1 illustrates this pattern, showing that these windows cluster at moderate-to-high wet proportions under the naive model, inflating dive predictions. In contrast, the DNN better distinguishes true dives from transient wet/dry transitions, thus outperforming the naive model overall by reducing false positives and improving specificity (see electronic supplementary material, figure S2). It follows that the DNN offers a more ecologically meaningful classification of dive behaviour by incorporating more nuanced immersion patterns. However, we note that even with this naive model (identifying dives as those locations that are not entirely dry or immersed), it performs surprisingly well at identifying diving locations in the data and may be a useful tool in analysing GLS datasets. However, it is worth considering the relative complexity of the models here—training the DNNs is much more computationally intensive and requires more voluminous data compared with a naive classifier.

Identifying the spatial locations of diving is also complicated by the disparate timescales of the GPS data and immersion data (figure 5). False positives cluster near true positives but are much more spread out elsewhere. If the aim is to identify regions in which foraging behaviours occur, then it can be argued that the occurrence of the former when implementing this georeferencing method is benign. The more isolated false positives, however, pose more of a problem as they have the potential to detrimentally influence mapping efforts. However, it is possible that this can be resolved simply by including more data and/or adjusting the threshold used to classify individual sites.

Our results suggest high classification accuracy and perform well on entirely withheld individuals, but we note that data were only available for a small sample of individuals and should therefore be viewed as a pilot demonstration of this methodology. Some individuals' behaviour appears harder to classify (ch_gps_12, figure 4), and this may be due to different patterns of behaviour, but also note that our low sample size may overestimate the rarity of this behaviour. Classifiers that can accurately predict diving from GLS data alone offer the possibility of identifying diving for the duration of GLS deployment, either throughout the annual cycle (e.g. [56]) or even over multiple years. At present, there are 387 studies comprising over 10 000 tracks using GLS data listed on the global Seabird Tracking Database [57]. It is unclear if any of the associated immersion data often collected alongside the processed tracks have been suitably archived. Our study highlights the potential importance of archiving these data for future studies exploring long-term behavioural patterns. While developing the underlying models requires co-deployment of devices/double-tagged individuals (to learn to predict dives from TDR/GPS-TDR using GLS), these models can then be used across this wealth of data to identify diving during migration and over-wintering. However, understanding how generalizable such models are (e.g. does diving during breeding correspond to diving during non-breeding) would be a valuable next step. A shared database of multi-tagged individuals may help provide critical ground-truth data to develop and validate further models.

By demonstrating that saltwater immersion data from GLS loggers can reliably predict diving behaviour in seabirds, this study highlights new possibilities for enriching the long-term monitoring of foraging behaviours across both historical and future datasets. This approach offers the opportunity to leverage the vast amounts of GLS data already collected, paving the way for more comprehensive and scalable studies in marine ecology.

Ethics. For birds that were captured and handled, methods were reviewed by the Zoological Society of London Ethics Committee and were approved by the British Trust for Ornithology special methods panel. Avian research on Diego Garcia was undertaken under permit number 0000SE19.

Data accessibility. Source code, the analysis and presentation are available on GitHub [58]. Data for the analysis are available at [59]. Please note that in order for the code to execute successfully, the Data/ directory must be in the same directory as the Code/ directory. The required directory structure can be viewed in the README.md file in the GitHub repository.

Supplementary material is available online [60].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. L.S.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft, writing—review and editing; P.C.: data curation, validation, writing—review and editing; R.D.: data curation, supervision, validation, writing—review and editing; M.A.C.N.: data curation, supervision, validation, writing—review and editing; A.T.: data curation, supervision,

validation, writing—review and editing; S.C.V.: data curation, project administration, resources, supervision, validation, writing—review and editing; H.W.: data curation, validation, writing—review and editing; R.F.: conceptualization, data curation, investigation, methodology, project administration, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was funded by the Bertarelli Foundation and contributed to the Bertarelli Programme in Marine Science. R.F. and M.A.C.N. were supported by Research England.

Acknowledgements. We are very grateful for the logistical support of Environment Officer Harri Morrall.

References

1. Hazen EL, Abrahms B, Brodie S, Carroll G, Jacox MG, Savoca MS, Scales KL, Sydeman WJ, Bograd SJ. 2019 Marine top predators as climate and ecosystem sentinels. *Front. Ecol. Environ*. **17**, 565–574. (doi:10.1002/fee.2125)
2. Ciancio JE, Yorio P, Buratti C, Colombo GÁ, Frere E. 2021 Isotopic niche plasticity in a marine top predator as indicator of a large marine ecosystem food web status. *Ecol. Indic*. **126**, 107687. (doi:10.1016/j.ecolind.2021.107687)
3. Carignan V, Villard MA. 2002 Selecting indicator species to monitor ecological integrity: a review. *Environ. Monit. Assess*. **78**, 45–61. (doi:10.1023/a:1016136723584)
4. Parsons M, Mitchell I, Butler A, Ratcliffe N, Frederiksen M, Foster S, Reid JB. 2008 Seabirds as indicators of the marine environment. *ICES J. Mar. Sci*. **65**, 1520–1526. (doi:10.1093/icesjms/fsn155)
5. Thaxter CB, Lascelles B, Sugar K, Cook A, Roos S, Bolton M, Langston RHW, Burton NHK. 2012 Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. *Biol. Conserv*. **156**, 53–61. (doi:10.1016/j.biocon.2011.12.009)
6. Bost C, Le Maho Y. 1993 Seabirds as bio-indicators of changing marine ecosystems: new perspectives. *Acta Oecologica Int. J. Ecol*. **14**, 463–470.
7. Einoder LD. 2009 A review of the use of seabirds as indicators in fisheries and ecosystem management. *Fish. Res*. **95**, 6–13. (doi:10.1016/j.fishres.2008.09.024)
8. Mallory ML, Robinson SA, Hebert CE, Forbes MR. 2010 Seabirds as indicators of aquatic ecosystem conditions: a case for gathering multiple proxies of seabird health. *Mar. Pollut. Bull*. **60**, 7–12. (doi:10.1016/j.marpolbul.2009.08.024)
9. Guilford T, Meade J, Willis J, Phillips RA, Boyle D, Roberts S, Collett M, Freeman R, Perrins CM. 2009 Migration and stopover in a small pelagic seabird, the Manx shearwater *Puffinus puffinus*: insights from machine learning. *Proc. R. Soc. B Biol. Sci*. **276**, 1215–1223. (doi:10.1098/rspb.2008.1577)
10. Bograd SJ, Block B, Costa D, Godley BJ. 2010 Biologging technologies: new tools for conservation. Introduction. *Endang. Species. Res*. **10**, 1–7. (doi:10.3354/esr00269)
11. Grecian WJ *et al*. 2016 Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biol. Lett*. **12**, 20160024. (doi:10.1098/rsbl.2016.0024)
12. Davies TE *et al*. 2021 Multispecies tracking reveals a major seabird hotspot in the North Atlantic. *Conserv. Lett*. **14**, e12824. (doi:10.1111/conl.12824)
13. Trevail A, Wood H, Carr P, Dunn RE, Nicoll MA, Votier SC, Freeman R. 2023 Multi-colony tracking reveals segregation in foraging range, space use, and timing in a tropical seabird. *Mar. Ecol. Prog. Ser*. **724**, 155–165. (doi:10.3354/meps14479)
14. Trevail AM *et al*. 2023 Tracking seabird migration in the tropical Indian Ocean reveals basin-scale conservation need. *Curr. Biol*. **33**, 5247–5256. (doi:10.1016/j.cub.2023.10.060)
15. Watanabe YY, Papastamatiou YP. 2023 Biologging and biotelemetry: tools for understanding the lives and environments of marine animals. *Annu. Rev. Anim. Biosci*. **11**, 247–267. (doi:10.1146/annurev-animal-050322-073657)
16. Guilford TC, Meade J, Freeman R, Biro D, Evans T, Bonadonna F, Boyle D, Roberts S, Perrins CM. 2008 GPS tracking of the foraging movements of Manx shearwaters *Puffinus puffinus* breeding on Skomer Island, Wales. *Ibis* **150**, 462–473. (doi:10.1111/j.1474-919x.2008.00805.x)
17. Cox SL, Miller PI, Embling CB, Scales KL, Bicknell AWJ, Hosegood PJ, Morgan G, Ingram SN, Votier SC. 2016 Seabird diving behaviour reveals the functional significance of shelf-sea fronts as foraging hotspots. *R. Soc. Open Sci*. **3**, 160317. (doi:10.1098/rsos.160317)
18. Shoji A, Elliott K, Fayet A, Boyle D, Perrins C, Guilford T. 2015 Foraging behaviour of sympatric razorbills and puffins. *Mar. Ecol. Prog. Ser*. **520**, 257–267. (doi:10.3354/meps11080)
19. Dean B, Freeman R, Kirk H, Leonard K, Phillips RA, Perrins CM, Guilford T. 2013 Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *J. R. Soc. Interface* **10**, 20120570. (doi:10.1098/rsif.2012.0570)
20. Dunn RE, Freeman R, Nicoll MA, Ramsden J, Trevail AM, Wood H, Votier SC. 2024 From route to dive: multi-scale habitat selection in a foraging tropical seabird. *Mar. Biol*. **171**, 124. (doi:10.1007/s00227-024-04445-y)
21. Barron DG, Brawn JD, Weatherhead PJ. 2010 Meta-analysis of transmitter effects on avian behaviour and ecology. *Methods Ecol. Evol*. **1**, 180–187. (doi:10.1111/j.2041-210x.2010.00013.x)
22. Phillips RA, Xavier JC, Croxall JP. 2003 Effects of satellite transmitters on albatrosses and petrels. *Auk* **120**, 1082–1090. (doi:10.2307/4090279)
23. Jackson S, Wilson RP. 2002 The potential costs of flipper-bands to penguins. *Funct. Ecol*. **16**, 141–148. (doi:10.1046/j.0269-8463.2001.00606.x)
24. Bodey TW, Cleasby IR, Bell F, Parr N, Schultz A, Votier SC, Bearhop S. 2018 A phylogenetically controlled meta-analysis of biologging device effects on birds: deleterious effects and a call for more standardized reporting of study data. *Methods Ecol. Evol.* **9**, 946–955. (doi:10.1111/2041-210X.12934)
25. Clewley GD, Cook A, Davies JG, Humphreys EM, O'Hanlon NJ, Weston E, Boulinier T, Ponchon A. 2021 Acute impacts from Teflon harnesses used to fit biologging devices to black-legged kittiwakes *Rissa tridactyla*. *Ringing Migr*. **36**, 69–77. (doi:10.1080/03078698.2022.2151065)
26. Lopez SL, Clewley GD, Johnston DT, Daunt F, Wilson JM, O'Hanlon NJ, Masden E. 2024 Reduced breeding success in great black-backed gulls (*Larus marinus*) due to harness-mounted GPS device. *Ibis* **166**, 69–81. (doi:10.1111/ibi.13247)
27. Gillies N *et al*. 2020 Short-term behavioural impact contrasts with long-term fitness consequences of biologging in a long-lived seabird. *Sci. Rep*. **10**, 1–10. (doi:10.1038/s41598-020-72199-w)
28. Newton I. 2010 *The migration ecology of birds*. London, UK: Elsevier.
29. Åkesson S, Klaassen R, Holmgren J, Fox JW, Hedenström A. 2012 Migration routes and strategies in a highly aerial migrant, the common swift *Apus apus*, revealed by light-level geolocators. *PLoS One* **7**, e41195. (doi:10.1371/journal.pone.0041195)
30. Halpin LR *et al*. 2021 Double-tagging scores of seabirds reveals that light-level geolocator accuracy is limited by species idiosyncrasies and equatorial solar profiles. *Methods Ecol. Evol*. **12**, 2243–2255. (doi:10.1111/2041-210x.13698)
31. Phillips R, Silk J, Croxall J, Afanasyev V, Briggs D. 2004 Accuracy of geolocation estimates for flying seabirds. *Mar. Ecol. Prog. Ser*. **266**, 265–272. (doi:10.3354/meps266265)
32. Jonsen ID, Flemming JM, Myers RA. 2005 Robust state–space modeling of animal movement data. *Ecology* **86**, 2874–2880. (doi:10.1890/04-1852)

33. Patterson TA, Basson M, Bravington MV, Gunn JS. 2009 Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *J. Anim. Ecol*. **78**, 1113–1123. (doi:10.1111/j.1365-2656.2009.01583.x)
34. Breed GA, Costa DP, Jonsen ID, Robinson PW, Mills-Flemming J. 2012 State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecol. Model*. **235–236**, 49–58. (doi:10.1016/j.ecolmodel.2012.03.021)
35. Nathan R, Spiegel O, Fortmann-Roe S, Harel R, Wikelski M, Getz WM. 2012 Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *J. Exp. Biol*. **215**, 986–996. (doi:10.1242/jeb.058602)
36. Martiskainen P, Järvinen M, Skön JP, Tiirikainen J, Kolehmainen M, Mononen J. 2009 Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Appl. Anim. Behav. Sci*. **119**, 32–38. (doi:10.1016/j.applanim.2009.03.005)
37. Grünewälder S, Broekhuis F, Macdonald DW, Wilson AM, McNutt JW, 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. (doi:10.1371/journal.pone.0049120)
38. Carroll G, Slip D, Jonsen I, Harcourt R. 2014 Supervised accelerometry analysis can identify prey capture by penguins at sea. *J. Exp. Biol*. **217**, 4295–4302. (doi:10.1242/jeb.113076)
39. Bishop CM. 1995 *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.
40. LeCun Y, Bengio Y, Hinton G. 2015 Deep learning. *Nature* **521**, 436–444. (doi:10.1038/nature14539)
41. Min S, Lee B, Yoon S. 2017 Deep learning in bioinformatics. *Briefings Bioinform*. **18**, 851–869. (doi:10.1093/bib/bbw068)
42. Christin S, Hervet É, Lecomte N. 2019 Applications for deep learning in ecology. *Methods Ecol. Evol*. **10**, 1632–1644. (doi:10.1101/334854)
43. Borowiec ML, Frandsen PB, Dikow RB, McKeeken A, Valentini G, White AE. 2022 Deep learning as a tool for ecology and evolution. *Methods Ecol. Evol*. **13**, 1640–1660. (doi:10.32942/OSF.IO/NT3AS)
44. Browning E, Bolton M, Owen E, Shoji A, Guilford T, Freeman R. 2018 Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Methods Ecol. Evol*. **9**, 681–692. (doi:10.1111/2041-210x.12926)
45. Chakravarty P, Cozzi G, Ozgul A, Aminian K. 2019 A novel biomechanical approach for animal behaviour recognition using accelerometers. *Methods Ecol. Evol*. **10**, 802–814. (doi:10.1111/2041-210x.13172)
46. Wilson RP *et al*. 2020 Estimates for energy expenditure in free-living animals using acceleration proxies: a reappraisal. *J. Anim. Ecol*. **89**, 161–172. (doi:10.1111/1365-2656.13040)
47. Nelson S, Nelson B. 1978 *The sulidae: gannets and boobies (no.154)*. New York, NY, USA: Oxford University Press.
48. Austin RE, De Pascalis F, Votier SC, Haakonsson J, Arnould JPY, Ebanks-Petrie G, Newton J, Harvey J, Green JA. 2021 Interspecific and intraspecific foraging differentiation of neighbouring tropical seabirds. *Mov. Ecol*. **9**, 27. (doi:10.1186/s40462-021-00251-z)
49. R Core Team. 2021 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.
50. Weimerskirch H, Le Corre M, Ropert-Coudert Y, Kato A, Marsac F. 2005 The three-dimensional flight of red-footed boobies: adaptations to foraging in a tropical environment? *Proc. R. Soc. B Biol. Sci*. **272**, 53–61. (doi:10.1098/rspb.2004.2918)
51. Diamond A. 1974 The red-footed booby on Aldabra Atoll, Indian Ocean. *Ardea* **62**, 196–218.
52. Pitman R, Ballance L. 1993 Booby prey-capturing behavior in the eastern tropical Pacific. In *Abstract. Pacific Seabird Group 20th Annual Meeting*. Seattle, WA: Pacific Seabird Group.
53. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002 SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res*. **16**, 321–357. (doi:10.1613/jair.953)
54. Roy A, Lanco Bertrand S, Fablet R. 2022 Deep inference of seabird dives from GPS-only records: performance and generalization properties. *PLoS Comput. Biol*. **18**, e1009890. (doi:10.1371/journal.pcbi.1009890)
55. Ma R, Kazama K, Yasutake YC, Kazama MT, Tsukamoto S, Watanuki Y. 2022 Dynamic body acceleration improves mapping of at-sea foraging behavior in black-tailed gulls *Larus crassirostris*. *J. Avian Biol*. **2022**, e02869. (doi:10.1111/jav.02869)
56. Freeman R, Dean B, Kirk H, Leonard K, Phillips RA, Perrins CM, Guilford T. 2013 Predictive ethoinformatics reveals the complex migratory behaviour of a pelagic seabird, the Manx shearwater. *J. R. Soc. Interface* **10**, 20130279. (doi:10.1098/rsif.2013.0279)
57. Birdlife International. 2024 Seabird tracking database: a platform for seabird researchers to share their tracking data with the research and conservation communities. See https://www.seabirdtracking.org/ (accessed 5 February 2024).
58. Swaby L, Freeman R. 2026 robinfreeman/seabirds_dive_predict: seabirds dive prediction (v1.0.0). Zenodo (doi:10.5281/zenodo.19220945)
59. Freeman R, Dunn R, Nicoll M, Trevail A, Votier S, Wood H. 2025 Data for: Deep neural networks to predict foraging behaviour: salt-water immersion data can accurately predict diving in seabirds. Harvard Dataverse, V1 (doi:10.7910/DVN/GHWM5U)
60. Swaby L, Carr P, Dunn R, Nicoll MAC, Trevail A, Votier SC *et al*. 2026 Supplementary material from: Deep neural networks to predict foraging behaviour: salt-water immersion data can accurately predict diving in seabirds. Figshare. (doi:10.6084/m9.figshare.c.8479352)