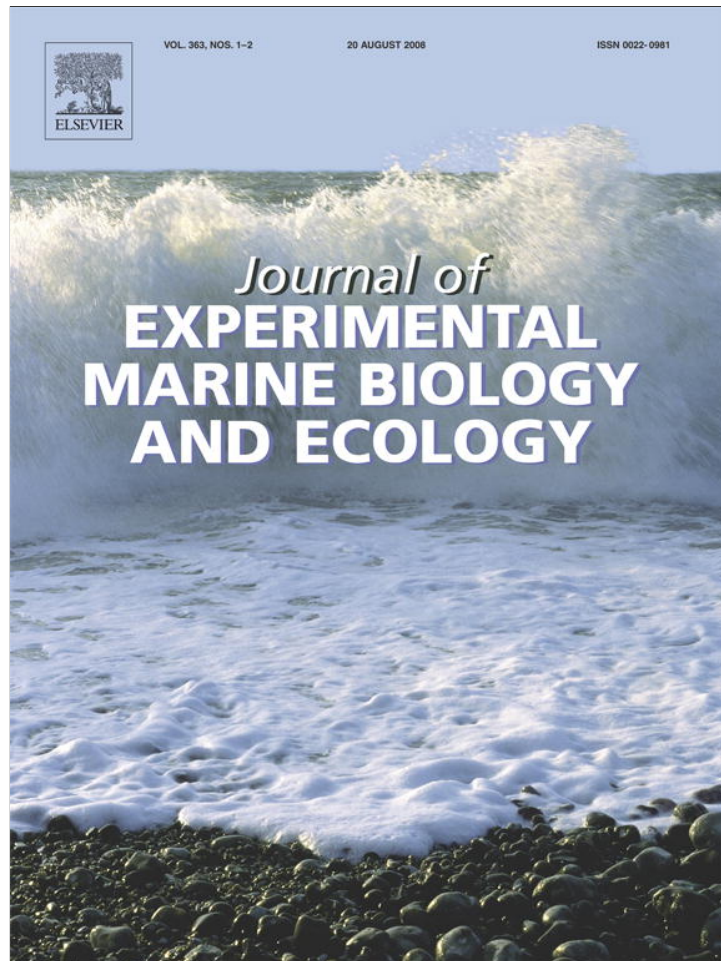


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Journal of Experimental Marine Biology and Ecology

journal homepage: [www.elsevier.com/locate/jembe](http://www.elsevier.com/locate/jembe)

## A validated approach for supervised dive classification in diving vertebrates

Michele Thums<sup>a,\*</sup>, Corey J.A. Bradshaw<sup>b,c</sup>, Mark A. Hindell<sup>a</sup><sup>a</sup> Antarctic Wildlife Research Unit, School of Zoology, University of Tasmania, Private Bag 05, Hobart, Tasmania 7001, Australia<sup>b</sup> Research Institute for Climate Change and Sustainability, School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia<sup>c</sup> South Australian Research and Development Institute, P.O. Box 120, Henley Beach, South Australia 5022, Australia

## ARTICLE INFO

## Article history:

Received 29 May 2008

Received in revised form 18 June 2008

Accepted 19 June 2008

## Keywords:

Dive classification

Diving

Elephant seal

Random forests algorithm

Supervised classification

Time-depth profiles

Time-depth recorders

## ABSTRACT

Supervised dive classification is a commonly used technique for categorising time-depth profiles of diving vertebrates. Such analyses permit the description and quantification of dive behaviour and foraging tactics, and highlight feeding locations. Ideally, classification functions should be validated, and this is commonly done visually. Visual classification is subjective, but it is currently one of the few measures available for validation. We develop several approaches to validate a supervised dive classification: (1) two people visually assigning dives and developing a dataset where both agree, and (2) use of dives from southern elephant seals identified as “drift dives” with characteristic velocity signatures. We classified the dives of three seals from their post-moult foraging trips and estimated the error associated with visual classification. We found classification error (disagreement) between classifiers up to 57%. We created a training dataset based on dives with agreement and applied this to a relatively new, automated classification method – the random forests (RF) algorithm. A supervised function developed using this algorithm estimated a classification error of 5% on elephant seal dives; classification error on underrepresented dive classes ranged from 2 – 12%. Testing this classification function on independent data produced a low error (1.6%). RF function errors were lower than for visual classification, and errors were similar to or better than those estimated using discriminant functions. Swim velocity parameters were the most important predictors, but their absence did not reduce the random forests function's effectiveness by much. Our results suggest that there is a temporal shift in diving behaviour as seals become more buoyant. We compared the temporal patterns in drift rate from the drift dives classified using the RF function with the drift dives validated via characteristic velocity signatures. This indicated that the classifications produced by RF function are valid even though some error persists, and we suggest adoption of this method for classifying dives in other air-breathing diving vertebrates.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

For many diving vertebrates such as penguins, turtles, seals and whales, a large proportion of time is spent underwater where direct observation of behaviour is difficult or impossible. Technical advances over the last 30 years have enabled the collection of data describing elements of sub-surface foraging, travelling and other activities (Hooker et al., 2007). Time-depth recorders (TDRs) have provided some of the simplest, yet voluminous and informative data to date. Interpreting TDR-collected data has commonly involved reconstructing time-depth profiles of individual dives and categorising them into different putative behavioural types related predominately to transit, foraging and rest based on their two-dimensional shape. Such analyses have provided a wealth of information on diving behaviour and have proven useful for understanding foraging strategies (Burns

et al., 1997; Hassrick et al., 2007; Putz and Cherel, 2005; Wilson, 1995), for inferring foraging/feeding behaviour (Austin et al., 2006; Bailleul et al., 2007a; Biuw et al., 2003, 2007; Le Boeuf et al., 2000; Robinson et al., 2007; Simeone and Wilson, 2003) and for highlighting feeding locations and the environmental gradients associated with successful foraging (Bailleul et al., 2007a,b; Biuw et al., 2007).

One dive type that has been used to infer prey acquisition is the drift diving behaviour of seals (*Mirounga* spp. and *Arctocephalus pusillus*). These are dives where the animal drifts passively in the water column for several minutes after first swimming to some depth (Crocker et al., 1997). Buoyancy during the drifting phase, and hence, relative body composition, can be measured as changes in the rate of vertical displacement. In seals, buoyancy is determined largely by the relative proportions of lipid and lean body tissue (Beck et al., 2000; Biuw et al., 2003; Webb et al., 1998); individuals with a high lipid:lean tissue ratio are more buoyant than individuals of similar mass but with a lower proportional lipid store (Beck et al., 2000; Biuw et al., 2003; Thums et al., in press).

Another commonly identified dive type is the square-shaped (also called U-shaped) dive thought to indicate foraging behaviour in seals

\* Corresponding author. Antarctic Wildlife Research Unit, School of Zoology, University of Tasmania, Private Bag 05, Hobart, Tasmania 7001, Australia. Tel.: +61 3 6226 2645; fax: +61 3 6226 2745.

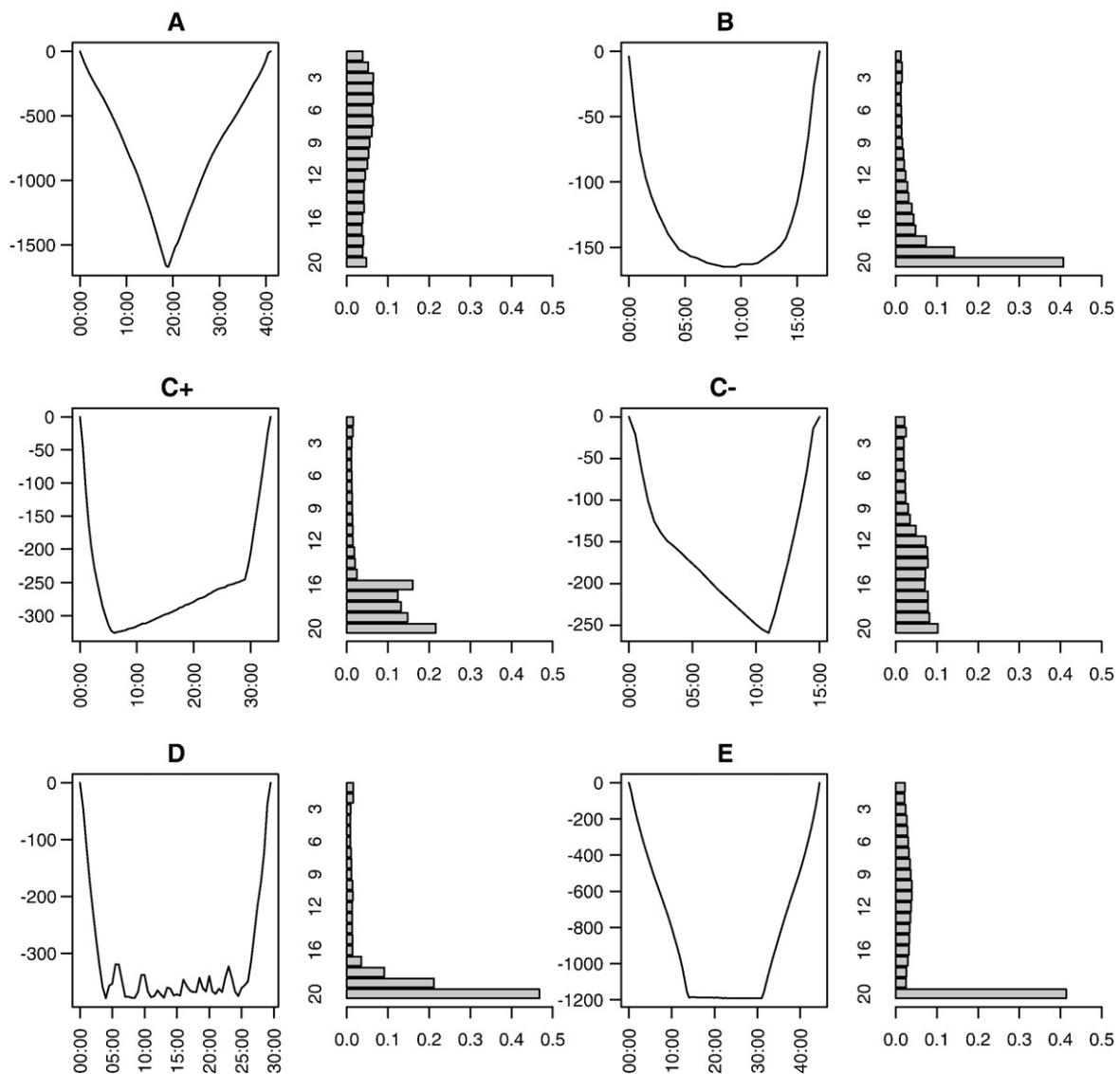
E-mail address: [mthums@utas.edu.au](mailto:mthums@utas.edu.au) (M. Thums).

and diving birds (Schreer et al., 2001), some cetaceans (Malcolm and Duffus, 2000; Nowacek et al., 2001), and benthic resting and foraging in sea turtles (Hochscheid et al., 2005; McMahon et al., 2007). A direct link between square dives and feeding has been found in some species by correlating drops in stomach temperature (Lesage et al., 1999; Wilson et al., 1995) and video image recordings of prey capture (Davis et al., 2003). The identification and quantification of square dives has been used to infer changes in foraging behaviour in space and time (Baechler et al., 2002; Le Boeuf et al., 2000; Schreer et al., 2001; Wilson and Wilson, 1990), to validate track based estimates of feeding locations (Robinson et al., 2007) and to indicate resting behaviour (Hays et al., 2004a; Hochscheid et al., 2005).

Early analyses of TDR data classified dive types visually based on their shape, but the associated subjectivity and effort required renders this method unsuitable for large datasets. Multivariate statistical methods can provide objective and fast classification with reasonable precision (Schreer and Testa, 1996). The two most commonly used statistical techniques are discriminant function and cluster analysis. Discriminant function requires *a priori* groupings to develop the classification function, so it is commonly used to validate the groups created by cluster analysis or visual classification (Schreer and Testa,

1995). The validation is thus a test of whether the dive types assessed by either visual or cluster methods are valid for the data, i.e., a high error rate may indicate an invalid group. In a sense, this is not a true validation because there is no “correct” solution. Further, it is an internal validation, only providing an estimate of how well a new dataset would be classified.

We developed a supervised classification function for two-dimensional dive data using more robust validation processes. We attempted to circumvent some of the common validation problems by (1) attempting to obtain a “correct” solution using two human classifiers visually assigning dives and developing a dataset where the two classifiers agree, (2) obtaining a truly “correct” solution for one type of dive, drift dives, independently verified based on characteristic velocity signatures, and (3) testing the method on an independent dataset (one that was not used in the development of the function). We also tested a relatively new automated classification method, the random forests algorithm (RF) (Breiman, 2001), which appears to perform well compared to other techniques such as discriminant functions and neural networks (Breiman, 2001). More importantly, the RF algorithm is relatively simple to implement because it does not require cross validation with a separate test dataset, it is resistant to



**Fig. 1.** Each dive class (A) V-shaped dives, (B) round dives (C+) positive drift dives, (C-) negative drift dives, (D) square dives with depth oscillations (wiggles) and (E) flat-bottomed, square dives plotted as depth (y-axis) versus time (x-axis), with its corresponding plot of time spent in each of 20 equally spaced depth bins (y-axis) as a proportion of the total dive duration (x-axis).

over-fitting, and it estimates the precision associated with each classification prediction as well as the relative weight of each dive characteristic in determining the particular classification (Liaw and Wiener, 2002). Variables calculable from speed turbines on some of the TDRs were also used to determine whether their inclusion in the methods used can provide greater predictive precision for the classification of dive types as suggested by Davis et al. (2003), Hassrick et al. (2007) and Lesage et al. (1999).

## 2. Methods

### 2.1. Data collection

We instrumented adult female southern elephant seals (*Mirounga leonina*) from Macquarie Island (54° 35' S, 158° 58' E) with velocity time–depth recorders (VTDR) (Wildlife Computers MK8, Redmond, Washington, USA) prior to the 2004 ( $n=12$ ) post-lactation (Oct–Dec) and 2002 ( $n=14$ ), 2004 ( $n=16$ ) and 2005 ( $n=4$ ) post-moult (Feb–Sep) foraging trips. Seals were captured and sedated following the procedures outlined in Field et al. (2002). The VTDRs were attached to the pelage above the shoulders following the procedures outlined in Hindell and Slip (1997) and sampled time, depth, light level, revolutions of a flow-driven turbine mounted on top of the instrument and temperature every 30 seconds for the duration of both foraging trips. Speed was derived from the logged revolutions of the turbine using the *post hoc* calibration method outlined in Fletcher et al. (1996).

### 2.2. Data extraction and dive variables

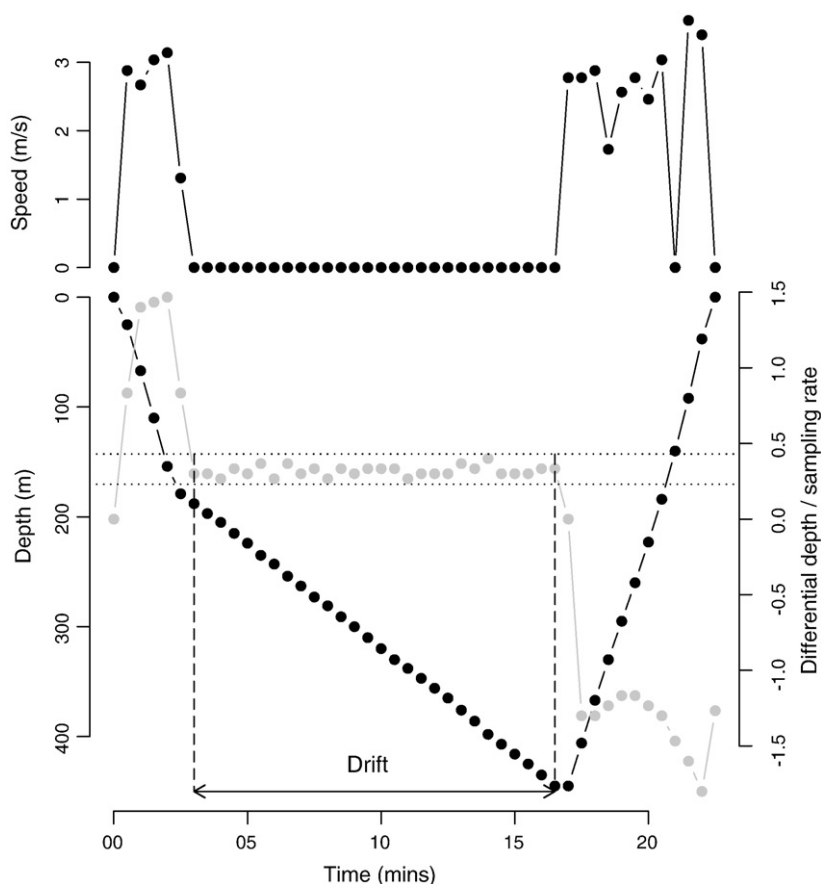
Raw data from the VTDRs were extracted using software provided by Wildlife Computers (Redmond, Washington, USA). Summary

parameters for each dive were then calculated using 'DIVE', a custom dive-analysis program (Stuart Greenhill, Murdoch University, Perth, Western Australia). These parameters were maximum depth (m), duration (min), total mean speed ( $\text{m s}^{-1}$ ), descent speed ( $\text{m s}^{-1}$ ), ascent speed ( $\text{m s}^{-1}$ ), descent rate ( $\text{m s}^{-1}$ ), ascent rate ( $\text{m s}^{-1}$ ), vertical and horizontal distances travelled during ascent and descent (m), number of wiggles (depth oscillations at the bottom of a dive) and time spent at the bottom of a dive (min). We also calculated the time spent in each of 20 equally spaced depth bins as a proportion of the total dive duration for each dive (Fig. 1). Similar to the method used by Schreer and Testa (1995), we also calculated 20 depth profiles for each dive by dividing dive duration into 20 equal parts and computing the proportion of each corresponding depth to the total depth of the dive, thereby standardising for varying depth. We also calculated the descent and ascent angles.

### 2.3. Development of classification function

We used the tree-based model *random forests*, developed by Breiman (2001). Random forests differs from a standard tree in that instead of splitting each node using the best split among all variables, each node is split using the best of a subset of predictors chosen randomly at that node. We used the R Language (R Development Core Team, 2007) interface to the original Fortran programs, *randomForest* (Liaw and Wiener, 2002).

In RF, cross-validation, or a separate unbiased estimate of error, is not required because it is estimated internally. Trees are constructed using a different bootstrap sample each time from the original data. About one-third of the cases are left out of the bootstrap sample and classifications are given to these data using the tree grown with the bootstrap sample. In this way, a test set classification is obtained for



**Fig. 2.** (Top) For one C-dive, swim speed is plotted against time. (Bottom) Dive depth against time (black), with the first differential of depth/sampling rate plotted against time (grey). Threshold differentials (black dotted lines) and drift segments (black dashed lines) are shown. Note:  $0 \text{ m s}^{-1}$  on the graph denotes anything  $\leq$  the stall speed of the turbine.



each case in about one third of the trees. These predictions are then aggregated and the error is calculated. This is known as the 'out-of-bag' (OOB) estimate of error and is generally unbiased in many tests (Breiman, 2001).

RandomForest also computes the contribution of each explanatory variable to the classification using the Gini Index (Breiman, 2002). Each time a node split occurs based on a particular variable, the Gini impurity criterion for the two descendent nodes is less than the parent node. The Gini index is calculated by summing the Gini decreases for each individual variable over all trees in the forest (Breiman, 2002). To assess the importance of the speed-derived variables on the classification success, we developed two RF classification functions, one including variables derived from speed sensors (total mean speed ( $m s^{-1}$ ), descent speed ( $m s^{-1}$ ), ascent speed ( $m s^{-1}$ ), horizontal distance (m) and descent and ascent angles) and another excluding the speed-derived variables.

2.4. Prior specification of classes for development of a training and validation dataset

RandomForest was run in supervised mode; thus, the dive classes had to be specified a priori to develop a training set of dives required to train the function and for validation. We used visual classification and three of the Mk8 TDR records were chosen for this purpose. These had data for their entire post-moult foraging trip ( $242 \pm 8$  days [mean  $\pm$  SD]) comprising 37224 dives in total ( $11796 \pm 592$ ). Due to the potential subjectivity inherent in this method, two people (classifiers) assigned the same dives to classes independently and only those dives that were assigned to the same class by both were used in the training set. This was subsetted to ensure that equal numbers of each of the dive classes were present. This was necessary because around 80% of southern elephant seal dives are square dives (Hindell et al., 1991), which would have underrepresented other dive classes. This resulted in a training dataset of 2226 dives, with 371 of each dive class used to develop the classification function. We further subsetted the drift dives by only accepting those independently verified as a drift dive by the presence of a speed reading  $\leq$  stall speed of the swim speed turbine during the putative drift phase (Fig. 2) (Thums et al., in press).

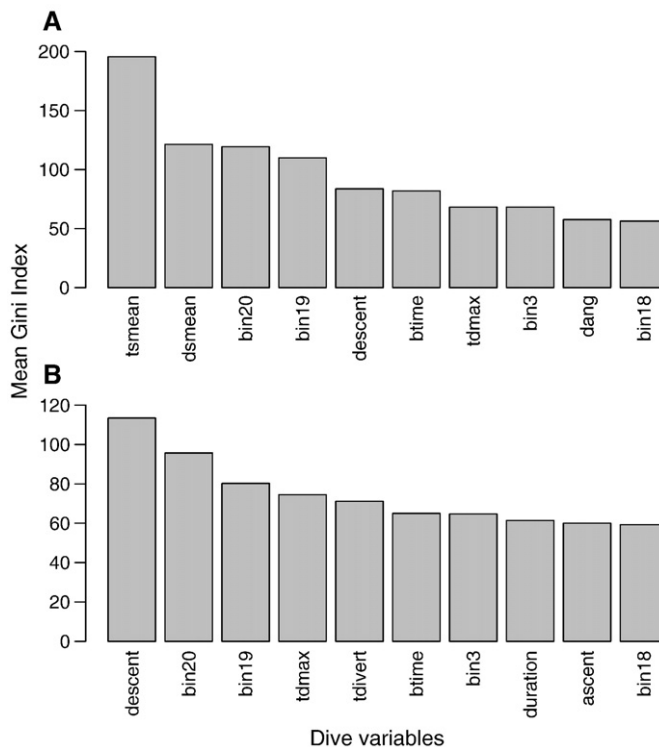
The classes used were (A) V-shaped dives, (B) round dives (C) drift dives, (D) square dives with depth oscillations (wiggles) and (E) flat-bottomed, square dives (Asaga et al., 1994; Crocker et al., 1997; Hassrick et al., 2007; Hindell et al., 1991; Jonker and Bester, 1994; Le Boeuf et al., 1992, 1993). We further demarcated drift dives, separating drift dives with a floating (dive type C<sup>+</sup>) versus sinking drift component (type C<sup>-</sup>). We therefore had a final total of six dive classes. Dives that could not be allocated visually to any class (4% of dives) were removed from the analysis.

**Table 1**  
The 'out-of-bag' (OOB) estimate of error from the Random Forests (RF) classification function with and without speed parameters, and the visual classification error calculated assuming human classifiers 1 and 2 were correct

Dive Class	Random Forests		Visual		RF (PL)
	+Speed	-Speed	Classifier 1	Classifier 2	-Speed
A	1.9	1.3	6	57	0
B	7.5	8.4	10	31	5.3
C <sup>+</sup>	2.1	1.6	14	46	NA
C <sup>-</sup>	4.0	4.3	11	10	0
D	11.8	12.4	8	2	1.6
E	1.9	2.2	2	12	NA
Total	4.9	5.0	8		1.6

The total error calculated for the visual classification is the % disagreement between the two classifiers. The final column shows the errors from testing the RF function on a post-lactation (PL) seal's dataset.

Dive category A=V-shaped dives; B=round dives; C<sup>+</sup>=positive drift dives; C<sup>-</sup>=negative drift dives; D=square dives with depth oscillations (wiggles); E=flat-bottomed, square dives.



**Fig. 3.** The importance of each of the top ten variables to the classification based on the Mean Gini Index computed by RandomForest (A) using speed variables or (B) without using speed variables. tsmean=mean total speed, dsmean=mean descent speed, bin20=time spent in the deepest depth bin as a proportion of total dive duration, bin19=time spent in the second deepest depth bin as a proportion of total dive duration, descent=descent rate, ascent=ascent rate, btime2=time spent at the bottom of the dive as a proportion of total dive duration, tdmx=maximum depth, bin3=time spent in the third shallowest depth bin as a proportion of total dive duration, dang=descent angle, bin18=time spent in the third deepest depth bin as a proportion of total dive duration, duration=total dive duration, tdivert=total vertical distance.

2.5. Error associated with visual dive classification

With two classifiers categorising dives visually to create the training and validation dataset, we were able to assess the agreement and thus obtain an idea of the errors associated with visual classification. It is not possible to know the "true" classification of a dive, so we calculated the errors (% disagreement) twice (once assuming Classifier 1 was correct, and again assuming Classifier 2 was correct). We therefore calculated a range of errors associated with each dive class. For drift dives, we calculated a definitive error because we could identify true periods of drift from the swim speed data (Thums et al., in press).

**Table 2**  
The mean and standard deviation (SD) of the percentage of dive types found in the records of the three post-moult seals that had their dives classified manually versus those classified by Random Forests

Dive Class	Post-moult			Post-lactation		
	Manual	RF	Δ	Manual	RF	Δ
A	1.5 $\pm$ 2.0	3.5 $\pm$ 4.2	2.0	0.1	0.2	0.1
B	6.9 $\pm$ 3.6	8.7 $\pm$ 3.1	1.8	2.2	2.7	0.5
C <sup>+</sup>	1.2 $\pm$ 0.5	3.9 $\pm$ 1.1	2.7	0	0.3	NA
C <sup>-</sup>	6.4 $\pm$ 3.8	7.5 $\pm$ 4.1	1.1	4.7	5.1	0.4
D & E	84.1 $\pm$ 6.2	76.5 $\pm$ 4.8	7.6	92.7	91.5	1.2

There was only 1 post-lactation dataset visually classified so only the percentage of each dive class is presented. Note: The post-lactation seal did not exhibit types E or C<sup>+</sup> dives, and there were only three type A dives present in the record.

Note: The post-lactation seal did not exhibit types E or C<sup>+</sup> dives, and there were only three type A dives present in the record.

## 2.6. External and independent validations

We tested the RF classification function by using it to predict the dive classes of an external dataset, i.e., the data from a seal from which dives were not included in the development of the classification functions. The external data were also collected during a different type of foraging trip (post-breeding) and from a different year (2004) than the training-set data. These data were visually classified by the two classifiers as before, and subsetted based on agreement. The RF classification function then predicted dive classes on this dataset. Using known drift dives (as above), we truly validated our classification function for this dive class.

## 2.7. Temporal patterns in rate of drifting during drift dives to assess the validity of RF dive classifications

Tracking the rate of drifting during drift dives is a useful indicator of foraging success over a seals foraging trip. We used these dives to assess the effect of misclassifications by comparing the temporal patterns in drift rate (see details below) obtained from the RF-classified drift dives to the patterns obtained from drift dives identified on the basis of having a drifting segment with speed  $\leq$  the stall speed of the turbine (stall speed effectively indicates no, or minimal, active propulsion) (Fig. 2(top)).

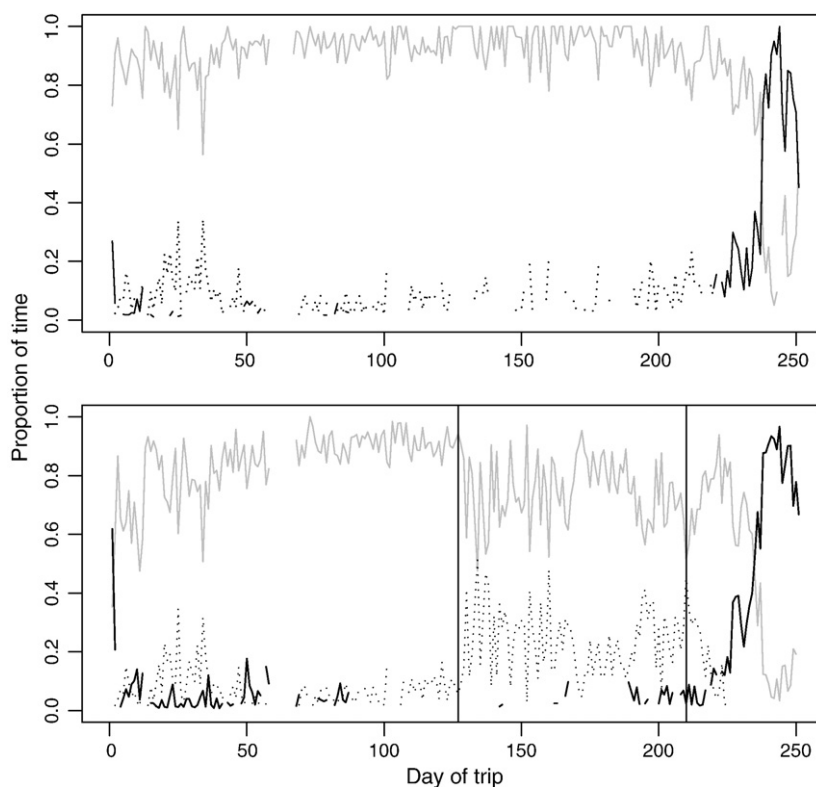
To calculate drift rate, the drift segment must be identified from the RF-classified drift dives. We used a custom function written in the R Language where the first differential of depth was divided by the sampling rate. We calculated a threshold region for each drift dive where the differences between the differentials remained low and constant (Fig. 2). Drift rate was calculated as the difference between the start and end depth within each drift component divided by the difference between the start and end time of the component. We only

examined drift dives  $> 50$  m to minimise the potential buoyancy effects of residual air in the lungs, and we only used drifting periods lasting longer than 3 min because a minimum of 6 consecutive records is needed to estimate drift rate.

To compare the temporal patterns in drift rate, we fitted a median smoothing spline to the data using the constrained quadratic regression B-spline from the COBS library in the R Package to the drift rates derived from both approaches for each day of the foraging trip. Function roughness was controlled by the number of internal knots rather than the smoothing parameter  $\lambda$ . The number of knots was chosen via the automated knot-selection procedure which uses a stepwise knot deletion and addition process and then makes adaptive choices using the default Schwarz Information Criterion (SIC) (He and Ng, 2006). We used the probabilities derived from `randomForest` as the weightings when fitting the spline.

## 3. Results

Twenty-five foraging trip records (18 post-moult [PM] and 7 post-lactation [PL]) were used in the analysis. Eight of the seals equipped with VTDRs were not recaptured and another 11 instruments malfunctioned. Mean foraging trip duration for these seals was  $68.6 \pm 12.1$  and  $234.4 \pm 6.0$  days for PL and PM trips, respectively. Speed sensor failures occurred occasionally and predominantly at the start of the trip when turbines were temporarily blocked by accumulated beach sand. We identified blockages by examining the total mean speed for all dives (not just the drift component); when these were zero, it was clear the turbine was blocked. A small number of these blockages occurred during the trip, presumably as a result of flotsam or large particulates becoming lodged in the turbine. There was no evidence to suggest progressive clogging and it appeared that the few blockages that did occur cleared quickly. We omitted these periods of blockage from analysis.



**Fig. 4.** The proportion of time spent on the three most common dive types plotted against day of trip for one post-moult seal using the visually classified dives where the two human classifiers agreed (top), and for the random forests-classified dives (bottom). Solid light grey line = foraging dives (dive types D and E), dotted black line = drift dives (types C- and C+), and black line = round dives (type B). The two vertical black lines indicate the period of time that the seal was positively buoyant.

### 3.1. Visual classification

Disagreement between the classifiers was high for dive classes A (6 – 57%), B (10 – 31%) and C+ (14 – 46%). Lower disagreement was found for the remaining dive classes and disagreement with C<sup>-</sup> dives varied the least (10 – 11%) (Table 1; Fig. 1). The overall disagreement was 8%. When comparing the dataset where the two classifiers agreed to the drift dives validated by speed, C<sup>+</sup> dives were classified incorrectly in 1% of cases, and C<sup>-</sup> dives were classified incorrectly in 5% of cases.

### 3.2. Validation of classification function

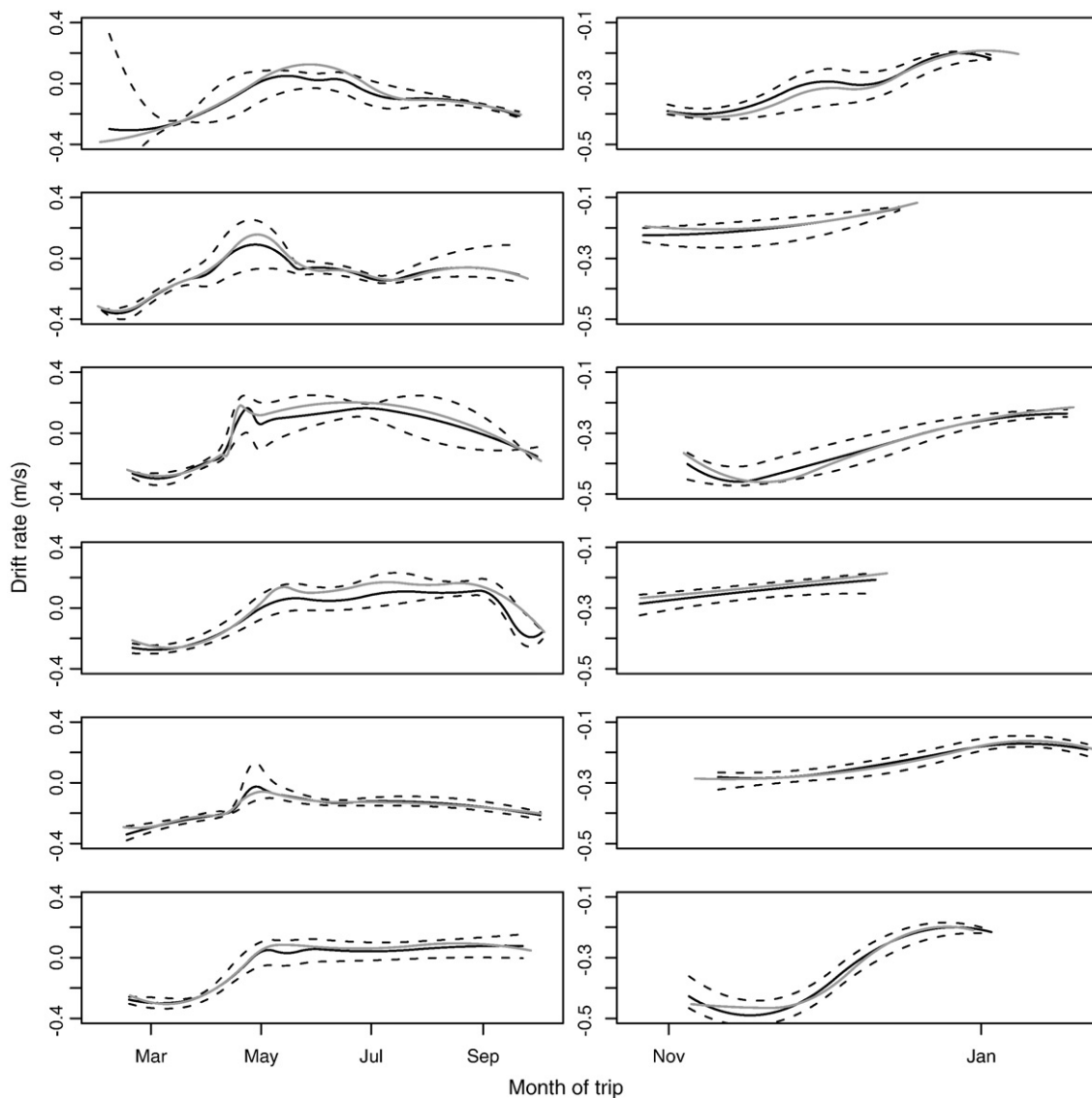
The overall OOB error estimate was 5% for both RF classification functions (with and without speed parameters), with little difference between them (Table 1). D dives had the greatest classification error (12%), followed by B dives (8%); the error for all other dive types <4%. Both the overall and individual dive classification errors were lower than for visual classification, although visual classification error for class D dives was lower (2 to 8%) than the RF classification function error (12%) (Table 1). True error for C<sup>+</sup> dives was 2%, and 4% for C<sup>-</sup> dives.

### 3.3. Importance of speed to the classification function

When included, speed-derived variables were the top-ranked predictors, particularly total mean speed and descent speed (Fig. 3A). Omitting speed-derived variables resulted in descent rate becoming the best predictor, although the relative importance of the most highly ranked versus subsequent variables was not as pronounced (Fig. 3B). In both RF classification functions, the proportion of time spent in the deepest depth bin also influenced classification success.

### 3.4. Validation with external dataset

Testing the RF classification function (without speed derived variables) on the PL dataset resulted in a 1.6% classification error, and no classification error on C<sup>-</sup> dives (Table 1). This seal did not perform any C<sup>+</sup> or E dives. It is common for post-lactation seals not to reach positive buoyancy and therefore not to perform C<sup>+</sup> dives (Crocker et al., 1997), and E dives are only exhibited by a small number of female elephant seals (they are normally associated with males) (Hindell et al., 1991).



**Fig. 5.** Median (black line) and 95% confidence intervals (black dashed lines) for the temporal spline function applied to the drift rates calculated from the true drift dives (speed on the drift segment  $\leq$  stall speed). Median predictions from the spline function applied to the drift rates calculated from the random forests-classified drift dives are shown as a grey line. Results for post-moult (left panels) and post-lactation (right panels) for 12 individual seals are shown.

### 3.5. Proportions of dive types found by visual versus RandomForest classification

Using the entire dataset (not just the training and validation datasets) where the two visual classifiers agreed, 84.1% of PM dives were square type dives (D and E) (Table 2). The next most common were B dives (6.9%), followed by C<sup>-</sup> dives (6.4%) (Table 2). In PL, a larger proportion of dives were D (92.7%), and C<sup>-</sup> dives were more common (4.7%) than B dives (2.2%). The proportion of RF-classified dives had reasonable agreement with the proportion of visually classified dives in PM for D, E and C<sup>-</sup> dives; however, the proportions calculated for the other dive classes were inflated, particularly for C<sup>+</sup> dives where RF classified more than triple the amount (3.9%) than were visually classified (1.3%) (Table 1). The proportions of RF-classified dives differed to the visually classified proportions during PM ( $\chi^2_3=3183.9$ ,  $P<0.001$ ), but there was little evidence that they differed during PL ( $\chi^2_3=5.32$ ,  $P=0.149$ ) (Table 2).

### 3.6. Time series of common dive classes

Most time spent on B dives occurred at the start and end of each of the foraging trips, but these dives also occurred intermittently throughout the trip (Fig. 4). Time spent on square dives (D and E) and drift dives (C<sup>+</sup> and C<sup>-</sup>) varied throughout the trip. The patterns were similar for both but for the RF classified dives the seal demonstrated a decrease in time spent on square dives and a corresponding increase in drift dives (Fig. 4). The temporal pattern in drift rates calculated from the RF classified drift dives was similar to or the same as the patterns from validated drift dives (Fig. 5), and all were within the 95% confidence interval of the predictions (Fig. 5).

## 4. Discussion

The classification of dive types essentially identifies different behavioural states of diving animals. Some of these have proven useful for inferring foraging (square dives) and inferring periods of successful feeding (drift dives). Analysing the fine-scale components of such dives can be useful for understanding foraging tactics and tracking these behaviours spatially is important for determining feeding foci. Our analysis has shown that the common method for conducting supervised dive classification, visual determination, may be prone to high errors. This is due to the inherent subjectivity exacerbated by a lack of clear category in which to place dives of intermediate shape. This is likely to result from dive types being inferred from 2-dimensional data, yet performed in 3-dimensional volume, or fine-scale changes in diving parameters that warp the shape of the profile. Internal validation of the classification function provides a more realistic estimate of the expected classification errors; we were successful in keeping error to around 5% for elephant seal dives, both with and without speed-derived variables. While not considerably different to the overall classification success using the visual technique (8%), an inspection of the error rates on the less common dive classes highlights the superiority of the automatic classification function (Table 1). Testing the classification function on a post-lactation dataset (external validation) produced a much lower classification error, both overall (1.6%) and for the individual dive classes (0–5%) (Table 1). This represents the true error rather than an estimate.

Even though speed variables were the most important predictors when available, their absence did not reduce the RF classification function's effectiveness greatly. There is an increasing range of parameters that can be measured by data-loggers attached to animals. For example flipper beat frequency (Hays et al., 2004b), state of consciousness (Houghton et al., 2008) and prey search and ingestion (Soto et al., 2008) can all be measured. These types of data will allow for further discrimination among dives and thus, a more informed

basis for ascribing behavioural function to dive profiles. In both functions the proportion of time spent in the deepest depth bin was an important predictor of dive type. This is due to the depth bins giving an overall impression of shape and the time spent in the deepest bin is the most characteristic of shape (Fig. 1).

The high disagreement between the two classifiers is indicative of the potential errors involved in visual classification. At best, these errors were as low as 2% (Table 1) and at worst, reached 57% (Table 1). This suggests that the use of one classifier in the development of supervised classification functions could lead to high errors in the training dataset. Our strategy of making a training dataset where both classifiers agreed reduces error greatly. Using our independent drift dive validation, the misclassification on these dives was only 1% for C<sup>+</sup>, and 5% for C<sup>-</sup> dives.

Discriminant function analysis is the most commonly used supervised classification procedure for this type of data. Past studies have estimated error at around 10% for five dive classes (Schreer and Testa, 1995), and 8% for six classes in Weddell seals (*Leptonychotes weddellii*) (Schreer and Testa, 1996), 8% for 7 dive classes in harbour seals (*Phoca vitulina*) (Baechler et al., 2002), 4% for five dive classes in grey seals (*Halichoerus grypus*) (Beck et al., 2000) and 7% for 5 classes in juvenile northern elephant seals (*M. angustirostris*) (Hassrick et al., 2007). Given that classification error rises with an increasing number of classes (Schreer and Testa, 1995), our RF classification function therefore produces errors that are similar to or better than the discriminant function. Given that the time-depth profiles of most diving animals are dominated by one dive class (Schreer et al., 2001), error applies mainly to the overrepresented dive class (Malcolm and Duffus, 2000). Thus, higher errors are likely for underrepresented dive classes based on discriminant functions. For example, Malcolm and Duffus (2000) found an overall error rate of 8.6% for 5 dive classes; however, the rarer dive classes had classification errors of 33 to 78%. In comparison, our RF classification function produced low error (2 to 8%) for even the rarest dive classes. This is a clear improvement over existing techniques especially with respect to subjective visual classification, which is still routinely used (e.g. Crocker et al., 2006; Elliott et al., 2008; Hassrick et al., 2007). Another strength of RF is that it assigns a probability to each dive classification – this is invaluable when incorporating dive data into state-space models of individual animal movement because accounting for the errors in observations allows for more meaningful biological inference (Patterson et al., 2008).

The highest error produced with the RF classification function was for D type dives during the post-moult foraging trip (12%). This may arise from fine-scale variation in aspects of the dive occurring over the course of the extended foraging trip. Indeed, descent and ascent rates are influenced by changes in buoyancy, and buoyancy can change monotonically during the foraging trip (Biuw et al., 2003; Thums et al., in press). Misclassification of these dives resulted in inflated proportions of the other dive classes predicted by the function, and D dives were most often misclassified as C dives. We observed that once seals reach positive buoyancy, they have a tendency to drift upwards on the depth oscillations at the bottom of D dives. Either this or some other effect of increasing buoyancy (i.e., lower descent rate) may result in these dives being classified as drift dives (Fig. 4) (Thums et al., in press). Drifting during such dives may permit the individual to maximise its time submerged by reducing energy (oxygen) expenditure associated with active propulsion. The relatively high error associated with classifying D dives may be due to the temporal shift in diving behaviour as seals become more buoyant, but may also result from the continuous gradient of diving behaviour that does not conform necessarily to discrete categories (Malcolm and Duffus, 2000; Schreer and Testa, 1995). There is also some question as to the validity of summarising 3-dimensional behaviour using only two dimensions (time and depth). Indeed, there is evidence demonstrating that behavioural interpretation of dive data can differ markedly based on



the number of dimensions examined (Davis et al., 2003; Harcourt et al., 2000).

This method can be applied to compressed dive data obtained from Satellite Relay Data Loggers (SRDLs) with overall classification error rates generally <10% (M. Biuw and colleagues, unpubl. data). These units relay positional and behavioural data which must be compressed due to the constrained bandwidth of the Argos satellite system (Myers et al., 2006). Data are frequently represented by a small number of time-depth points (usually 4–5) representing the inflection points where the dive trajectory changes most rapidly (Fedak et al., 2001).

Our comparison of the temporal patterns in drift rate from the RF-classified drift dives with the patterns from validated drift dives (Fig. 5) indicated the RF method's classification is valid. Even though some classification error persists, summarising patterns using splines or means provides a reasonable approximation of the underlying trend in the data. Thus, the use of RF provides a relatively precise, simple and fast method for classifying diving behaviour. Objectively distinguishing different dive types is also important for a wide range of air-breathing vertebrates including turtles, whales and birds (e.g. Baird et al., 2006; Elliott et al., 2008; Rice and Balazs, 2008), so we recommend using the RF approach for these taxa.

## Acknowledgements

We thank C. Holland and S. Wall for assistance conducting the visual dive classifications and data collection, and T. Patterson and M. Biuw for helpful discussion on analyses. M. Thums was supported by an Australian Post Graduate Award. [SS]

## References

- Asaga, T., Naito, Y., Le Boeuf, B.J., Sakurai, H., 1994. Functional analysis of dive types of female northern elephant seals. In: Le Boeuf, B.J., Laws, R.M. (Eds.), *Elephant Seals: Population Ecology, Behavior, and Physiology*. University of California Press, pp. 310–327.
- Austin, D., Bowen, W.D., McMillan, J.I., Iverson, S.J., 2006. Linking movement, diving, and habitat to foraging success in a large marine predator. *Ecology* [Ecology] 87, 3095–3108.
- Baechler, J., Beck, C.A., Bowen, W.D., 2002. Dive shapes reveal temporal changes in the foraging behaviour of different age and sex classes of harbour seals (*Phoca vitulina*). *Can. J. Zool.* 80, 1569–1577.
- Bailleul, F., Charrassin, J.-B., Monestiez, P., Roquet, F., Biuw, M., Guinet, C., 2007a. Successful foraging zones of southern elephant seals from Kerguelen Islands in relation to oceanographic conditions. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 362, 2169–2181.
- Bailleul, F., Charrassin, J.-B., Ezratty, R., Girard-Ardhuin, F., McMahon, C.R., Field, I.C., Guinet, C., 2007b. Southern elephant seals from Kerguelen Islands confronted by Antarctic Sea ice. Changes in movements and diving behaviour. *Deep Sea Res.* 54, 343–355.
- Baird, R.W., Webster, D.L., McSweeney, A.D.L., Schorr, G.S., Barlow, J., 2006. Diving behaviour of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales in Hawai'i. *Can. J. Zool.* 84, 1120–1128.
- Beck, C.A., Bowen, W.D., Iverson, S.J., 2000. Seasonal changes in buoyancy and diving behaviour of adult grey seals. *J. Exp. Biol.* 203, 2323–2330.
- Biuw, M., McConnell, B.J., Bradshaw, C.J.A., Burton, H.R., Fedak, M.A., 2003. Blubber and buoyancy: monitoring the body condition of free-ranging seals using simple dive characteristics. *J. Exp. Biol.* 206, 3405–3423.
- Biuw, M., Boehme, L., Guinet, C., Hindell, M.A., Costa, D., Charrassin, J.-B., Roquet, F., Bailleul, F., Meredith, M., Thorpe, S., Tremblay, Y., McDonald, B., Park, Y.-H., Rintoul, S., Bindoff, N., Goebel, M.E., Crocker, D., Lovell, P., Nicholson, J., Monks, F., Fedak, M., 2007. Variations in behaviour and condition of a Southern Ocean top predator in relation to in-situ oceanographic conditions. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13705–13710.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
- Breiman, L., 2002. Manual on setting up, using and understanding random forests V3.1.
- Burns, J.M., Schreer, J.F., Castellini, M.A., 1997. Physiological effects of dive patterns and foraging strategies in yearling Weddell seals (*Leptonychotes weddellii*). *Can. J. Zool.* 75, 1796–1810.
- Crocker, D.E., Le Boeuf, B.J., Costa, D.P., 1997. Drift diving in female northern elephant seals: implications for food processing. *Can. J. Zool.* 75, 27–39.
- Crocker, D.E., Costa, D.P., Le Boeuf, B.J., Webb, P.M., Houser, D.S., 2006. Impact of El Niño on the foraging behavior of female northern elephant seals. *Mar. Ecol. Prog. Ser.* 309, 1–10.
- Davis, R.W., Fuiman, L.A., Williams, T.M., Horning, M., Hagey, W., 2003. Classification of Weddell seal dives based on 3-dimensional movements and video-recorded observations. *Mar. Ecol. Prog. Ser.* 264, 109–122.
- Elliott, K.H., Woo, K., Gaston, A.J., Benvenuti, S., Dall'Antonia, L., Davoren, G.K., 2008. Seabird foraging behaviour indicates prey type. *Mar. Ecol. Prog. Ser.* 354, 289–303.
- Fedak, M.A., Lovell, P., Grant, S.M., 2001. Two approaches to compressing and interpreting time-depth information as collected by time-depth recorders and satellite-linked data recorders. *Mar. Mamm. Sci.* 17, 94–110.
- Field, I.C., Bradshaw, C.J.A., McMahon, C.R., Harrington, J., Burton, H.R., 2002. Effects of age, size and condition of elephant seals (*Mirounga leonina*) on their intravenous anaesthesia with tiletamine and zolazepam. *Vet. Rec.* 151, 235–240.
- Fletcher, S., Le Boeuf, B.J., Costa, D.P., Tyack, P.L., Blackwell, S.B., 1996. Onboard acoustic recording from diving northern elephant seals. *J. Acoust. Soc. Am.* 100, 2531–2539.
- Harcourt, R.G., Hindell, M.A., Bell, D.G., Waas, J.R., 2000. Three-dimensional dive profiles of free-ranging Weddell seals. *Polar Biol.* 23, 479–487.
- Hassrick, J.L., Crocker, D.E., Zeno, R.L., Blackwell, S.B., Costa, D.P., Le Boeuf, B.J., 2007. Swimming speed and foraging strategies of northern elephant seals. *Deep Sea Res.* 54, 369–383.
- Hays, G.C., Metcalfe, J.D., Walne, A.W., 2004a. The implications of lung regulated buoyancy control for dive depth and duration. *Ecology* 85, 1137–1145.
- Hays, G.C., Metcalfe, J.D., Walne, A.W., Wilson, R.P., 2004b. First records of flipper beat frequency during sea turtle diving. *J. Exp. Mar. Biol. Ecol.* 303, 243–260.
- He, X., Ng, P., 2006. COBS: Qualitatively Constrained Smoothing via Linear Programming. <http://www.cba.nau.edu/pin-ng/cobs.html>.
- Hindell, M.A., Slip, D.J., 1997. The importance of being fat: maternal expenditure in the southern elephant seal *Mirounga leonina*. In: Hindell, M., Kemper, C. (Eds.), *Marine mammal research in the Southern Hemisphere*. Surrey Beatty and Sons Pty Ltd., Chipping Norton, Australia, pp. 72–77.
- Hindell, M.A., Slip, D.J., Burton, H.R., 1991. The diving behaviour of adult male and female southern elephant seals, *Mirounga leonina* (Pinnipedia: Phocidae). *Aust. J. Zool.* 39, 595–619.
- Hochscheid, S., Bentivegna, F., Hays, G.C., 2005. First records of dive durations for a hibernating sea turtle. *Biol. Lett.* 1, 82–86.
- Hooker, S.K., Biuw, M., McConnell, B.J., Miller, P.J.O., Sparling, C.E., 2007. Bio-logging science: logging and relaying physical and biological data using animal-attached tags. *Deep Sea Res.* 54, 177–182.
- Houghton, J.D.R., Cedras, A., Myers, A.E., Liebsch, N., Metcalfe, J.D., Mortimer, J.A., Hays, G.C., 2008. Measuring the state of consciousness in a free-living diving sea turtle. *J. Exp. Mar. Biol. Ecol.* 356, 115–120.
- Jonker, F.C., Bester, M.N., 1994. The diving behaviour of adult southern elephant seal, *Mirounga leonina*, cows from Marion Island. *S. Afr. J. Antarct. Res. [S. Afr. J. ANTARCT. RES.]* 24, 1–2.
- Le Boeuf, B.J., Naito, Y., Asaga, T., Crocker, D.E., Costa, D.P., 1992. Swim speed in a female northern elephant seal: metabolic and foraging implications. *Can. J. Zool.* 70, 768–795.
- Le Boeuf, B.J., Crocker, D.E., Blackwell, S.B., Morris, P.A., Thorson, P.H., 1993. Sex differences in foraging in northern elephant seals. In: Boyd, I.L. (Ed.), *Marine Mammals: advances in behavioural and population biology*. Oxford University Press, London, pp. 149–178.
- Le Boeuf, B.J., Crocker, D.E., Costa, D.P., Blackwell, S.B., Webb, P.M., Houser, D.S., 2000. Foraging ecology of northern elephant seals. *Ecol. Monogr.* 70, 353–382.
- Lesage, V., Hammill, M.O., Kovacs, K.M., 1999. Functional classification of harbor seal (*Phoca vitulina*) dives using depth profiles, swimming velocity, and an index of foraging success. *Can. J. Zool.* 77, 74–87.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2/3, 18–22.
- Malcolm, C.D., Duffus, D.A., 2000. Comparison of subjective and statistical methods of dive classification using data from a time-depth recorder attached to a gray whale (*Eschrichtius robustus*). *J. Cetacean Res. Manag.* 2, 177–182.
- McMahon, C.R., Bradshaw, C.J.A., Hays, G.C., 2007. Satellite tracking reveals unusual diving characteristics for a marine reptile, the olive ridley turtle *Lepidochelys olivacea*. *Mar. Ecol. Prog. Ser.* 329, 239–252.
- Myers, A.E., Lovell, P., Hays, G.C., 2006. Tools for studying animal behaviour: validation of dive profiles relayed via the Argos satellite system. *Anim. Behav.* 71, 989–993.
- Nowacek, D.P., Johnson, M.P., Tyack, P.L., Shorter, K.A., McLellan, W.A., Pabst, D.A., 2001. Buoyant balaenids: the ups and downs of buoyancy in right whales. *Proc. R. Soc. Lond., B Biol. Sci.* 268, 1811–1816.
- Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State-space models of individual animal movement. *Trends Ecol. Evol.* 23, 87–94.
- Putz, K., Cherel, Y., 2005. The diving behaviour of brooding king penguins (*Aptenodytes patagonicus*) from the Falkland Islands: variation in dive profiles and synchronous underwater swimming provide new insights into their foraging strategies. *Mar. Biol.* 147, 281–290.
- R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rice, M.R., Balazs, G.H., 2008. Diving behavior of the Hawaiian green turtle (*Chelonia mydas*) during oceanic migrations. *J. Exp. Mar. Biol. Ecol.* 356, 121–127.
- Robinson, P.W., Tremblay, Y., Crocker, D.E., Kappes, M.A., Kuhn, C.E., Shaffer, S.A., Simmons, S.E., Costa, D.P., 2007. A comparison of indirect measures of feeding behaviour based on ARGOS tracking data. *Deep Sea Res.* 54, 356–368.
- Schreer, J.F., Testa, J.W., 1995. Statistical classification of diving behavior. *Mar. Mamm. Sci.* 11, 85–93.
- Schreer, J.F., Testa, J.W., 1996. Classification of Weddell seal diving behavior. *Mar. Mamm. Sci.* 12, 227–250.
- Schreer, J.F., Kovacs, K.M., Hines, R.J.O., 2001. Comparative diving patterns of pinnipeds and seabirds. *Ecol. Monogr.* 71, 137–162.
- Simeone, A., Wilson, R.P., 2003. In-depth studies of Magellanic penguin (*Spheniscus magellanicus*) foraging: can we estimate prey consumption by perturbations in the dive profile? *Mar. Biol.* 143, 825–831.

- Soto, N.A., Johnson, M.P., Madsen, P.T., Diaz, F., Dominguez, I., Brito, A., Tyack, P., 2008. Cheetahs of the deep sea: deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands). *J. Anim. Ecol.* doi:10.1111/j.1365-2656.2008.01393.x.
- Thums, M., Bradshaw, C.J., Hindell, M.A., in press. Tracking changes in relative body composition of southern elephant seals using swim speed data. *Mar. Ecol. Prog. Ser.*
- Webb, P.M., Crocker, D.E., Blackwell, S.B., Costa, D.P., Leboeuf, B.J., 1998. Effects of buoyancy on the diving behavior of northern elephant seals. *J. Exp. Biol.* 201, 2349–2358.
- Wilson, R.P., 1995. Foraging ecology. In: Williams, T.D. (Ed.), *The penguins*. Oxford University Press, Oxford, pp. 81–106.
- Wilson, R.P., Wilson, M.-P.T., 1990. Foraging ecology of breeding *Spheniscus* penguins. In: Davis, L.S., Darby, J.T. (Eds.), *PENGUIN BIOLOGY*. Academic, San-Diego, pp. 181–206.
- Wilson, R.P., Pütz, K., Grémillet, D., Culik, B.M., Kierspel, M., Regel, J., Bost, C.A., Lage, J., Cooper, J., 1995. Reliability of stomach temperature changes in determining feeding characteristics of seabirds. *J. Exp. Biol.* 198, 1115–1135.