

Enriching historical biologging datasets on seabirds using deep neural networks: A transformer-based approach to infer energy expenditure proxy from GPS and environmental data

Supporting Information

1 Trajectory plots for all the GPS data used in the study

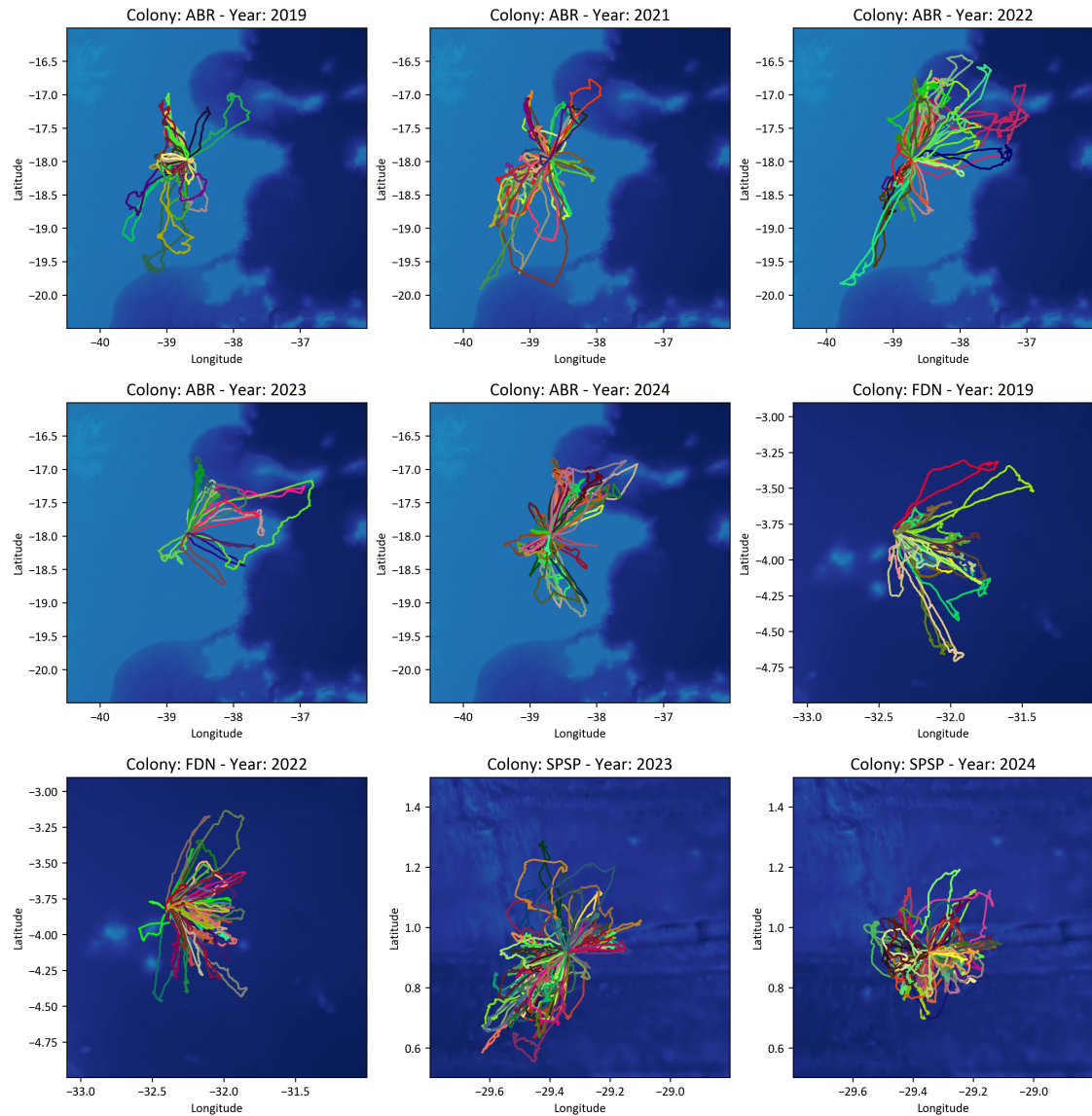


Figure 1: **Plot of GPS trajectories over bathymetry for every data used in this study, separated by fieldwork.** ABR: Abrolhos, FDN: Fernando de Noronha, SPSP: São Pedro e São Paulo.

2 On cuboid attention's complexity

In self-attention, the computational complexity is $\mathcal{O}(N^2)$, where N denotes the size of the input. In the case of inputs of shape (H, W) , with $H \times W$ the spatial grid, this corresponds to $N = H \times W$, and the complexity becomes $\mathcal{O}(H^2W^2)$. Such complexity requires significant computing resources, often making it impossible to train the model without adequate hardware. Cuboid attention is a variant of the attention mechanism proposed to address computational complexity. Cuboid attention works by decomposing the input data into non-overlapping cuboids of size (b_H, b_W) , and applying self-attention independently within each cuboid before merging them to recover the output data. This approach reduces the self-attention complexity and, in our setting, reduced the overall complexity by at least a factor of 100 (exact calculation is difficult because (b_H, b_W) change in each layer).

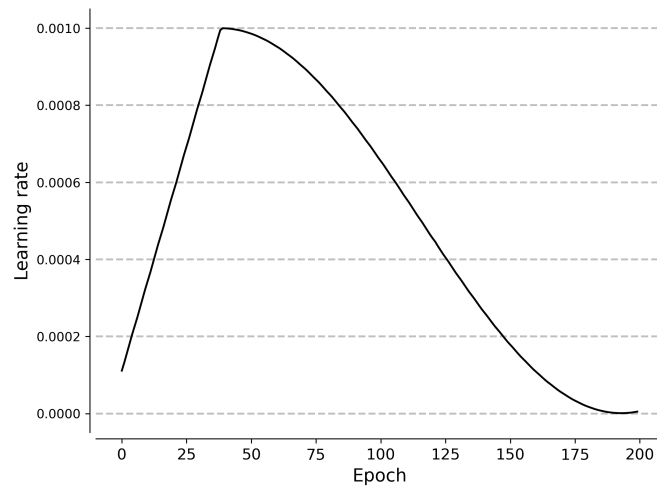
$$O\left(\frac{H}{b_H} \frac{W}{b_W} (b_H b_W)^2\right) \approx O(WH(b_H b_W))$$

3 Training parameters used for ODBAFormer

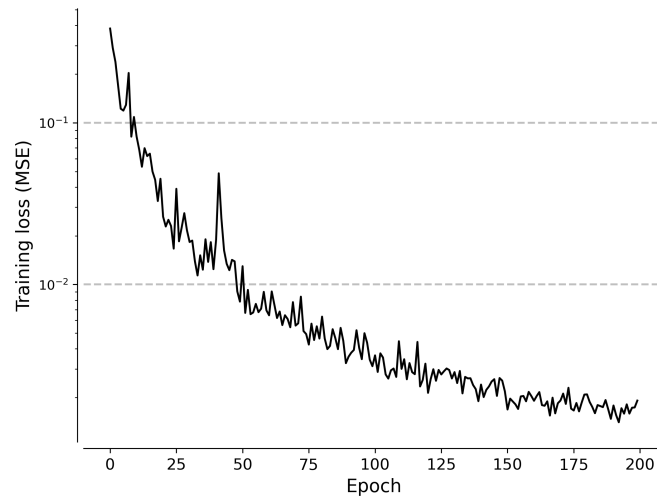
Table 1: **Summary of model hyperparameters and training configuration.** This table reports the main hyperparameters used for training, including dataset characteristics, architectural details, optimization strategy, and cuboid attention settings.

Hyperparameter	Value
Input dataset size (singles-species model)	298
Input dataset size (two-species model)	524
Evaluation dataset size (Masked booby)	23
Evaluation dataset size (Brown Booby)	36
Input shape (H, W)	(100, 100)
Number of channels C	1 – 4 depending on number of covariates
Epochs	200
Batch size	8
Learning rate	0.001
Learning rate scheduler	cosine
Loss function	MSE
Optimizer	AdamW
β_1	0.9
β_2	0.999
Parameter number	9.6 millions
Model size	38.358 Mb
Number of attention heads	4
Cuboid decomposition strategy	Along each axis (T, H, W)
Number of global vectors	8

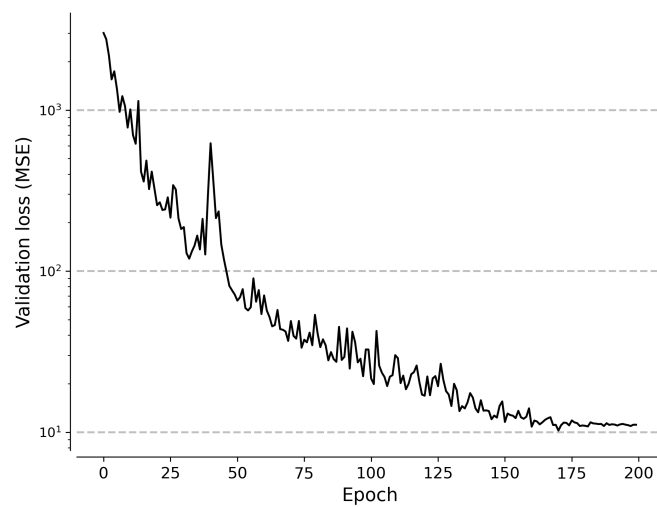
4 Loss and learning rate curves



(a) Learning rate evolution.



(b) Training loss (MSE) evolution.



(c) Validation loss (MSE) evolution.

Figure 2: **Training dynamics for the single-species dataset with bathymetry as covariate.**

5 HMM parameters

5.1 Masked Booby evaluation dataset, Fernando de Noronha 2019

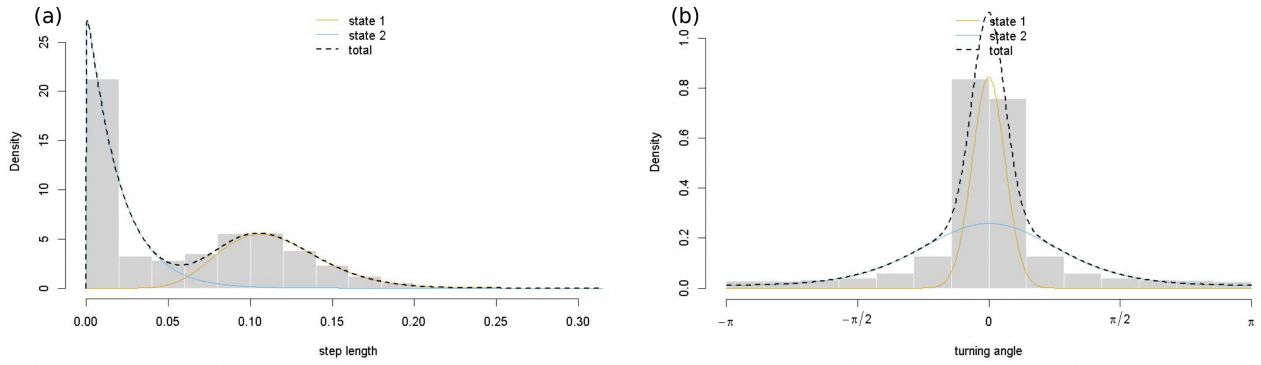


Figure 3: **Histograms and HMM fitted distributions for the Masked Boobies evaluation dataset. (a)** step length; **(b)** turning angle. State 1 is the 'traveling' state and state 2 is the 'foraging' state.

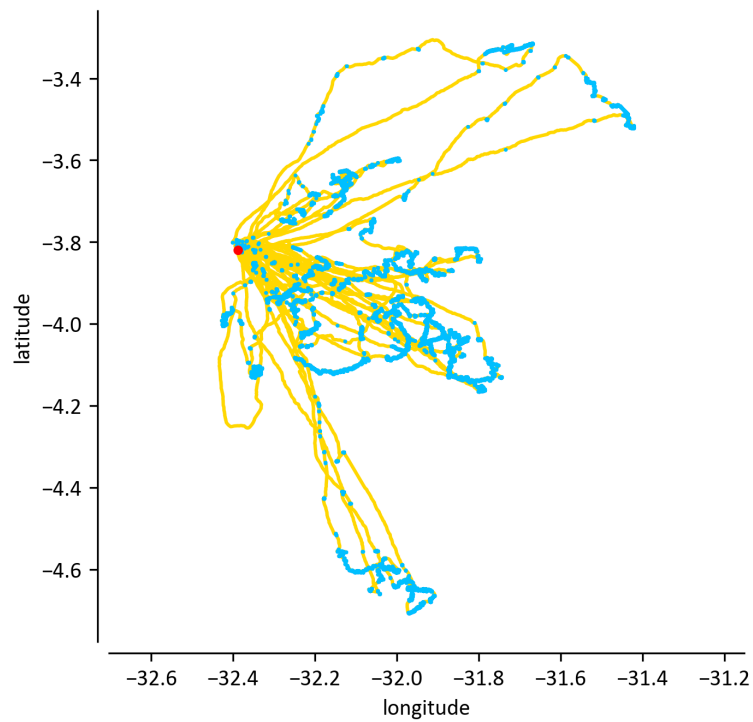


Figure 4: **Trajectories with segmented states for the Masked Boobies evaluation dataset. Yellow:** 'traveling' state, **blue:** 'foraging' state, **red dot:** colony position.

Table 2: **HMM parameters values for the Masked Booby evaluation dataset.** State 1 is the 'traveling' state and state 2 is the 'foraging' state. Initial parameters, parameters of the fitted distributions, and state transition probabilities are reported.

Parameter	Value
Number of states	2
Step length distribution	Gamma
Turning angles distribution	Von Mises
Initial step length means (state 1 ; state 2)	(0.1 ; 0.01)
Initial step length standard deviations (state 1 ; state 2)	(0.5 ; 0.01)
Initial step length zero mass (state 1 ; state 2)	(0.01 ; 0.01)
Initial angle mean (state 1 ; state 2)	(0 ; 0)
Initial angle concentration (state 1 ; state 2)	(0.1 ; 0.5)
Step length mean state 1	0.1145
Step length standard deviation state 1	0.0313
Step length zero mass state 1	2.23e-09
Angle mean state 1	-0.0023
Angle concentration state 1	26.19
Step length mean state 2	0.0193
Step length standard deviation state 2	0.0190
Step length zero mass state 2	0.0030
Angle mean state 2	0.0020
Angle concentration state 2	1.56
State 1 initial probability	0.254
State 2 initial probability	0.746
Transition probability 1→1	0.946
Transition probability 1→2	0.054
Transition probability 2→2	0.962
Transition probability 2→1	0.038

5.2 Brown Booby evaluation dataset, Abrolhos 2019

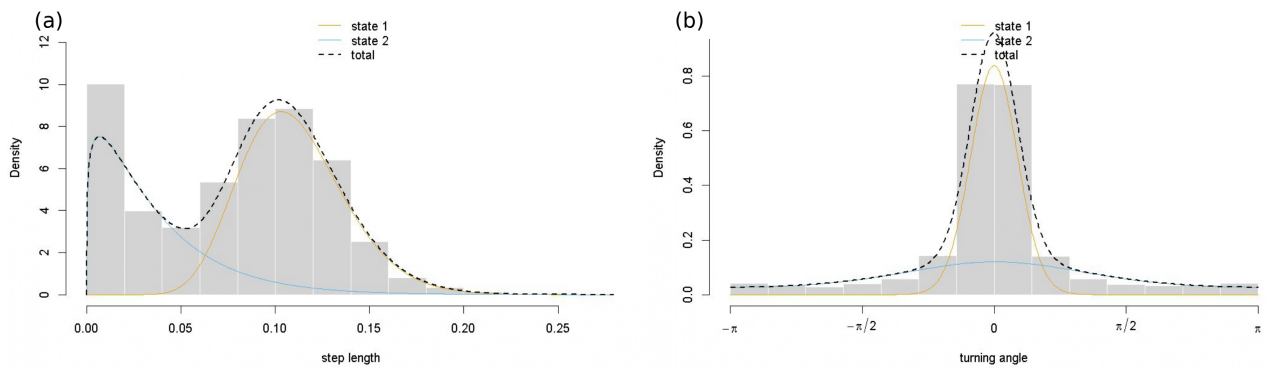


Figure 5: **Histograms and HMM fitted distributions for the Brown Booby evaluation dataset.** (a) step length; (b) turning angle. State 1 is the 'traveling' state and state 2 is the 'foraging' state.

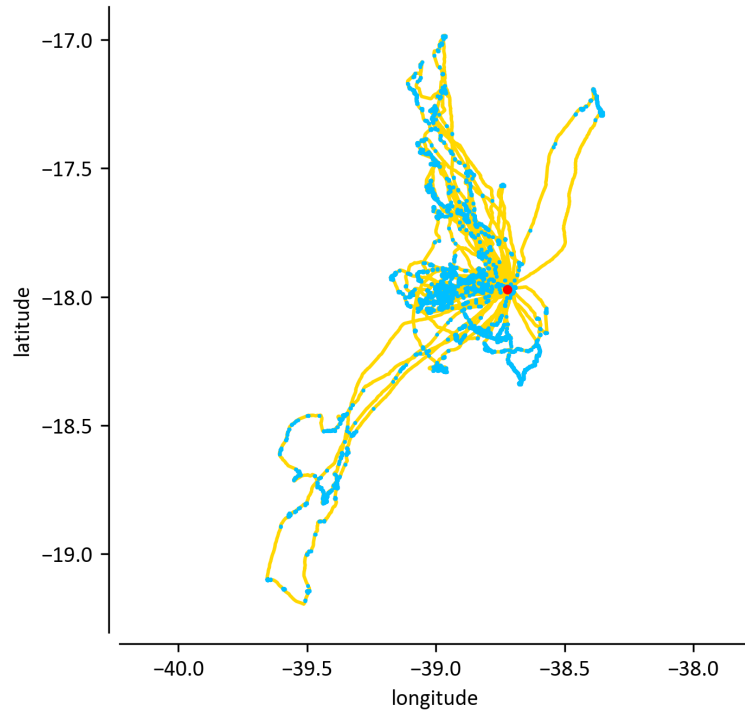


Figure 6: **Trajectories with segmented states for the Brown Booby evaluation dataset.** Yellow: 'traveling' state, blue: 'foraging' state, red dot: colony position.

Table 3: **HMM parameters values for the Brown Booby evaluation dataset.** State 1 is the 'traveling' state and state 2 is the 'foraging' state. Initial parameters, parameters of the fitted distributions, and state transition probabilities are reported.

Parameter	Value
Number of states	2
Step length distribution	Gamma
Turning angles distribution	Von Mises
Initial step length means (state 1 ; state 2)	(0.1 ; 0.01)
Initial step length standard deviations (state 1 ; state 2)	(0.5 ; 0.01)
Initial step length zero mass (state 1 ; state 2)	(0.01 ; 0.01)
Initial angle mean (state 1 ; state 2)	(0 ; 0)
Initial angle concentration (state 1 ; state 2)	(0.1 ; 0.5)
Step length mean state 1	0.1102
Step length standard deviation state 1	0.2078
Step length zero mass state 1	4.76e-10
Angle mean state 1	-0.0008
Angle concentration state 1	12.94
Step length mean state 2	0.0363
Step length standard deviation state 2	0.0328
Step length zero mass state 2	0.1298
Angle mean state 2	0.0131
Angle concentration state 2	0.7458
State 1 initial probability	0.227
State 2 initial probability	0.773
Transition probability 1→1	0.967
Transition probability 1→2	0.047
Transition probability 2→2	0.953
Transition probability 2→1	0.032

6 Experiment with distance to colony as a covariate

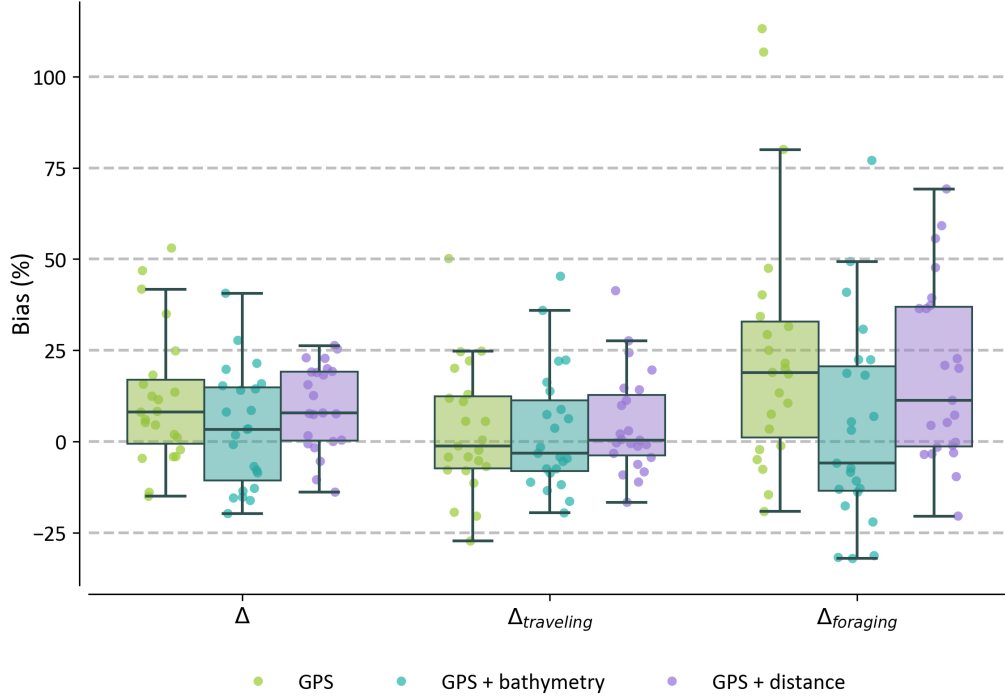


Figure 7: Distributions of ODBA prediction performances for each covariate combination, including distance to colony.

Table 4: Comparison of ODBAFormer performance for different covariate combinations as inputs, including distance to colony. Metrics include relative error in cumulated ODBA (Δ), relative error in cumulated ODBA per inferred behavioral state ($\Delta_{traveling}$, $\Delta_{foraging}$), and Earth Mover’s Distance (EMD) between inferred and real ODBA values distributions. Values are reported as mean $\pm$ standard deviation. Bold values indicate the best-performing covariate combination for each metric.

Evaluation metric	GPS only	GPS + bathymetry	GPS + distance
Δ (%)	11.5 $\pm$ 18.2	3.3 $\pm$ 16.1	9.6 $\pm$ 11.8
$\Delta_{traveling}$ (%)	3.0 $\pm$ 17.6	2.9 $\pm$ 16.7	4.6 $\pm$ 13.9
$\Delta_{foraging}$ (%)	24.8 $\pm$ 34.7	3.8 $\pm$ 27.8	18.7 $\pm$ 24.8
EMD	34.8 $\pm$ 17.7	30.3 $\pm$ 11.5	28.4 $\pm$ 8.7

7 Inter-individual variability in inference results

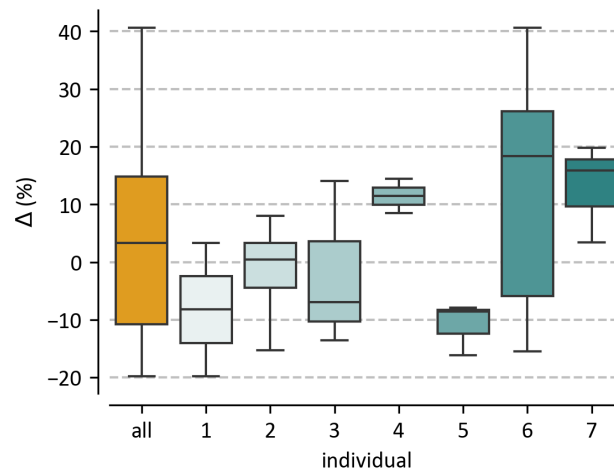


Figure 8: **Distributions of ODBA prediction performances for all the Masked Booby evaluation dataset and partitioned by individual.** Covariate used: bathymetry. 23 trips total; individual 1: 2 trips, individual 2: 4 trips, individual 3: 3 trips, individual 4: 2 trips, individual 5: 3 trips, individual 6: 6 trips, individual 7: 3 trips. For all trips considered independently, Q25, median, and Q75 of the errors are -10.7 %, 3.3 %, and 14.8 %, respectively, whereas the averages across individuals are -3.9 %, 3.2 %, and 7.6 %.