



## OPEN The role of individual variability on the predictive performance of machine learning applied to large bio-logging datasets

Marianna Chimienti<sup>1</sup>✉, Akiko Kato<sup>1</sup>, Olivia Hicks<sup>1,2</sup>, Frédéric Angelier<sup>1</sup>, Michaël Beaulieu<sup>3</sup>, Jazel Ouled-Cheikh<sup>4,5</sup>, Coline Marciau<sup>1,6</sup>, Thierry Raclot<sup>7</sup>, Meagan Tucker<sup>8</sup>, Danuta Maria Wisniewska<sup>9</sup>, André Chiaradia<sup>8,10</sup> & Yan Ropert-Coudert<sup>1,10</sup>

Animal-borne tagging (bio-logging) generates large and complex datasets. In particular, accelerometer tags, which provide information on behaviour and energy expenditure of wild animals, produce high-resolution multi-dimensional data, and can be challenging to analyse. We tested the performance of commonly used artificial intelligence tools on datasets of increasing volume and dimensionality. By collecting bio-logging data across several sampling seasons, datasets are inherently characterized by inter-individual variability. Such information should be considered when predicting behaviour. We integrated both unsupervised and supervised machine learning approaches to predict behaviours in two penguin species. The classified behaviours obtained from the unsupervised approach *Expectation Maximisation* were used to train the supervised approach *Random Forest*. We assessed agreement between the approaches, the performance of *Random Forest* on unknown data and the implications for the calculation of energy expenditure. Consideration of behavioural variability resulted in high agreement (> 80%) in behavioural classifications and minimal differences in energy expenditure estimates. However, some outliers with < 70% of agreement, highlighted how behaviours characterized by signal similarity are confused. We advise the broad bio-logging community, approaching these large datasets, to be cautious when upscaling predictions, as this might lead to less accurate estimates of behaviour and energy expenditure.

Extremely large, complex and multidimensional data sets, difficult to handle and analyze with common methodological approaches (so-called big data), are collected and produced by several scientific and industrial fields and require tools to facilitate efficient management, visualization, analysis, validation and interpretation<sup>1</sup>. In biology, big data have emerged with studies on gene and protein sequences, gene expression, evolutionary trees, microscopy images, 3D structures and interaction networks<sup>2-5</sup>. Within global change biology, ecology and remote sensing, the recent use of big data is improving our understanding of interactions between environmental changes and biological systems<sup>6</sup>. In animal movement ecology, using animal-borne tags (bio-logging) can quickly accumulate large amount of data. The quantity of collected data is large due to high sampling resolution, the number of sensors (e.g., a bio-logging tag can include location and pressure sensors, magnetometer, accelerometer and environmental sensors), length of deployment and number of species considered (e.g., more than one year of accelerometer data recordings, decades of data tracking)<sup>7-11</sup>. Tri-axial accelerometers, for example, provide

<sup>1</sup>Centre d'Etudes Biologiques de Chizé, UMR7372 CNRS - La Rochelle Université, 405 Route de Prissé La Charrière, 79360 Villiers-en-Bois, France. <sup>2</sup>British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. <sup>3</sup>German Oceanographic Museum, Stralsund, Germany. <sup>4</sup>Institut de Recerca de la Biodiversitat (IRBio) and Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals (BEECA), Facultat de Biologia, Universitat de Barcelona., Av. Diagonal 643, 08028 Barcelona, Spain. <sup>5</sup>Institut de Ciències del Mar (ICM-CSIC), Departament de Recursos Marins Renovables, Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain. <sup>6</sup>Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS 7001, Australia. <sup>7</sup>Institut Pluridisciplinaire Hubert Curien, UMR7178, CNRS-Université de Strasbourg, Strasbourg, France. <sup>8</sup>Conservation Department, Phillip Island Nature Parks, Cowes, VIC, Australia. <sup>9</sup>Sound Communication and Behaviour Group, Department of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. <sup>10</sup>These authors jointly supervised this work: André Chiaradia and Yan Ropert-Coudert. ✉email: marianna.chimienti@cebc.cnrs.fr

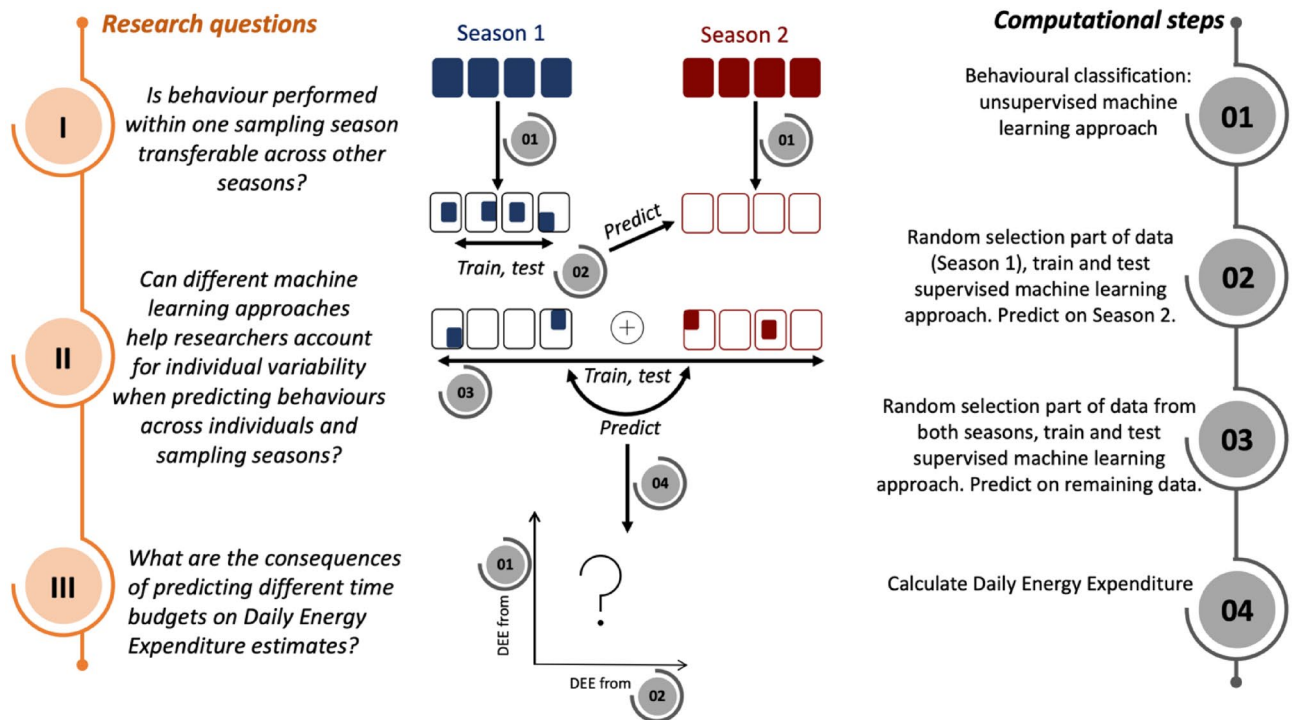
3-dimensions (called surge, sway and heave) of high-resolution acceleration data to obtain measurements of effort and relative energy expenditure related to different behaviours and activities performed by individual animals<sup>12</sup>. Several variables are calculated from raw acceleration data indicating body orientation and acceleration, such as body pitch, roll, dynamic acceleration for each axis, Vectorial Dynamic Body Acceleration (VeDBA), (see<sup>12,13</sup> and our Method section for overview). A common proxy of energy expenditure is activity-specific Dynamic Body Acceleration (DBA)<sup>12</sup>. When validated with both direct and indirect measurements of energetic expenditure (heart rates, isotope elimination—Doubly Labelled Water (DLW),—respiratory chambers), DBA can provide estimates of energy consumption (e.g., daily energy expenditure, DEE hereafter) across habitats, generating “energy landscapes”<sup>14–16</sup>. Energetic expenditure attributed to landscape features is currently used to quantify behavioural and fitness responses to habitat availability and human-derived risks at both individual and population level<sup>17–19</sup>.

To extract relevant information from large and complex datasets, the use of Artificial Intelligence tools (e.g. machine learning and deep learning approaches) has steadily grown in importance over the last decade to the point that they are applied in nearly every research field. In the bio-logging field, machine learning has been applied across marine and terrestrial systems to classify types of behaviour performed by animals equipped with fast-sampling accelerometers. Both unsupervised and supervised machine learning approaches, have been used on both data collected in the wild as well as in captivity<sup>20–25</sup>. Usually, unsupervised approaches are used when validation data are not available, in contrast to a supervised approach making use of pre-labelled known behavioural activities recorded. While unsupervised approaches (e.g., Expectation Maximisation, k-means) independently detect behaviours and allow for the detection of unknown behaviours<sup>20,26</sup>, supervised approaches (e.g., Random Forest, Support Vector Machine) are fast and reliable on known behaviours<sup>27,28</sup>. Whilst both approaches have their strengths, they also have their individual and shared weaknesses and can be complementary. Unsupervised machine learning approaches require manual post-labelling, which does not scale well with the exponential increase of data volume. By using pre-labelled training datasets to predict on unknown data, supervised approaches can automatically predict on large volume of data, but are, instead, limited by the information acquired via the training. Integrating both unsupervised and supervised machine learning approaches can make predictions of behaviours across large datasets robust and feasible<sup>29</sup>.

Behavioural validation on elusive species, challenging to monitor in the wild, might be a lacking, or possible only on some behaviours and for short time periods, producing datasets which can be too small to enable robust learning. Moreover, behaviours recorded/characterized in captivity may not be suitable to validate data collected in the wild that are likely to be noisier and/or contain unexpected signals due to the combination of individual variability in movement mechanics, environmental variability, and properties of the medium in which individuals are moving. To address some of these issues, researchers have provided more detailed validation data, recording variation in accelerometer signals of animals walking, for example, on different types of terrain<sup>30</sup>. Depending on the research question tackled, the resolution and type of data collected, number and types of behaviours needed to estimate (few coarse vs many high-resolution behaviours), machine learning approaches might estimate time-activity budgets differently, and hence, lead to differences in DEE values, if DBA values are grouped by activity type (as done in<sup>31,32</sup>, for example). Therefore, in the past few years, this field has taken a more critical look at the available approaches and begun to assess which methods are most appropriate in different scenarios<sup>13,28,33</sup>.

As researchers continue to collect data, accelerometer datasets, obtained from large numbers of individuals across years, are inherently characterised by individual and environmental variability<sup>34,35</sup>. The ideal setting for obtaining behavioural validation and data acquisition over a large sample size and sampling years to train machine learning models might not always be feasible in the wild, making inference of behavioural activities and their changes even more challenging. Both unsupervised and supervised machine learning approaches need to deal with individual and environmental variabilities, and the effect of individual variability on the performance of machine learning algorithms is largely unknown<sup>24</sup>. The large volume of complex bio-logging data calls for continuous refinements in analytical methods to make upscaling processes (such as inferring behaviours over larger proportions of novel data compared with training and testing data) reliable, robust and as efficient as possible. Here we explore the feasibility of combining different unsupervised and supervised machine learning methods to address the following issues (Fig. 1): (I) Assuming acceleration data are sampled from individuals belonging to the same population, can behaviours detected across individuals be transferred from one sampling period to another sampling period? (II) is the combination of unsupervised and supervised machine learning a viable approach to include individual variability when predicting behaviours across individuals and sampling seasons? As these approaches might result in differing activity budgets, we also address a third question: (III) what are the consequences of varying budgets for the calculation of DEE?

To explore these broad research questions, we use accelerometer data collected over two breeding seasons from two species of penguins foraging in different environments as a case study: Adélie penguins (*Pygoscelis adeliae*) and little penguins (*Eudyptula minor*). The Adélie penguin is the most abundant seabird species in Antarctica, with around 4 million pairs around the continent<sup>36</sup>. Adélie penguins have high underwater flexibility, frequently diving to shallow depths of 5–10 m, likely to feed just under the ice, but also reaching maximum depths of 115–120 m<sup>37</sup>. This species mainly feed on Antarctic and ice krill (*Euphausia superba* and *E. crystallorophias*) and Antarctic silverfish, *Pleuragramma antarctica*<sup>38</sup>. The Little penguin is distributed in southern Australia and New Zealand and its overall population is estimated to be under 500,000 breeding pairs<sup>39</sup>. Little penguins are high trophic level generalist predators feeding in dynamic coastal environments<sup>40,41</sup> mainly on sardine (*Sardinops sagax*) and anchovy (*Engraulis australis*), but also red cod (*Pseudophycis bachus*), barracouta (*Thyrssites atun*) and jack mackerel (*Trachurus declivis*)<sup>42</sup>. Individuals forage in the shallower part of the water column, diving on average within 20 m but also reaching 57 m<sup>43,44</sup>. Both species’ are sensitive to human disturbance and climatic variability<sup>45,46</sup>. The feeding ecology and energetic requirements of these species at fine scales can help us to predict their vulnerability to environmental change. To answer the above-mentioned research questions applied to these two penguin species, we follow four main computational steps (Fig. 1). (1) We first use an unsupervised



**Figure 1.** Conceptual overview of research questions and approach followed in this study. Both unsupervised and supervised machine learning approaches are applied on large accelerometer datasets to detect and predict behavioural activities. We quantify the behavioural variation included in the training datasets used for the supervised approach and test their predictive performance. We finally validate DEE estimates resulted from the behavioural activity budgets from the two approaches based on energetic validation obtained on Adélie penguin (*Pygoscelis adeliae*) from Hicks et al.<sup>32</sup>.

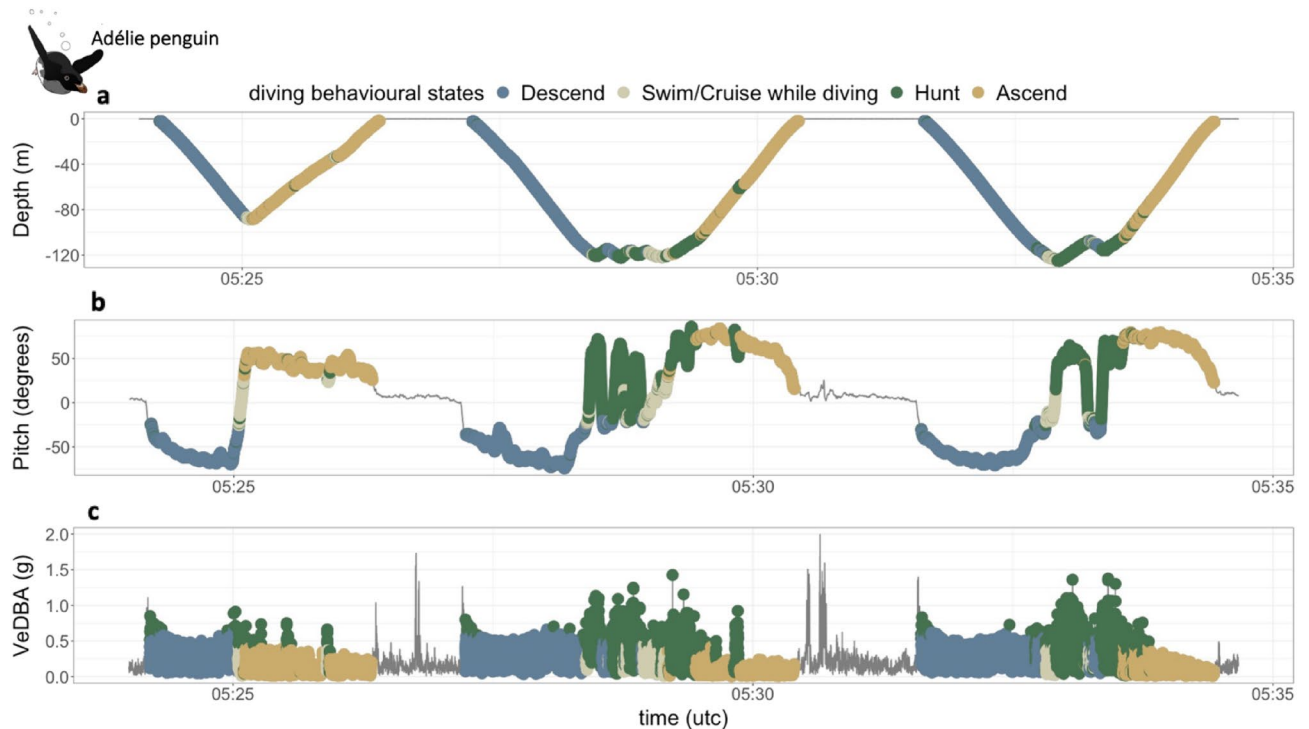
machine learning approach to quantify the individual variability in behaviour across the datasets within seasons. (2 and 3) By randomly selecting parts of the datasets, we then include such variability in training datasets fed to a supervised approach and test its predictive performance and agreement with the unsupervised approach. For Adélie penguins, for which energetic validation is available<sup>32</sup>, we compare DEE estimates (4), resulting from the behavioural activity budgets estimated by both unsupervised and supervised machine learning approaches.

## Results

The two studied species feed within the water column, hence temperature-depth data were also recorded along with acceleration data and processed (see Methods: *Data Preparation*). The same set of variables were used for both species: VeDBA, pitch, the standard deviation of raw heave at 2 s and change in depth for detecting diving behaviours; VeDBA, pitch, the standard deviation of raw heave at 10 s and the standard deviation of roll at 30 s for detecting subsurface and land/ice behaviours (SI Appendix Table S1). The unsupervised machine learning algorithm, *Expectation Maximisation* (EM), was run first (Fig. 1), followed by the supervised machine learning algorithm *Random Forest*.

**Unsupervised machine learning approach.** Twelve behavioural classes were identified for Adélie penguins. Behaviours such as “descend”, “ascend”, “hunt” and “swim/cruise” were identified from the diving subsets, describing activities performed while diving (Fig. 2). For *Season 1*, we identified an additional swimming behaviour, similar to “swim/cruise”, named “swim/cruise type 2” in 5 foraging trips (out of 107) (SI Appendix Fig. S1). The same additional swimming behaviour was also found in 3 foraging trips (out of 53) in *Season 2*. On land/ice, we identified behaviours such as “walking”, “standing”, “lying down/toboggan” and “preen/high flap on land/ice” (SI Appendix Fig. S2). Finally at the sub-surface/air interface, we identified “slow swim”, “swim/porpoise”, while moving horizontally at the sub-surface, and “preen/high flap on water” behaviours (SI Appendix Figs. S3 and S4).

Ten behavioural classes were identified for little penguins. As for diving Adélie penguins, we identified “descend”, “ascend”, “hunt” and “swim/cruise” (Fig. 3). For *Season 2*, we additionally identified the swimming behaviour “swim/cruise type 2” in one foraging trip (out of 39). While at the sub-surface/air interface we identified “slow swim” and “swim/porpoise” while moving horizontally at the sub-surface and “preen/high flap on water” behaviours plus one additional resting behaviour not found in Adélie penguins, termed “still” and characterized by the animals floating on the water surface without any detectable activity other than a change in posture (SI Appendix Fig. S5). For both species “swim/cruise type 2” was included in the “swim/cruise” general



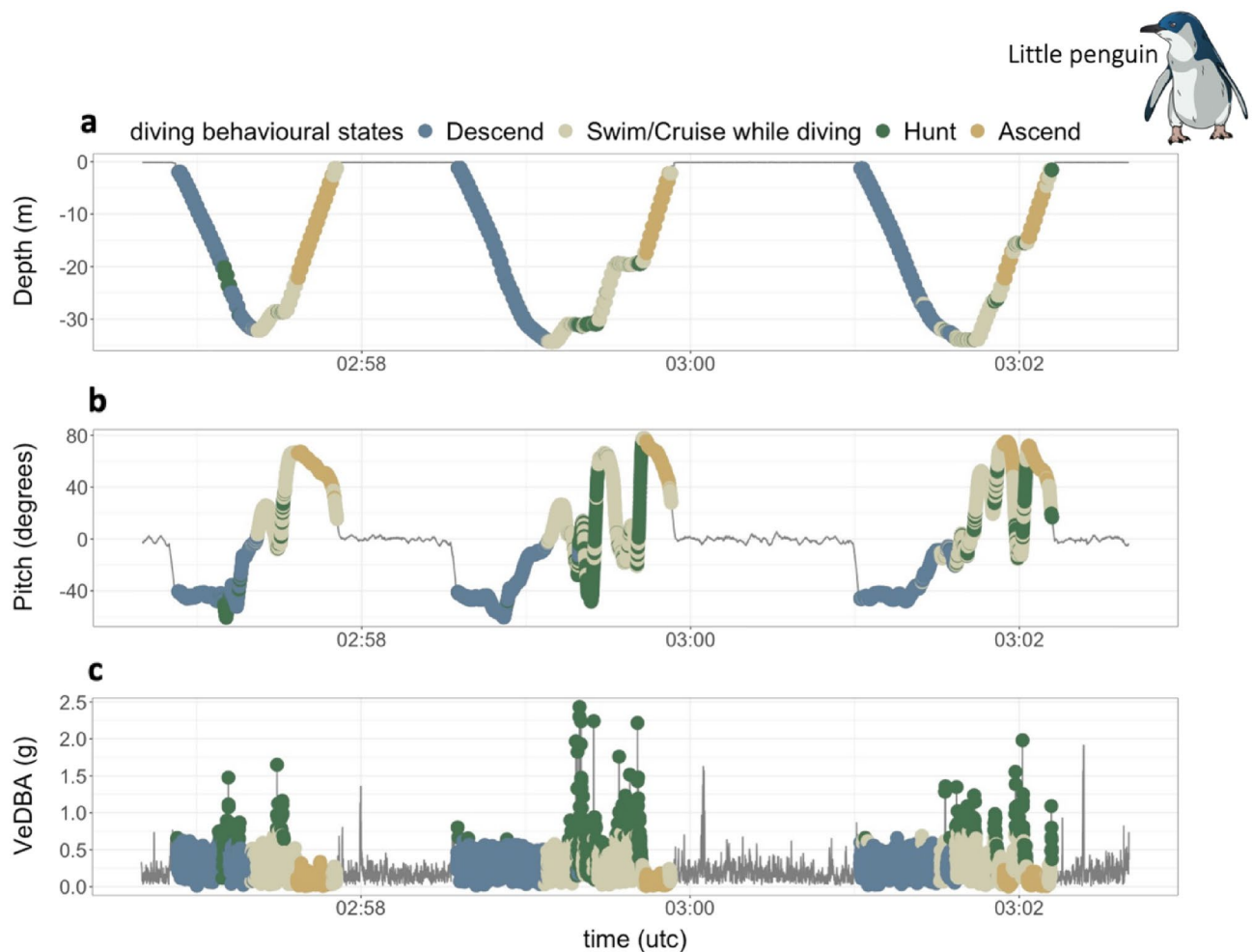
**Figure 2.** Example of behavioural classification in Adélie penguins (*Pygoscelis adeliae*) while diving. (a) diving profiles, (b) Pitch (degrees) indicating body posture while diving, (c) Vectorial Dynamic Body Acceleration (VeDBA, g) indicating overall acceleration while moving underwater. Colours indicate the behavioural states identified while diving.

behaviour for the *Random Forest* (RF) models, because of its similar biological meaning to “swim/cruise” and because it was identified in very few individuals (see Methods for more details).

**Supervised machine learning approach.** Repeatability scores for the training datasets from both species generally showed a lower inter-individuals’ variation in VeDBA than in pitch (Fig. 4). Including both sampling seasons within the training datasets resulted in higher variation in pitch for some behaviours and mainly for Adélie penguins, as “lie down/toboggan” and “stand” performed on land/ice, or “ascend” and “descend” (Fig. 4). The Out-Of-Bag (OOB) error estimated by the RF algorithm (measuring the prediction error of the model) showed a minimal decrease with increasing number of trees for both species (SI Appendix Table S2). Based on the minimal decrease, the model with 1000 trees was selected for both species to avoid any potential overfitting. Pitch, depth, change in depth, the standard deviation of raw heave at 2 s, the standard deviation of the roll at 30 s and the standard deviation of heave at 10 s were the six most important variables for Adélie penguins, with VeDBA taking the seventh position (SI Appendix Fig. S6). Depth, the standard deviation of the roll at 30 s, the standard deviation of heave at 2 s, the change in depth, VeDBA and pitch were the six most important variables for little penguins (SI Appendix Fig. S7). For Adélie penguins, behaviours such as “walk”, “swim/porpoise”, “slow surface swim” and “preen/flap on the water” were characterised by a higher degree of confusion compared to “ascend” and “descend”, for example. Hence observations originally labelled as “walk” could be confused as “stand” or “preen/flap on land” by the RF. The degree of confusion across behaviours was also confirmed with the Receiver Operating Characteristic curve (ROC), with curves closer to the top left corner indicating a better performance (SI Appendix Table S3 and Fig. S8). Similarly, for little penguins, behaviours such as “swim/cruise” while diving, “swim/porpoise” and “still” had a higher degree of confusion, and could have been labelled as other behaviours and displayed a lower ROC (SI Appendix Tables S4 and Fig. S9).

The overall accuracy (hereafter agreement between the two methods, as proportion of identical behavioral prediction estimated by the two approaches within the given dataset) was high, above 0.8 (Fig. 5). However, a few outliers indicated low agreement on some individuals and a slight drop in the level of agreement was observed between the part of datasets used for training and the part used for predicting, so unknown to the RF. Specifically, for Adélie penguins, the agreements between the methods EM and RF were  $0.85 \pm 0.06$  for the model trained on *Season 1* only, and  $0.81 \pm 0.05$  for the prediction on *Season 2*. For the model trained using data from both seasons, the agreement levels were  $0.87 \pm 0.06$  and  $0.88 \pm 0.02$  for the part of the datasets respectively from *Season 1* and *Season 2* used for training; and  $0.81 \pm 0.07$  and  $0.83 \pm 0.07$  for the part of the datasets used for predicting (Fig. 5).

Similarly, for little penguins, the agreement between the methods were  $0.86 \pm 0.09$  for the model trained on *Season 1* only and  $0.81 \pm 0.08$  for the prediction on *Season 2*. For the model trained using data from both seasons, the agreement levels were  $0.85 \pm 0.09$  and  $0.87 \pm 0.05$  for the part of the datasets respectively from *Season 1* and *Season 2* used for training, and  $0.80 \pm 0.1$  and  $0.84 \pm 0.08$  for the part of the datasets used for predicting (Fig. 5).



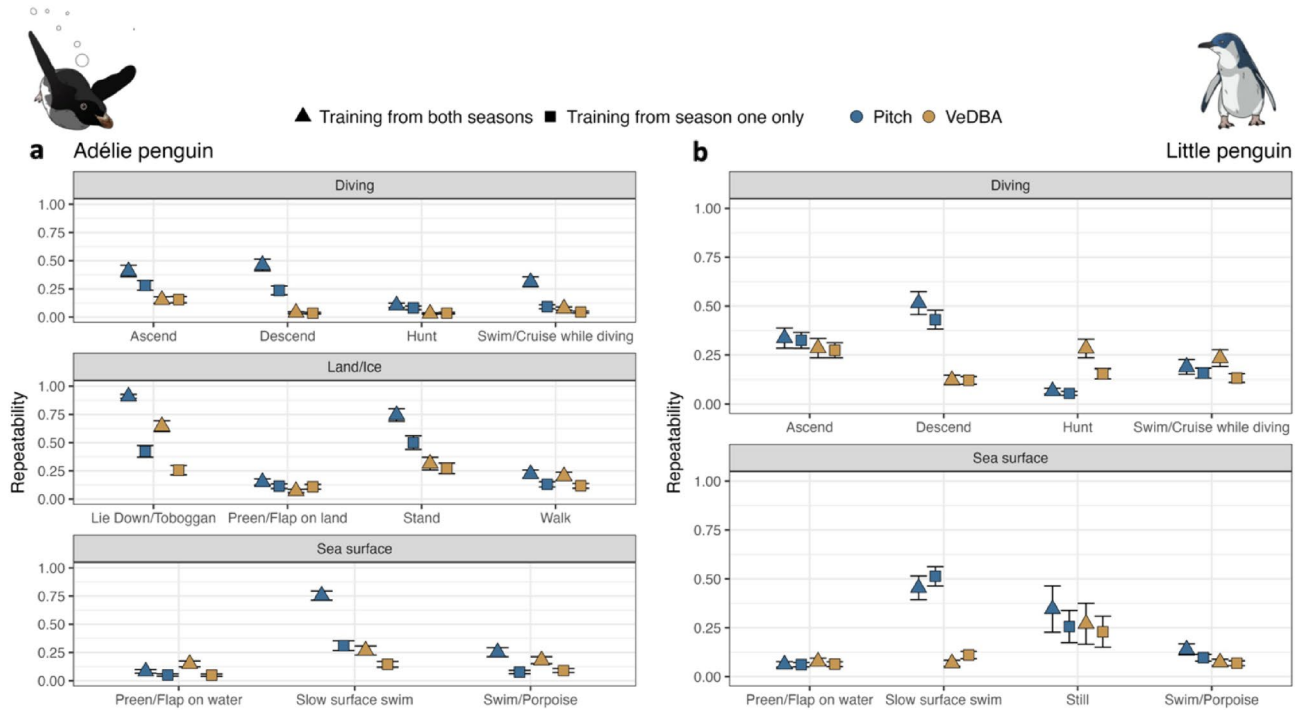
**Figure 3.** Example of behavioural classification in Little penguins (*Eudyptula minor*) while diving. (a) diving profiles, (b) Pitch (degrees) indicating body posture while diving, (c) Vectorial Dynamic Body Acceleration (VeDBA, g) indicating overall acceleration while moving underwater. Colours indicate the behavioural states identified while diving.

### Effect of differences in behavioural classification on time-budgets and estimates of energy expenditure.

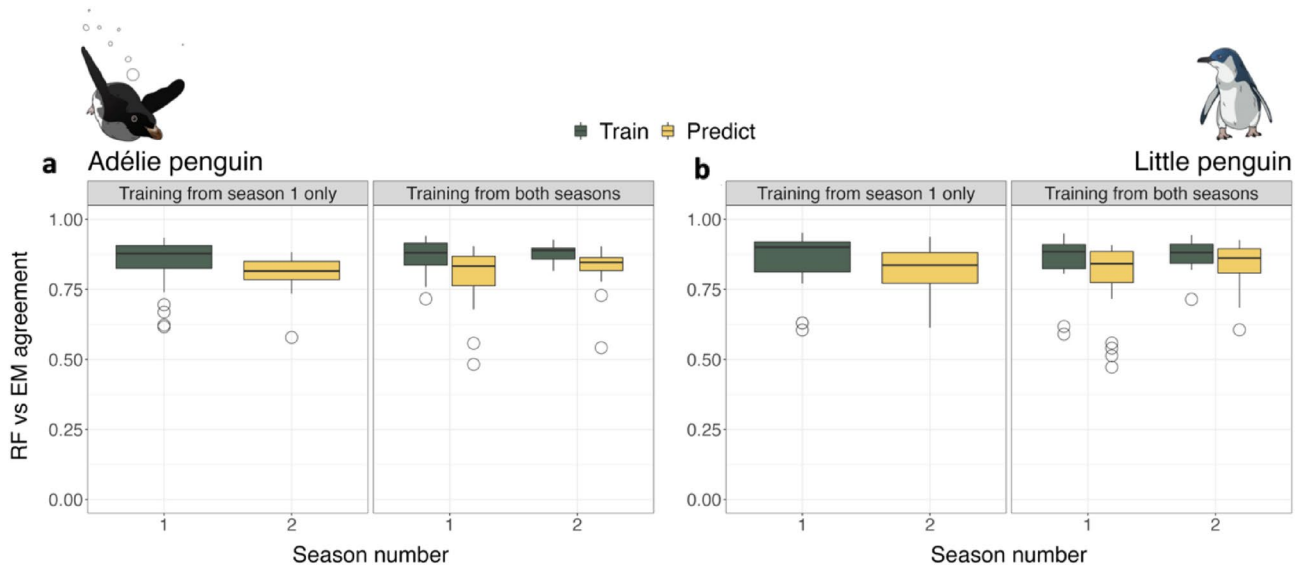
As predicted by the confusion matrices and overall agreements, the differences in behavioural classification by the two machine learning approaches, albeit small, gave rise to differences in fine-scale activity budgets (SI Appendix Figs. S10–S15). For both species, depending on the type of training used (from *Season 1* only or both seasons), some predicted budgets as “slow surface swim”, “still” and “stand” could vary more than others like “lie down/toboggan”, “ascend” and “descend” (SI Appendix Figs. S10–S15). For Adélie penguin, energy calculations were possible for 89 trips (38 individuals performing 1 to 3 trips each) as sex was not recorded for some individuals. Trip duration was  $24.2 \pm 6.6$  h. Resulting behavioural classification was grouped in the three coarse behavioural components used in Hicks et al.<sup>32</sup>, “Preen/Flap on water”, “Water” and “Land/Ice”, showing that the overall proportions of time spent within the components and resulting DEEs estimated with the two methods were similar (Fig. 6). However, both differences in proportions of time spent in the water and preening on the water surface affected the relationship between estimates by the two methods ( $p$  value < 0.001, Fig. 6 and SI Appendix Table S5). The model with a difference in the proportion of time spent preening on the water as covariate had the lowest AIC (adjusted R-squared: 0.98, Table S5). It showed that, while DEE estimates by the two methods are similar for the same individual, the DEE estimated by the EM method could be slightly higher than the one estimated by the RF method (Fig. 6). As consequence, total energy expended per trip followed the same trend, being lower when estimated by the RF compared to the EM (Table S5).

### Discussion

Our study revealed that behavioural classification obtained from big data recorded with accelerometers can be transferred across individuals within the same population and upscaled. Using a large sample size and including behavioural variability within training datasets resulted in a high level of accuracy (> 80%) between the supervised and unsupervised approaches. However, a decrease in accuracy is expected when predicting on unknown data using the supervised machine learning approach *Random Forest*. The low predictive performance has been

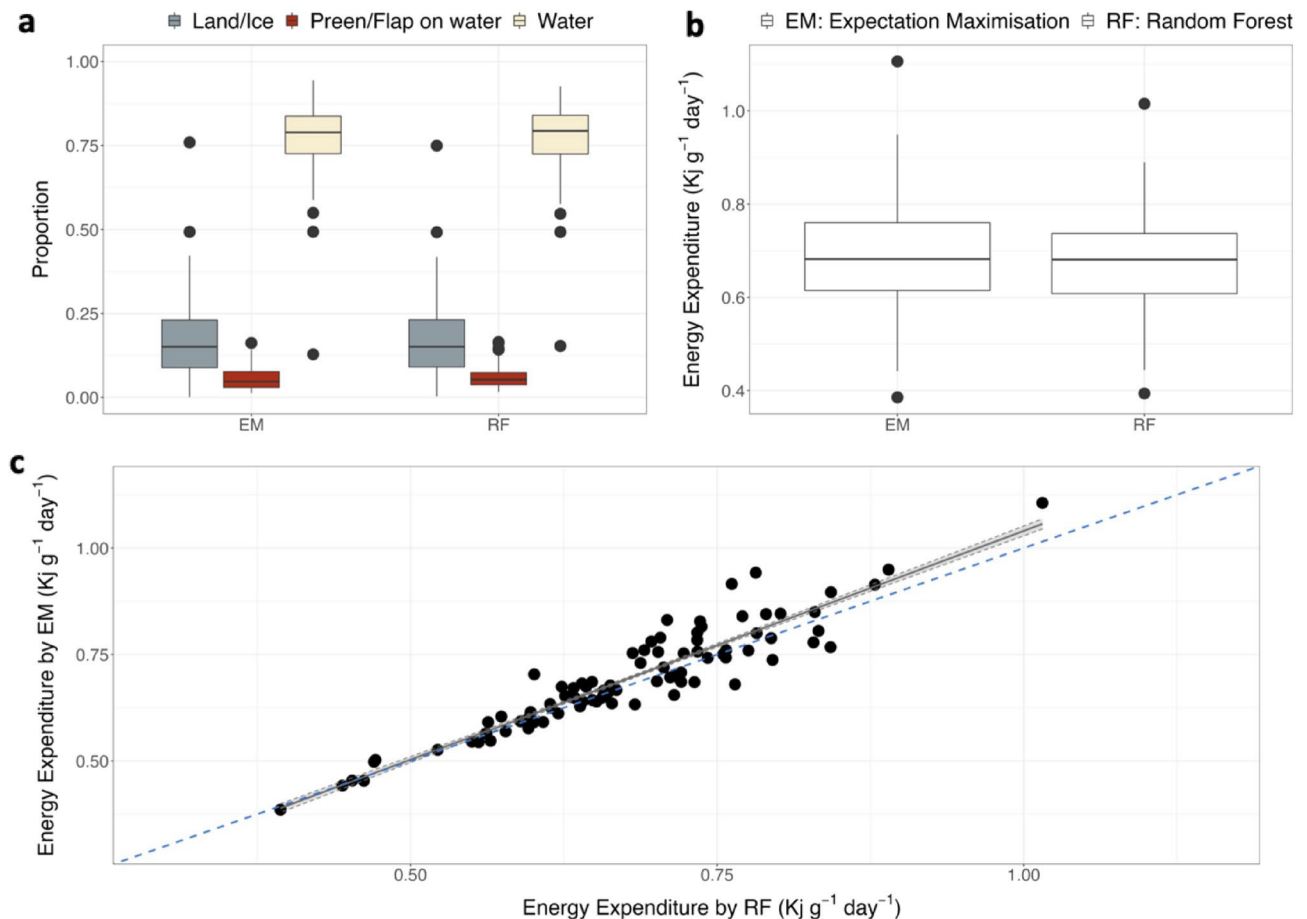


**Figure 4.** Repeatability scores. Values are calculated for pitch (blue) and VeDBA (yellow) on training datasets obtained from *Season 1* (squares) and both seasons (triangles) for both species, Adélie penguin (*Pygoscelis adeliae*) (a) and little penguin (*Eudyptula minor*) (b).



**Figure 5.** Agreement between machine learning approaches. Overall agreement between behavioural classification performed by the unsupervised machine learning algorithm *Expectation Maximisation* (EM) and one returned by the supervised machine learning algorithm *Random Forest* (RF), ran on data collected during foraging trips of Adélie penguins (*Pygoscelis adeliae*) (a), and little penguins (*Eudyptula minor*) (b).

suggested to lie within individual intraspecific differences<sup>47,48</sup>. Our reduced accuracy of 3–6%, between the models with different training datasets, can be associated to reduced sample size used in the model including individuals from both seasons. This drop, is, however, small compared to previously published values (30–35%), resulting from our overall larger sample size<sup>48</sup>. Few individuals with low predictive performance (around 50% of agreement between approaches) could be identified in our case, highlighting that some behaviours were classified differently by the two methods. Hence, individual variation was not completely accounted for within the training sets. Lower predictive performance could be associated, for example, to differences in diving angles during the descent phase, leading one algorithm to associate this posture to a “swim/cruise” behaviour rather than “descend” as in the other. We advise the broad scientific community approaching the large datasets collected with



**Figure 6.** Daily Energy Expenditure estimated using results from two machine learning approaches. Energy Expenditure obtained accounting for the activity budgets predicted by unsupervised machine learning algorithm *Expectation Maximisation* (EM) and the supervised machine learning algorithm *Random Forest* (RF) for Adélie penguin (*Pygoscelis adeliae*). (a) activity budgets used for calculating energy expenditure as in Hicks et al.2020, (b) boxplots of the distributions of the resulting Energy Expenditure, (c) comparison and regression between the Energy Expenditure calculated with the two approaches.

accelerometer data to be cautious when I) using training datasets obtained from few individuals and II) predicting behaviours over large numbers of individuals. Comparison between unsupervised and supervised approaches can help detect individual variability within the individuals sampled. Increasing number of individuals, data points randomly assigned to the training datasets, as well as number of variables could improve model accuracy but at the cost of making models increasingly slow and memory hungry. Implementing feedback mechanisms and further parallelisation of machine learning tasks would be a step towards more efficient management of these increasingly large datasets.

One of the main aims in ecology is to understand the behavior and physiology of, and environmental conditions experienced by, animals as they move through and interact with their environment. The use of bio-logging technology has allowed researchers to observe and quantify behaviours of species difficult to monitor otherwise. It is, however, difficult to obtain behavioural validation on species' behaviours and variations. Although camera recorders are beginning to be widely used to validate behaviours, they usually collect data for short periods of time and on a few individuals, still not capturing inter-individual variability. Moreover, for the specific case of underwater images, data are affected by light and water turbidity and are still difficult to process automatically<sup>49</sup>. The use of deep learning techniques is making it easier and faster to extract information from images<sup>50</sup>. The variables used in our models have been validated for predicting hunting behaviour in similar species<sup>51</sup>.

Signal overlap across similar behaviours (such as subsurface swimming behaviours in our case) and class imbalance (as some behaviours might be under-represented within the training datasets), might confuse<sup>28,52</sup> and lead to disagreement between approaches used. Information on behaviours collected from deployments using camera loggers combined with accelerometer tags could help clarify and deal with confusion between some behavioural classes. Penguins are elusive species, moving in different habitats (land, ice, underwater) and dynamic environmental conditions. Furthermore, the two species considered in this study are marine predators foraging on lower trophic levels (e.g. feeding on fish and krill). Their behavioural variation as a result of encountered environmental conditions (e.g. prey availability) makes them sentinels of climate and ecosystem changes<sup>53</sup> emphasizing the need to monitor their behaviour and understand how individuals modify it. Adélie penguins

showed more inter-variability in behaviour than little penguins. However, for both species, predictions from models trained on data from both seasons showed similar agreement scores to the predictions made by models trained on data from a single season. Species, types of behaviour performed, and environmental conditions can shape the degree of variability recorded<sup>54–58</sup>. Changes in sea-ice conditions (as in extent and characteristics, for example), might lead Adélie penguins to move differently<sup>59</sup> and perhaps show high variability in behaviour. On a broader perspective, body conditions, age and changes in prey type and pursuing tactics, for example, can lead to changes in underwater movements, thereby affecting individual fitness (survival and reproduction)<sup>59–61</sup>. By implementing machine learning approaches able to analyse the vast amount of data collected with bio-logging devices recording high-resolution movements, it would be possible to start assessing the effect of environmental variability on individuals' behaviours and energetics.

In bio-logging and movement ecology, analyses on behavioural variation linked to environmental conditions are carried out primarily on location data, which currently constitute the most extensive animal movement dataset in terms of the number of individuals and species tracked<sup>55</sup>. Accelerometer datasets are also becoming suitable for these analyses<sup>62</sup>. However, differences in tag attachments across multiple sampling seasons (due to different researchers attaching the loggers, to double tagging experiments or tag positions) can cause discrepancies in accelerometer-derived variables, resulting in variability in the recorded behaviour<sup>12,63,64</sup>. We used pitch and VeDBA as candidates for body position and body motion. We accounted for possible differences in tag attachment by adjusting pitch values considering the subsurface slow swim/swim as the most common behaviour from which the tag “null position” could be identified. Standardizing accelerometer data across sampling seasons will help identify the effects of individual variability, and different environmental conditions on both animal behaviours and energy expenditure<sup>64</sup>.

Acceleration-based proxies for energy expenditure can provide new insights in optimal behavioural and evolutionary foraging theory, fitness and survival-related consequences of changes in energy expenditure<sup>12,65–67</sup>. It can potentially be used to explain how large-scale patterns at population levels may arise. An increasing amount of studies show that dynamic body acceleration, activity-specific dynamic body acceleration and activity budgets derived by accelerometer tags can predict energy expenditure, as measured via doubly labelled water in free-living animals<sup>16,32,68,69</sup>. Depending on the relationship found, as in number, type and resolution of behaviours estimated, the formulas linking accelerometer derived metrics with energy expenditure might be sensitive to different time budgets estimated via other methods. In our case, some of the fine-scale behavioural classes were grouped in coarser categories, leading to minimal differences in coarse time budgets and energy expenditure estimates. For Adélie penguins, our daily energy expenditure values predicted for *Season 1* fell within the range of values validated for *Season 2*, despite the use of slightly different behavioural classification methods<sup>32</sup>. Also the estimates for the total energy expended per trip fall within the range presented in Balance et al.<sup>70</sup>. However, this might not always be the case. We advise caution when using and mixing machine learning approaches to predict behaviours and energy expenditure, as this might cause variation in the resulting energy estimations.

Behavioural classification methods, used in ecology and bio-logging fields, based on a good understanding and careful examination of the acceleration signals can help improving computational time and accuracy of approaches used. Both unsupervised and supervised machine learning approaches applied to accelerometer data can provide insights into the behavioural ecology of wild animals from different perspectives but come with both benefits and costs. The unsupervised machine learning approach *Expectation Maximisation* allowed us to independently estimate behaviours for single individuals without using a priori knowledge of types of behaviour performed. Checking the behavioural classification of single foraging trips can be, however, time consuming and might lead to misinterpretation with increased behavioural complexity. The supervised machine learning approach, *Random Forest*, allowed us to upscale our analysis, build and test large training datasets across the two species sampling from the variability of behaviours available within our datasets. Random Forest, in fact, allows for quantification of the uncertainty around each estimated behaviour by means of confusion matrices. While computational speed was gained, it led to a potential loss of information on individual variability. These aspects are of vital importance when using behavioural classification's model outputs for testing theories on animals' fitness, energetics, and behavioural adaptation, as well as for informing conservation research.

Finally, training datasets accounting for behavioural variability could be applied to increasingly larger databases and/or used for real-time processing of accelerometer data on board of bio-logging devices to facilitate transmission via satellite systems<sup>27</sup>. While our framework has been tested on two penguin species, it is transferable to other species and systems. Methodological advances in computer science are allowing ecologists to handle and analyse large datasets more efficiently, but interpretation and transferability of results are tasks that need improvements. Refining the methodological applications calls for more integration between ecology and data science fields, along with standardizing the heterogeneous datasets collected, upscaling both data collection and analysis and interpretation of complex ecological patterns.

## Methods

**Ethics statement.** Field work protocol for both species was approved by the Terres Australes et Antarctiques Françaises (TAAF), the French regional ethic committee 54 for the Antarctic-based studies, and the ethics committee of the Phillip Island Nature Park Animal Experimentation Committee with a research permit issued by the Department of Environment, Land, Water and Planning of Victoria, Australia. All experiments were performed in accordance with relevant guidelines and regulations.

**Data collection.** Adélie penguins (*Pygoscelis adeliae*) were studied at the colony on Ile des Pétrils, at the Dumont d'Urville station, in Terre Adélie, East Antarctica (66°40' S; 140°01' E), over two breeding seasons, 2019–2020 (*Season 1, reference season*) and 2018–2019 (*Season 2, testing season*). The birds that we considered



were breeders during the chick guarding stage, in which one parent guards the chicks on the nest while the other is foraging at sea to bring food back to its offspring. Only one member of a pair was captured on or when leaving the nest, when both adults were attending the nest before/during a changeover. Little penguins (*Eudyptula minor*) were studied at Penguin Parade® on Phillip Island, Victoria, Australia (38° 31' S, 145° 09' E). Birds considered were breeding in artificial nest boxes and were sampled across incubation, guard and post-guard stages over two breeding seasons, 2020–2021 (*Season 1, reference season*) and 2019–2020 (*Season 2, testing season*). For both species, data-loggers (Axy-Trek, Technosmart, Italy, 40 × 20 × 8 mm, 14 g) were attached to the central dorsal region of the bird<sup>71</sup>. The loggers continuously sampled tri-axial acceleration at 100 Hz for Adélie penguins and 25 Hz for Little penguins, as well as pressure (in millibars) at 1 Hz and GPS locations (for Adélie penguins: 60 s; for little penguins in 2019–2020 300 s during incubation, 10 s during guard and 60 s during post guard, and in 2020–2021 30, 60 and 300 s during incubation, 10 s during guard and 60 s during post guard). GPS locations were only used during data preparation.

**Data preparation.** Acceleration data were checked and subsampled at 25 Hz, to be able to compare the two species across the analysis. Raw acceleration data are recorded over three axes: surge (back-forward acceleration), sway (lateral acceleration) and heave (dorso-ventral acceleration). Using a 2 s running mean over the raw acceleration signals for both species, we first calculated static acceleration, providing a measure of the angle of the instrumented animals. The time window applied for the running mean was arbitrarily chosen based on Shepard et al.<sup>72</sup> as a compromise to reduce noise and avoid masking behaviourally driven signals. By subtracting the static acceleration values from the raw values, we calculated dynamic acceleration for the three axes, measuring the change in velocity due to animal body motion. From the static and dynamic acceleration, vertical body orientation (pitch), lateral body orientation (roll), and Vectorial Dynamic Body Acceleration (VeDBA) were then calculated, following Hicks et al.<sup>32</sup>; but see also R code. It was assumed that different behaviours could be performed at different temporal resolutions<sup>20</sup>. Hence the following additional variables were also calculated: standard deviation of the raw acceleration values from the three axes over a running mean of 2, 10, 30 and 60 s, standard deviation of roll over 30 s and standard deviation of pitch and VeDBA over 60 s.

Pressure in millibars was converted to water depth (m) by subtracting the mode of the pressure distribution from the raw pressure and dividing by hundred. According to the resolution of the loggers and waves at the surface, only dives > 1 m and > 2 m, respectively, for Little and Adélie penguins, were considered diving behaviour<sup>32,73</sup>. Depth change rate was also calculated. Single versus multiple foraging trips were then defined by looking at GPS locations, depth, pitch, and temperature channels to detect when the animals were first entering the sea and when they were back at the colony. For Adélie penguins, we obtained 107 foraging trips (47 individuals) for *Season 1* and 53 foraging trips (48 individuals) for *Season 2*. For little penguins, we obtained 59 foraging trips (53 individuals) for *Season 1* and 39 foraging trips (38 individuals) for *Season 2*. Data analysis was performed in R version 4.0.4<sup>74</sup> on a processor of 3 GHz, 10-Cores and 128 GB of RAM. Distributions are presented as mean ± sd.

**Unsupervised machine learning approach.** Both species can perform several behaviours while moving in the wild. Specifically: “walk”, “stand”, “lie down”, “tobogganing” while on land (ice for Adélie penguins only), slow swim, preen, swim or porpoising at the subsurface level and ultimately descend, ascend, hunt and swim/cruise while diving<sup>75</sup>. To be able to detect these behaviours without a priori knowledge available, we followed the approach implemented in<sup>20</sup>, which used the unsupervised machine learning algorithm “Expectation Maximisation”, (hereafter EM). The algorithm is implemented in the R package “*Rmixmod*”<sup>76</sup> performing clustering analysis fitting a mixture model of multivariate Gaussian or multinomial components to given data. The algorithm selects an initial setting for the parameters, denoted  $h_i$ . Given a joint distribution over observed variables  $X$  and latent variables  $Z$ , governed by the chosen parameters, the algorithm maximises the likelihood function. The values of the latent variables in  $Z$  are given by the posterior distribution considering the expected value of the log-likelihood under the posterior distribution of the latent variable. In the Mstep, the algorithm evaluates  $h_{new}$  checking for convergence of either the log-likelihood or the parameter values. If the convergence criterion is not satisfied, the algorithm returns to select the initial settings for the parameters recalculating  $h_i$ <sup>20,77</sup>.

We have simplified the analytical procedure given the animals move in different dimensions (horizontal, as space, and vertical, as depth) and environments (land/ice and water). Foraging trips were split into coarse subsets representing when the animals were on ice (for Adélie only), at the water surface and diving, according to rules considering pitch (i.e. > 60 degrees to detect a penguin walking), and/or temperature (generally > 0°C when walking on ice) and depth as mentioned above (see R code and<sup>32</sup>). The subsets were then visually checked for potential errors. For each subset (2 or 3 per foraging trip, whether ice/land events were detected or not), variables were scaled and centered and the EM was randomly initialised by setting “*random*” within the “*mixmodStrategy*” function. Initialisation from a random position is a standard way to initialise the EM algorithm. This random initial position is obtained by choosing at random centres in the dataset. Based on the knowledge of the behavioural ecology of the species, EM runs for the diving subset were set to detect a minimum of three behavioural clusters and a maximum of eight. EM runs for the subsurface and ice subsets were set to detect two to five behavioural clusters. If only subsurface or ice events were detected, EM runs were set to run to detect from two to eight behavioural clusters. Variables fed in the algorithm were selected following an iterative approach consisting of considering the general knowledge of the environments where the data were collected, the general behaviours known about the study species (walk, preen and dive) and the statistical properties of the variables calculated<sup>20</sup>, but see also *Results* section. Foraging trips for each individual and season were run separately, and all behavioural clustering combinations recorded. Resulting clusters from each foraging trip analysed with the EM algorithm were visually assessed and associated to a behavioural class summarising the general behaviour performed (i.e., “descend”, “hunt”, “walk”).

**Supervised machine learning approach.** A set number of observations for each behavioural cluster was randomly drawn from each foraging trip to prepare the training dataset for the supervised machine learning approach. Only the Little penguins' dataset had recordings across different breeding stages. Breeding is not synchronized in this species and dive parameters overlap between breeding stages<sup>78</sup>, hence each foraging trip had the same probability to be sampled. The number of sampled observations was selected as the minimum between half of the observations from the smallest behavioural cluster and 1000. This way, it was possible to always leave at least half of the samples of each behavioural cluster for the test set and use at most half of the samples for training. The 1000 value is arbitrary and is only used to ensure that the number of samples taken from each trip is not diverging too much (i.e., each trip gets to contribute to the classifier with similar amount of data). To compare individual trips and seasons and account for possible differences in tag positioning, pitch values were adjusted by subtracting the median value of the Pitch considered as near the surface (values equal and above 2 m and less than 60 degrees for Adélie penguins, and values equal and above 1 m for little penguins). As the behavioural cluster "Swim/Cruise type 2" (see Results section) was performed by a few individuals for both species resulting in too few observations within the training, it was included in the general "Swim/Cruise" behavioural class.

To test the level of individual variation contained in the training datasets, we ran a repeatability analysis on both datasets using the R package "rptR"<sup>79</sup>. We used pitch and VeDBA as candidate variables to assess the repeatability on the overall body orientation and acceleration for each behaviour. Both pitch and VeDBA were used as response variable in two separate models, the reference "group" was set as individual bird with a set of consecutive trips. The repeatability index ranges from 0 to 1: a low repeatability (near 0) reflects either a high intra-individual variation or a low inter-individual variation.

We used the machine learning algorithm "Random Forest", (hereafter RF), implemented within the R packages "tidymodels" and "ranger"<sup>80,81</sup>. The RF algorithm was trained over twelve variables (see Table S1 and Results section) and parameterised on different numbers of trees (specifically 100, 500, 1000 and 1500). The parameter "mtry" indicating the number of variables randomly sampled as candidates at each split, around the sqrt(n max variables), was set to = 4. The randomly sampled training dataset obtained from all trips was split into further training (75%) and testing (25%) and a tenfold cross validation procedure performed on the final training data. Model selection was done by looking at the Out-Of-Bag error (OOB) and the Receiver Operating Characteristic curve (ROC) showing the performance of the classification model and the Area under the ROC Curve (AUC, see R code). The ROC curve shows the trade-off between sensitivity (or TPR) and specificity (1 - FPR).

**Comparing the classification approaches across sampling seasons.** We set up two sets of models to answer our research questions I and II (Fig. 1). The first model, containing random samples from each trip within the same season (*Season 1*), was built and validated on *Season 1* and used to predict on the second season (*Season 2*). The second model contained random samples from trips from both seasons (*Season 1* and *Season 2*). As we collected different foraging trips per season, the number of trips to be randomly sampled was selected as half of the total number of trips from the smallest season in order to account for differences in sampling, limit the size of the final training dataset and leave trips for prediction. For example, *Season 1* for Adélie penguins accounted for 107 trips while *Season 2* had 53 trips. Hence the number of trips from which observations were sampled from each season were 26. The resulting model was used to predict the behavioural classes on the remaining trips from both seasons. This procedure was run for each species separately.

**Effect of using different behavioural classification approaches on time-budgets and energy expenditure estimates.** Time budgets resulting from both EM and RF methods were calculated and compared. To answer question III (Fig. 1), we used the existing DEE vs activity-specific VeDBA relationship built and validated for Adélie penguins using data included in *Season 2*<sup>32</sup>. The model developed accounted for both VeDBA signatures, as well as time budgets. Thus, it was the ideal tool to fulfil our objective of investigating the consequences of using different behavioural classification methods. For the analysis, we only considered Adélie penguin data collected in *Season 1*. The RF algorithm used here, was trained solely on data from *Season 1*. We calculated DEEs for each trip using time budgets based on behavioural classes predicted by the EM and the RF, as well as total energy expended per trip, to be able to compare with other published estimates<sup>70</sup>. Following the best DEE model<sup>32</sup>, we also aggregated resulting behavioural classes in three coarse components: "Preen/Flap on water", "Dive" (including sub-surface swimming behaviours, as well as behaviours within dives) and "Land/Ice". We included the error estimates published in<sup>32</sup> and individual sex which was required for the calculation and assigned following morphometric measurements. We used linear models to test the agreement between the DEE calculations from the two methods. We also tested different models by considering the difference in estimated coarse components as covariates. We checked for collinearity among explanatory variables, as well as model residuals. Model selection was done using Akaike's Information Criterion (AIC).

## Data availability

Datasets used for this manuscript are available at: <https://doi.org/10.57745/IT6ITA>. Please cite as "CHIMIANTI, Marianna; KATO, Akiko; CHIARADIA, Andre; ROPERT COUDERT, Yan, 2022, "Data for - MuFFIN - Modelling Foraging Fitness in Marine Predators", <https://doi.org/10.57745/IT6ITA>, Recherche Data Gouv, V1". Sample dataset and all R codes used to manipulate and analyze the data used are available at [https://github.com/MariaChimi/MuFFIN\\_MSCA/releases/tag/v.02.Submission](https://github.com/MariaChimi/MuFFIN_MSCA/releases/tag/v.02.Submission).

Received: 6 July 2022; Accepted: 12 October 2022

Published online: 17 November 2022

## References

1. Tsai, C.-W., Lai, C.-F., Chao, H.-C. & Vasilakos, A. V. Big data analytics: a survey. *J. Big Data* **2**, 21 (2015).
2. Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
3. Manzoni, C. *et al.* Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.* **19**, 286–302 (2018).
4. Lichtman, J. W., Pfister, H. & Shavit, N. The big data challenges of connectomics. *Nat. Neurosci.* **17**, 1448–1454 (2014).
5. Altaf-Ul-Amin, M., Afendi, F. M., Kiboi, S. K. & Kanaya, S. Systems biology in the context of big data and networks. *Biomed. Res. Int.* **2014**, 428570 (2014).
6. Xia, J., Wang, J. & Niu, S. Research challenges and opportunities for using big data in global change biology. *Glob. Change Biol.* **26**, 6040–6061 (2020).
7. Hindell, M. A. *et al.* Tracking of marine predators to protect Southern Ocean ecosystems. *Nature* **580**, 87–92 (2020).
8. Hussey, N. E. *et al.* Ecology. Aquatic animal telemetry: A panoramic window into the underwater world. *Science* **348**, 1255642 (2015).
9. Kays, R., Crofoot, M. C., Jetz, W. & Wikelski, M. Terrestrial animal tracking as an eye on life and planet. *Science* **348**, aaa2478 (2015).
10. Sherub, S., Fiedler, W., Duriez, O. & Wikelski, M. Bio-logging, new technologies to study conservation physiology on the move: a case study on annual survival of Himalayan vultures. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **203**, 531–542 (2017).
11. Nathan, R. *et al.* Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* **375**, eabg1780 (2022).
12. Wilson, R. P. *et al.* Estimates for energy expenditure in free-living animals using acceleration proxies: A reappraisal. *J. Anim. Ecol.* **89**, 161–172 (2020).
13. Patterson, A., Gilchrist, H. G., Chivers, L., Hatch, S. & Elliott, K. A comparison of techniques for classifying behavior from accelerometers for two species of seabird. *Ecol. Evol.* **9**, 3030–3045 (2019).
14. Masello, J. F. *et al.* How animals distribute themselves in space: energy landscapes of Antarctic avian predators. *Mov. Ecol.* **9**, 24 (2021).
15. Shepard, E. L. C. *et al.* Energy landscapes shape animal movement ecology. *Am. Nat.* **182**, 298–312 (2013).
16. Elliott, K. H., Le Vaillant, M., Kato, A., Speakman, J. R. & Ropert-Coudert, Y. Accelerometry predicts daily energy expenditure in a bird with high activity levels. *Biol. Lett.* **9**, 20120919 (2013).
17. Nickel, B. A., Suraci, J. P., Nisi, A. C. & Wilmers, C. C. Energetics and fear of humans constrain the spatial ecology of pumas. *Proc. Natl. Acad. Sci. USA* **118**, e2004592118 (2021).
18. Eisaguirre, J. M., Booms, T. L., Barger, C. P., Lewis, S. B. & Breed, G. A. Novel step selection analyses on energy landscapes reveal how linear features alter migrations of soaring birds. *J. Anim. Ecol.* **89**, 2567–2583 (2020).
19. Wittemyer, G., Northrup, J. M. & Bastille-Rousseau, G. Behavioural valuation of landscapes using movement data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180046 (2019).
20. Chimienti, M. *et al.* The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data. *Ecol. Evol.* **6**, 727–741 (2016).
21. Hounslow, J. L. *et al.* Assessing the effects of sampling frequency on behavioural classification of accelerometer data. *J. Exp. Mar. Biol. Ecol.* **512**, 22–30 (2019).
22. Glass, T. W., Breed, G. A., Robards, M. D., Williams, C. T. & Kiehl, K. Accounting for unknown behaviors of free-living animals in accelerometer-based classification models: Demonstration on a wide-ranging mesopredator. *Ecol. Inf.* **60**, 101152 (2020).
23. Wang, Y. *et al.* Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. *Mov. Ecol.* **3**, 2 (2015).
24. Chakravarty, P., Cozzi, G., Ozgul, A. & Aminian, K. A novel biomechanical approach for animal behaviour recognition using accelerometers. *Methods Ecol. Evol.* <https://doi.org/10.1111/2041-210X.13172> (2019).
25. Clarke, T. M. *et al.* Using tri-axial accelerometer loggers to identify spawning behaviours of large pelagic fish. *Mov. Ecol.* **9**, 26 (2021).
26. Zhang, J., O'Reilly, K. M., Perry, G. L. W., Taylor, G. A. & Dennis, T. E. Extending the functionality of behavioural change-point analysis with k-Means clustering: a case study with the little penguin (*Eudyptula minor*). *PLoS ONE* **10**, e0122811 (2015).
27. Korpela, J. *et al.* Machine learning enables improved runtime and precision for bio-loggers on seabirds. *Commun. Biol.* **3**, 633 (2020).
28. Jeantet, L. *et al.* Behavioural inference from signal processing using animal-borne multi-sensor loggers: a novel solution to extend the knowledge of sea turtle ecology. *R. Soc. Open Sci.* **7**, 200139 (2020).
29. Wang, G. Machine learning for inferring animal behavior from location and movement data. *Ecol. Inf.* **49**, 69–76 (2019).
30. Dunford, C. E. *et al.* Surviving in steep terrain: a lab-to-field assessment of locomotor costs for wild mountain lions (*Puma concolor*). *Mov. Ecol.* **8**, 34 (2020).
31. Jeanniard-du-Dot, T., Guinet, C., Arnould, J. P. Y., Speakman, J. R. & Trites, A. W. Accelerometers can measure total and activity-specific energy expenditures in free-ranging marine mammals only if linked to time-activity budgets. *Funct. Ecol.* **31**, 377–386 (2017).
32. Hicks, O. *et al.* Acceleration predicts energy expenditure in a fat, flightless, diving bird. *Sci. Rep.* **10**, 21493 (2020).
33. Dentinger, J. E. *et al.* A probabilistic framework for behavioral identification from animal-borne accelerometers. *Ecol. Model.* **464**, 109818 (2022).
34. Chakravarty, P., Maalberg, M., Cozzi, G., Ozgul, A. & Aminian, K. Behavioural compass: animal behaviour recognition using magnetometers. *Mov. Ecol.* **7**, 28 (2019).
35. Hammond, T. T., Palme, R. & Lacey, E. A. Ecological specialization, variability in activity patterns and response to environmental change. *Biol. Lett.* **14**, 20180115 (2018).
36. Lynch, H. J. & LaRue, M. A. First global census of the Adélie Penguin. *Auk* **131**, 457–466 (2014).
37. Riaz, J., Bestley, S., Wotherspoon, S., Freyer, J. & Emmerson, L. From trips to bouts to dives: temporal patterns in the diving behaviour of chick-rearing Adélie penguins East Antarctica. *Mar. Ecol. Prog. Ser.* **654**, 177–194 (2020).
38. Cherel, Y. Isotopic niches of emperor and Adélie penguins in Adélie Land, Antarctica. *Mar. Biol.* **154**, 813–821 (2008).
39. Little Penguin (*Eudyptula minor*) - BirdLife species factsheet. at <<http://datazone.birdlife.org/species/factsheet/little-penguin-eudyptula-minor/details>>
40. Carroll, G., Harcourt, R., Pitcher, B. J., Slip, D. & Jonsen, I. Recent prey capture experience and dynamic habitat quality mediate short-term foraging site fidelity in a seabird. *Proc. Biol. Sci.* **285**, 20180788 (2018).
41. Meyer, X. *et al.* Oceanic thermal structure mediates dive sequences in a foraging seabird. *Ecol. Evol.* **10**, 6610–6622 (2020).
42. Cavallo, C. *et al.* Quantifying prey availability using the foraging plasticity of a marine predator, the little penguin. *Funct. Ecol.* <https://doi.org/10.1111/1365-2435.13605> (2020).
43. Ropert-Coudert, Y., Chiaradia, A. & Kato, A. An exceptionally deep dive by a Little Penguin *Eudyptula minor*. *Mar. Ornithol.* **34**, 71–74 (2006).

44. Ropert-Coudert, Y., Kato, A., Wilson, R. P. & Cannell, B. Foraging strategies and prey encounter rate of free-ranging Little Penguins. *Mar. Biol.* **149**, 139–148 (2006).
45. Rodríguez, A., Chiaradia, A., Wasiak, P., Renwick, L. & Dann, P. Waddling on the dark side: ambient light affects attendance behavior of little penguins. *J. Biol. Rhythms* **31**, 194–204 (2016).
46. Ropert-Coudert, Y. *et al.* Happy feet in a hostile world? the future of penguins depends on proactive management of current and expected threats. *Front. Mar. Sci.* **6**, 248 (2019).
47. Shuert, C. R., Pomeroy, P. P. & Twiss, S. D. Assessing the utility and limitations of accelerometers and machine learning approaches in classifying behaviour during lactation in a phocid seal. *Anim. Biotelemetry* **6**, 14 (2018).
48. Dickinson, E. R. *et al.* Limitations of using surrogates for behaviour classification of accelerometer data: refining methods using random forest models in Caprids. *Mov. Ecol.* **9**, 28 (2021).
49. Conway, A. M., Durbach, I. N., McInnes, A. & Harris, R. N. Frame-by-frame annotation of video recordings using deep neural networks. *Ecosphere* **12**, e03384 (2021).
50. Ravindran, S. Five ways deep learning has transformed image analysis. *Nature* **609**, 864–866 (2022).
51. Del Caño, M. *et al.* Fine-scale body and head movements allow to determine prey capture events in the Magellanic Penguin (*Spheniscus magellanicus*). *Mar. Biol.* **168**, 84 (2021).
52. Johnson, J. M. & Khoshgoftaar, T. M. Survey on deep learning with class imbalance. *J. Big Data* **6**, 27 (2019).
53. Hazen, E. L. *et al.* Marine top predators as climate and ecosystem sentinels. *Front. Ecol. Environ.* **17**, 565–574 (2019).
54. Sánchez, S. *et al.* Within-colony spatial segregation leads to foraging behaviour variation in a seabird. *Mar. Ecol. Prog. Ser.* **606**, 215–230 (2018).
55. Patrick, S. C., Martin, J. G. A., Ummenhofer, C. C., Corbeau, A. & Weimerskirch, H. Albatrosses respond adaptively to climate variability by changing variance in a foraging trait. *Glob. Change Biol.* <https://doi.org/10.1111/gcb.15735> (2021).
56. Bonar, M. *et al.* Geometry of the ideal free distribution: individual behavioural variation and annual reproductive success in aggregations of a social ungulate. *Ecol. Lett.* **23**, 1360–1369 (2020).
57. Michelot, C., Kato, A., Raclot, T. & Ropert-Coudert, Y. Adélie penguins foraging consistency and site fidelity are conditioned by breeding status and environmental conditions. *PLoS ONE* **16**, e0244298 (2021).
58. Mahoney, P. J. *et al.* Navigating snowscapes: scale-dependent responses of mountain sheep to snowpack properties. *Ecol. Appl.* **28**, 1715–1729 (2018).
59. Watanabe, Y. Y., Ito, K., Kokubun, N. & Takahashi, A. Foraging behavior links sea ice to breeding success in Antarctic penguins. *Sci. Adv.* **6**, eaba4828 (2020).
60. Lescroël, A. *et al.* Working less to gain more: when breeding quality relates to foraging efficiency. *Ecology* **91**, 2044–2055 (2010).
61. Zimmer, I., Ropert-Coudert, Y., Kato, A., Ancel, A. & Chiaradia, A. Does foraging performance change with age in female little penguins (*Eudyptula minor*)?. *PLoS ONE* **6**, e16098 (2011).
62. Hertel, A. G., Royauté, R., Zedrosser, A. & Mueller, T. Biologging reveals individual variation in behavioural predictability in the wild. *J. Anim. Ecol.* **90**, 723–737 (2021).
63. Dickinson, E. R., Stephens, P. A., Marks, N. J., Wilson, R. P. & Scantlebury, D. M. Best practice for collar deployment of tri-axial accelerometers on a terrestrial quadruped to provide accurate measurement of body acceleration. *Anim. Biotelemetry* **8**, 9 (2020).
64. Garde, B. *et al.* Ecological inference using data from accelerometers needs careful protocols. *Methods Ecol. Evol.* <https://doi.org/10.1111/2041-210X.13804> (2022).
65. Watanabe, Y. Y., Ito, M. & Takahashi, A. Testing optimal foraging theory in a penguin-krill system. *Proc. Biol. Sci.* **281**, 20132376 (2014).
66. Grémillet, D. *et al.* Energetic fitness: Field metabolic rates assessed via 3D accelerometry complement conventional fitness metrics. *Funct. Ecol.* **32**, 1203–1213 (2018).
67. Chimienti, M. *et al.* Quantifying behavior and life-history events of an Arctic ungulate from year-long continuous accelerometer data. *Ecosphere* **12**, e03565 (2021).
68. Sutton, G. J., Botha, J. A., Speakman, J. R. & Arnould, J. P. Y. Validating accelerometry-derived proxies of energy expenditure using the doubly-labelled water method in the smallest penguin species. *Biol. Open* **10**, bio055475 (2021).
69. Pagano, A. M. & Williams, T. M. Estimating the energy expenditure of free-ranging polar bears using tri-axial accelerometers: A validation with doubly labeled water. *Ecol. Evol.* **9**, 4210–4219 (2019).
70. Ballance, L. T., Ainley, D. G., Ballard, G. & Barton, K. An energetic correlate between colony size and foraging effort in seabirds, an example of the Adélie penguin *Pygoscelis adeliae*. *J. Avian Biol.* **40**, 279–288 (2009).
71. Wilson, R. P. *et al.* Long-term attachment of transmitting and recording devices to penguins and other seabirds. *Wildl. Soc. Bull.* **25**, 101–106 (1997).
72. Shepard, E. L. C. *et al.* Identification of animal movement patterns using tri-axial accelerometry. *Endanger. Species Res.* **10**, 47–60 (2008).
73. Kato, A., Ropert-Coudert, Y., Grémillet, D. & Cannell, B. Locomotion and foraging strategy in foot-propelled and wing-propelled shallow-diving seabirds. *Mar. Ecol. Prog. Ser.* **308**, 293–301 (2006).
74. Team, R. C. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://url.org/www.R-project.org/> (2021).
75. Ainley, D. *The Adélie Penguin: Bellwether of Climate Change* (New York: Columbia University Press) (2006).
76. Langrognet, F. *et al.* Rmixmod: Classification with Mixture Modelling. (2020).
77. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, NY. (2006).
78. Amélineau, F. *et al.* Intra- and inter-individual changes in little penguin diving and isotopic composition over the breeding season. *Mar. Biol.* **168**, 62 (2021).
79. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644 (2017).
80. Kuhn, M. & Wickham, H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. <https://www.tidymodels.org> (2020).
81. Wright, M. N. & Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).

## Acknowledgements

We thank the editor Dr Kyle Elliott and two anonymous reviewers for providing constructive comments to earlier versions of this manuscript. We thank Phillip Island Nature Parks for their continued support and commitment to penguin research. We would like to thank Dr Cara Gallagher for designing the animals' logos added to the figures representing the two study species.

## Author contributions

Conceptualization: M.C., A.K., A.C., Y.R.C. Methodology: M.C., A.K., Y.R.C. Data Collection: O.H., F.A., M.B., J.O., C.M., T.R., M.T., D.W. Visualization: M.C. Supervision: A.K., A.C., Y.R.C. Writing—original draft: M.C., A.K., A.C., Y.R.C. Writing—review & editing: all authors.

## Funding

This study was supported by: H2020-Marie Skłodowska-Curie Individual Fellowship, project number 890284, Modelling Foraging Fitness in Marine predators (MuFFIN), awarded to Marianna Chimienti, Logistic and/or financial support was provided by: French Polar Institute Paul-Emile Victor (IPEV) through the project 1091, Zone Atelier Antarctique et Subantarctique – LTER France of the CNRS, PEW Foundation, WWF-UK through R. Downie, This study is also contribution to the program SENSEI funded by the BNP Paribas Foundation.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22258-1>.

**Correspondence** and requests for materials should be addressed to M.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022