The following supplement accompanies the article

Diving classification and behavior of free-ranging female southern elephant seals based on three-dimensional movements and video-recorded observations

McGovern KA*, Rodríguez DH, Lewis MN, Davis RW

*Corresponding author: krismcgov@gmail.com

Marine Ecology Progress Series 620: 215–232 (2019)

Supplement 1

1. Individual model development

To select the initial variables to use in LDA and QDA, several transformations (cubed, squared, log, square root, exponent, third[1]) for each descriptor were computed, standardized, and added to the predictor matrix. This was done to increase the potential for variables to be included in the model, as both LDA and QDA assume multivariate normality, and transformations can sometimes render non-normal data normal. The Shapiro-Wilks test in R was used to select variables that had a normal distribution (p-value > 0.05) for each of the three classes. These multivariate normal data were then entered into an initial feature selection process for use in QDA. LDA makes an additional assumption of equal covariance matrices between classes. The Bartlett test was used in R to select variables that met this assumption (p-value > 0.05). The resulting data was input into an initial feature selection process for use in LDA. RFA makes no assumptions regarding multivariate normality or equal covariance, so the original predictors were entered into the initial feature selection process for use in RFA.

[1] The third transformation raises the value to the 1/3 power.

The initial feature selection process was identical for each of the supervised classification models. A loop was written that incorporated the recursive feature elimination function from the caret package in R, which selected the predictors that contributed most to the separation between classes using 10-fold cross-validation. This function produced a list of predictors from most influential to least influential, according to how much each predictor contributed to variance between classes. During the first iteration of the loop, the most significant predictor was selected and all collinear variables were eliminated from the dataset. During each successive iteration of the loop, the next most significant predictor was selected and all collinear variables were eliminated. This process was repeated until there were no collinear variables present in the dataset. This process was completed for each dataset (original dataset for RFA, multivariate normal dataset for QDA, multivariate normal with equal covariance dataset for LDA) before training all classifiers.

To develop the LDA and QDA classifiers, the corresponding selected predictors were entered into a stepwise discriminant analysis (direction = both) using the stepclass function in the klaR package in R (Weihs, Ligges, Luebke, & Raabe, 2005) to determine the optimal predictors to include in each model. LDA or QDA was specified in the function input. The selected predictors for each were then used to develop the corresponding classifier using the MASS package in R using jackknived (leave-one-out) cross-validation (Venables & Ripley, 2013). The LDA was also fit in SPSS (IBM Corp, 2016) to calculate the eigenvalues and standardized canonical coefficients and to test the discriminant functions.

To develop the RFA classifier, the rfe function in the caret package in R was used to determine the optimum number of predictors, ranging from sizes 1:30. The outer resampling method used was repeated (10 repeats) 10-fold cross-validation. The number of predictors chosen for use in the classifier was determined based on Accuracy and the Kappa statistic. To tune the values for mtry (number of predictors sampled at each node split) and ntree (number of trees in forest), 150 random forest models were fit to the data using the randomforest package in R (Liaw & Wiener, 2002) with mtry = 1:15 and ntree = 50, 100, 200, 300, 400, 500, 1000, 1500, 2000, and 2500. To avoid overfitting the classifier to the training data, each tree in each forest was fit with 13 (equal to 70% of the dives in the training set for the class with the smallest sample size) randomly selected and bootstrapped samples from each class. The out-of-bag (OOB) error estimate was calculated for each model by using each tree in the forest to predict dive class for the dives that were not used to train that particular tree, then averaging the prediction error across all of the trees in the forest. The OOB error rate was subtracted from 1 to obtain an OOB correct classification rate.

The kmeans cluster analysis was completed using the kmeans function in R. The elbow method and silhouette method were both used to determine the optimal number of clusters. The input variables used for the random forest model were used for the kmeans analysis, as kmeans also does not make any assumptions based on normality or equality of covariance matrices. The resulting clusters were labeled as “foraging”, “resting”, or “transit” based on the primary class of video dives occurring in each cluster.

2. Individual model results

2.1 Linear discriminant analysis

Five predictors were selected by the stepclass function to develop the LDA classifier. The predictors chosen were transformations of 1) speed variance during the first half of the dive (in 4 Hz, speedcalcsmoothvar_4firsthalf), 2) stroking rate variance during the first half of the dive (varstrokingratefirsthalf), 3) rate of change in the y-axis accelerometer at the bottom of the dive (in 16 Hz, yaccelROC_16bottom), 4) variance in the x-axis accelerometer during the first half of the dive (in 16 Hz, xaccelvar_16firsthalf), and stroking rate variance during the second half of the dive (varstrokingratesecondhalf) (Table S1). The coefficients of the linear discriminants are listed in Table S2. Fig. S1 depicts the decision boundaries for the resulting LDA classifier. The classifier correctly classified 76.9% of the training data (89.8%, 58.8%, and 57.1% of foraging, resting, and transit dives, respectively) (Table S3). The eigenvalues of the first and second discriminant functions were 0.421 and 0.298, respectively (Table S4). The proportion of the trace explained by the first and second discriminant functions was 58.5% and 41.5%, respectively. Chi-square tests for the first and second discriminant functions together and the second discriminant function solo were both significant at the $p<0.001$ significance level (χ^2 statistic of 86.9 and 37, respectively) (Table S5). A plot of the canonical discriminant function coefficients standardized from 0 to 1 showed that varstrokingratefirsthalf contributed the most to the separation of classes on discriminant axis 1, and that yaccelROC_16bottom contributed most to the separation of classes on discriminant axis 2 (Fig. S2). The LDA classifier was tested on the test dataset with an overall accuracy of 0.803 (95% CI: 0.687, 0.891) and a Kappa statistic of 0.638 (Table S1). The classification had an overall accuracy significantly better than the no information rate (NIR) at the 0.01 level with a p-value of 0.0002. Balanced accuracy was 0.825, 0.804, and 0.805 for foraging, transit, and resting, respectively (Table S3).

Table S1. Group means for LDA (linear discriminant analysis) variables.

	Foraging	Resting	Transit
sqrt(varstrokingratefirsthalf)	0.13 ± 0.68	-0.92 ± 0.99	-0.05 ± 0.79
third(xaccelvar_16firsthalf)	0.02 ± 0.64	-0.20 ± 0.63	-0.60 ± 0.63
log(yaccelROC_16bottom)	0.35 ± 0.79	-0.59 ± 0.87	-0.26 ± 0.71
third(speedcalcvar_4firsthalf)	0.17 ± 0.99	0.74 ± 0.61	-0.52 ± 0.86
third(varstrokingratesecondhalf)	0.31 ± 0.89	-0.12 ± 0.71	0.29 ± 1.02

Table S2. Coefficients of linear discriminants for LDA (linear discriminant analysis) variables.

	LD1	LD2
sqrt(varstrokingratefirsthalf)	-1.32	0.26
third(xaccelvar_16firsthalf)	0.90	0.29
log(yaccelROC_16bottom)	0.71	1.03
third(speedcalcsmoothvar_4firsthalf)	0.65	-0.25
third(varstrokingratesecondhalf)	0.36	0.38

Table S3. LDA (linear discriminant analysis) classification performance on training dataset.

	Prediction			
Reference	Foraging	Resting	Transit	Correctly classified
Foraging	79	0	9	0.898
Resting	5	10	2	0.588
Transit	18	0	24	0.571

Table S4. Eigenvalues and % variance of first two linear discriminant functions.

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.421a	58.5	58.5	0.544
2	.298a	41.5	100	0.479

Table S5. Test of linear discriminant functions.

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	0.542	86.911	10	0
2	0.77	37.04	4	0

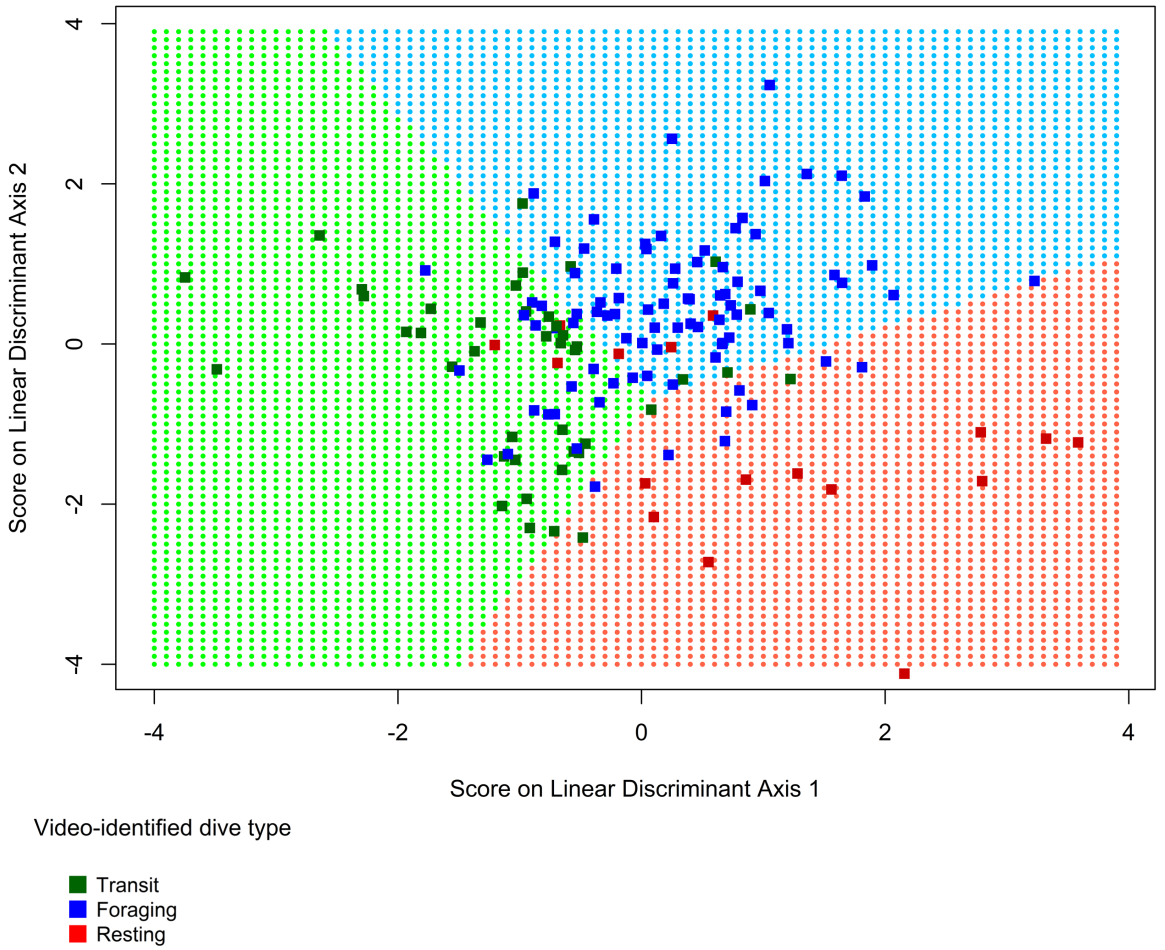

Figure S1. Decision boundaries for LDA (linear discriminant analysis) classifier, overlaid with training dataset color coded by known dive type. Correctly classified dives are located in the same color region (e.g. red on red for resting dives).

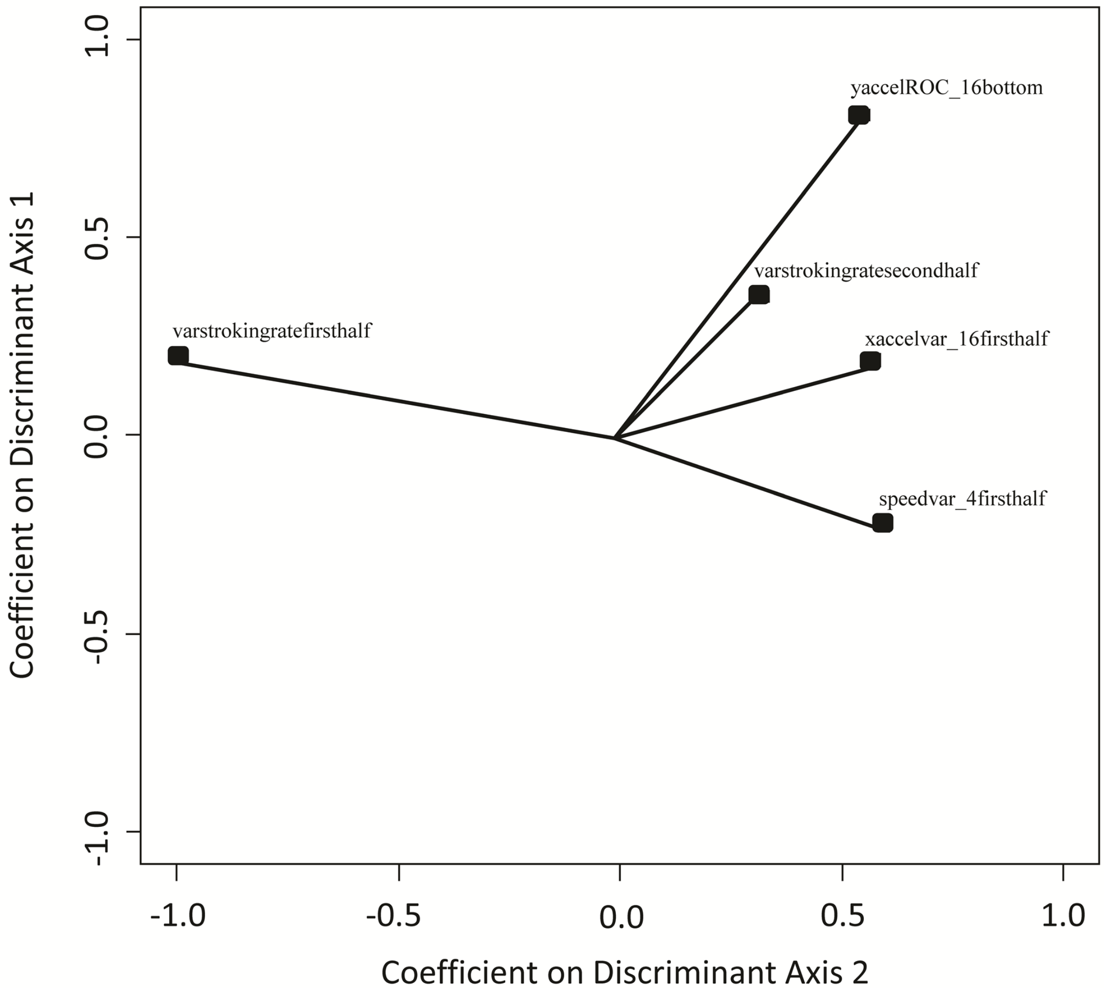

Figure S2. Plot of standardized canonical linear discriminant function coefficients.

2.2 Quadratic discriminant analysis

Three predictors were selected by the stepclass function to train the QDA classifier. The predictors chosen were transformations of: 1) variance in the x-axis accelerometer during the bottom phase of the dive (in 16 Hz, xaccelvar_16bottom), 2) variance in the y-axis accelerometer (in 16 Hz, yaccelvar_16), and 3) rate of change in the z-axis accelerometer (in 1 Hz, zaccelROC). Group means for each of the predictors are listed in Table S6. The classifier correctly classified 87% of the training data (95.5%, 65.5%, and 78.6% of foraging, resting, and transit dives, respectively (Table S7). Fig. S3 depicts the QDA classification as a 3D plot of the three predictor variables. The 3D plot is shown as three biplots to better visualize the decision boundaries of the classifier. The larger spheres are the video-identified dives used to train the classifier. The smaller spheres are the remainder of the dataset, colored by predicted dive type as classified by the QDA classifier. The QDA classifier was tested on the test dataset with an overall accuracy of 0.864 (95% CI: 0.757, 0.936) and a Kappa statistic of 0.755 (Table 2.1). The classification had an overall accuracy significantly better than the NIR at the 0.01 level with a p-value of 1.4e-06. Balanced accuracy was 0.906, 0.92, and 0.826 for foraging, transit, and resting, respectively (Table 2.3).

Table S6. Group means for QDA (quadratic discriminant analysis) variables.

	Foraging	Resting	Transit
log(xaccelvar_16bottom)	0.32	-0.80	-0.54
log(zaccelROC)	0.20	-0.55	-1.05
third(yaccelvar_16)	0.09	0.66	-0.65

Table S7. Performance of QDA (quadratic discriminant analysis) on training dataset.

	Prediction			
Reference	Foraging	Resting	Transit	Correctly classified
Foraging	84	0	4	0.955
Resting	1	11	5	0.647
Transit	9	0	33	0.786

Table S8. Performance of random forest model on training set (OOB).

	Prediction			
Reference	Foraging	Resting	Transit	Correctly classified (OOB)
Foraging	82	0	6	0.93
Resting	0	16	1	0.94
Transit	5	0	37	0.88

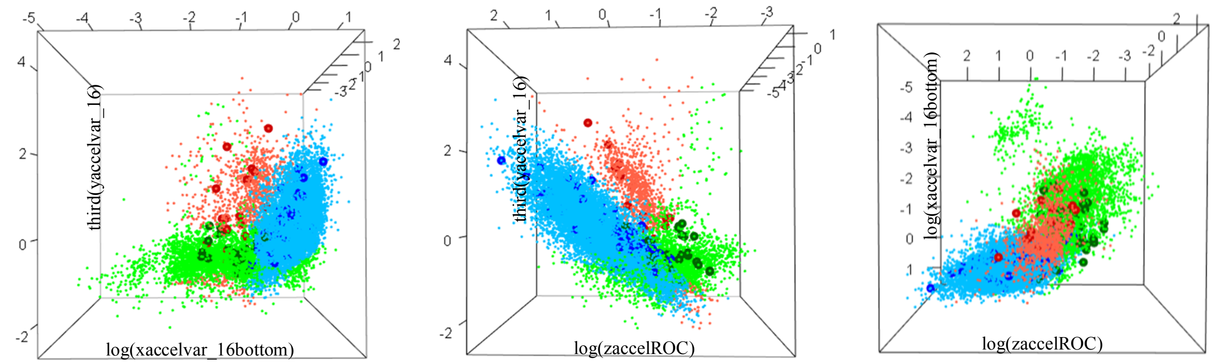

Figure S3. 3D plot of QDA (quadratic discriminant analysis) classification with all three predictors, viewed as three biplots. Large spheres are video-identified dives used to train the model; smaller spheres are the remainder of the dives, colored according to QDA classifier-predicted dive type.

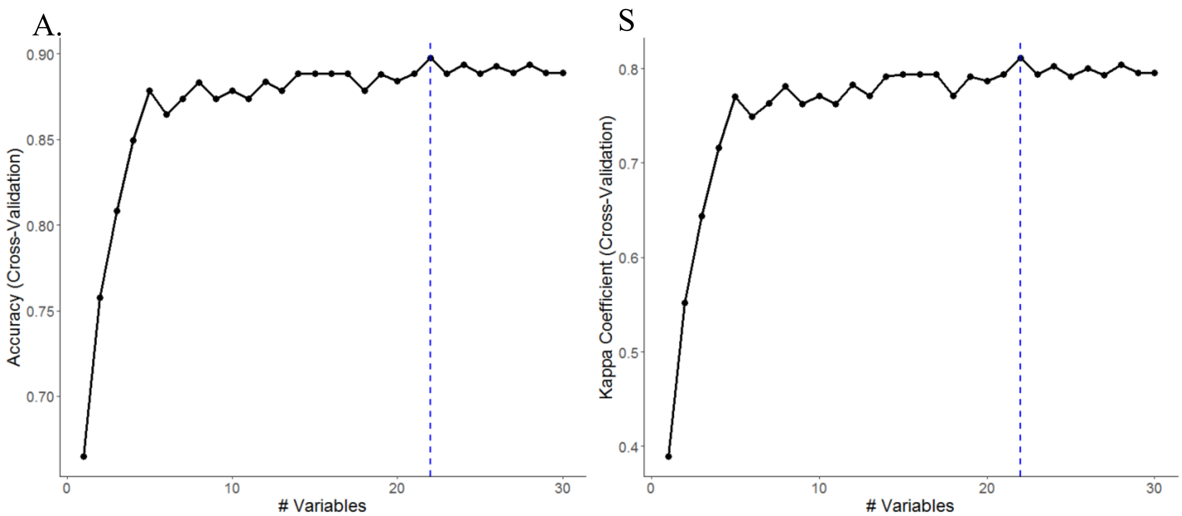

Figure S4. Plot of A. Accuracy and B. Kappa coefficient by number of predictors in random forest model. Blue line indicates optimal number of predictors for each metric.

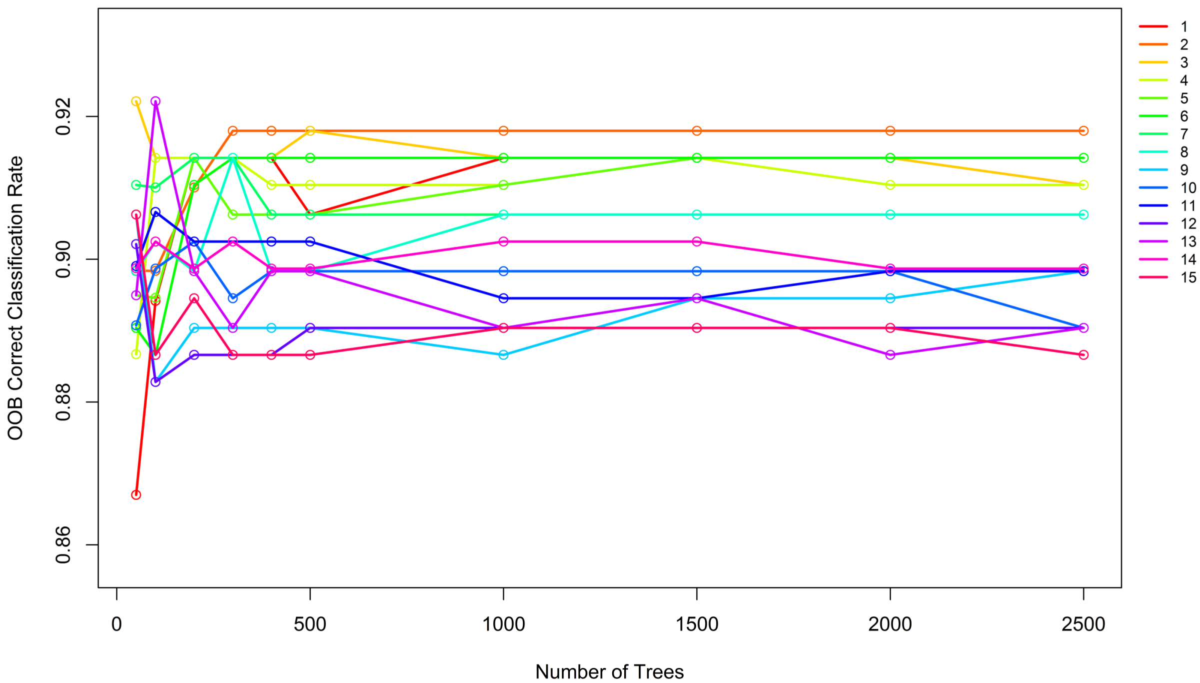

Figure S5. Out of Bag Correct Classification Rate for varying values of ntree and mtry for use in tuning the random forest model.

2.3 Random forest classification

Accuracy and the Kappa statistic indicated that 22 predictors was the optimal number of predictors to include in the classifier (Fig. S4). An mtry of 2 performed consistently well across ntree values of 400:2500, with an OOB correct classification rate of 0.918 (Fig. S5). An mtry of 2 and ntree of 1500 (mid-range of well-performing values) were selected for use in the final model. The selected 22 variables were input into the randomforest package in R with these parameters to train the RFA classifier. The OOB estimate of error rate for the training set in the random forest classification model was 8.16%. The OOB error rate by class was 0.068, 0.059, and 0.119 for foraging, resting, and transit, respectively (Table S8). The 22 variables included in the random forest model are listed from most important to least important in Fig. S6, according to the mean decrease in the Gini coefficient. The random forest model was tested on the test dataset with an overall accuracy of 0.909 (95% CI: 0.813, 0.966) and a Kappa statistic of 0.842 (Table 2.1). The random forest classification had an overall accuracy significantly better than the NIR at the 0.01 level with a p-value of 9.7e-09. Balanced accuracy was 0.94 for each of the three classes (Table 2.3).

2.4 KMeans cluster Analysis

Fig. S7 depicts the results of the kmeans cluster analysis plotted on the first two principal components overlaid with the training set dives. The elbow and silhouette method both indicated that the optimal number of clusters for kmeans cluster analysis was k=3 (Figs. S8, S9). The first two principal components explain 52.6% of the point variability (Fig. S10). Based on the percent of known dive classes that were assigned to each cluster, the cluster classes were defined as "transit", "foraging", and "resting", for clusters 1, 2, and 3, respectively. Classification was tested using all of the video dives with an overall accuracy of 0.812 (95% CI: 0.753, 0.862) and a Kappa statistic of 0.667 (Table 2.1). The kmeans cluster analysis assigned clusters had an overall accuracy significantly better than the NIR at the 0.01 level with a p-value of 1.9e-06. Balanced accuracy was 0.871, 0.74, and 0.864 for foraging, resting, and transit, respectively (Table 2.3).

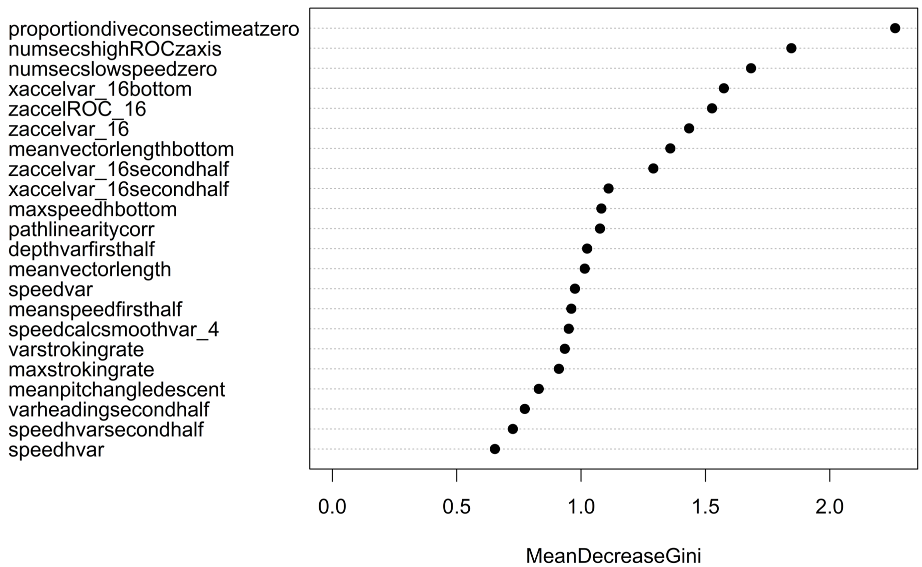

Figure S6. Mean decrease in the Gini coefficient (a relative measure of how great of a role a predictor plays in separating the data into classes) for predictors in the random forest model.

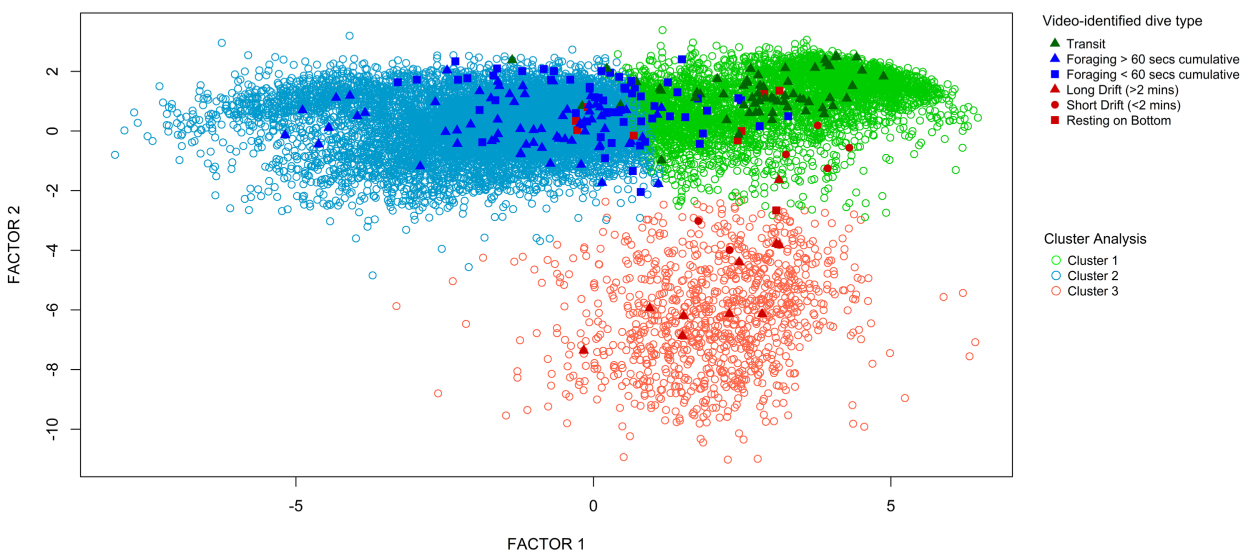

Figure S7. KMeans cluster analysis results plotted on principal components 1 and 2, overlaid with known video-recorded dive types.

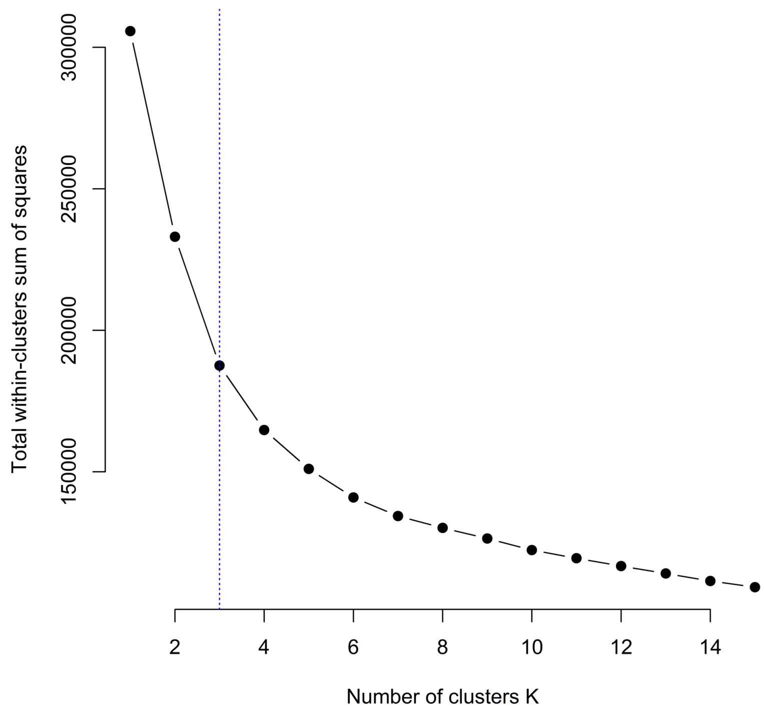

Figure S8. Elbow plot for kmeans analysis. Blue line demarcates optimal number of clusters.

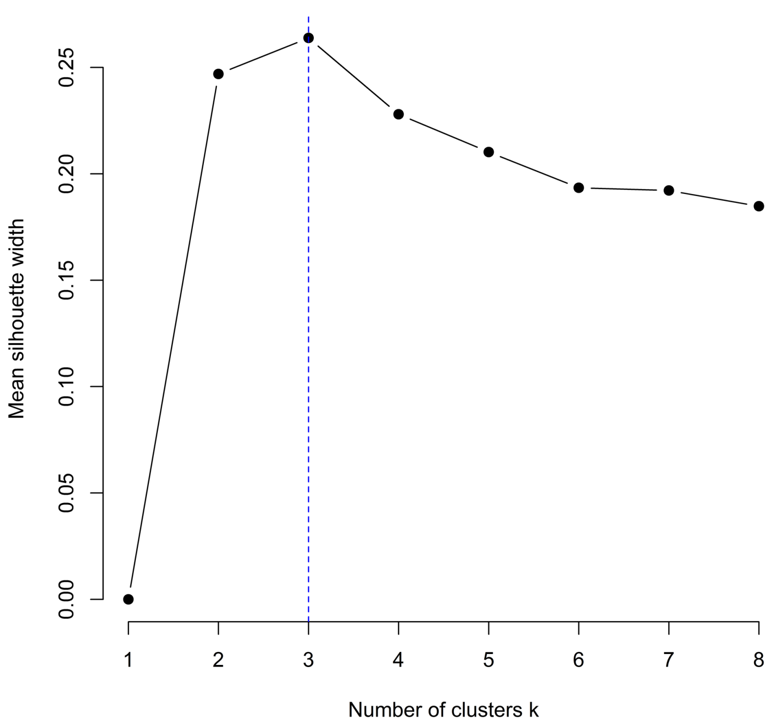

Figure S9. Silhouette width plot for kmeans analysis. Blue line demarcates optimal number of clusters.

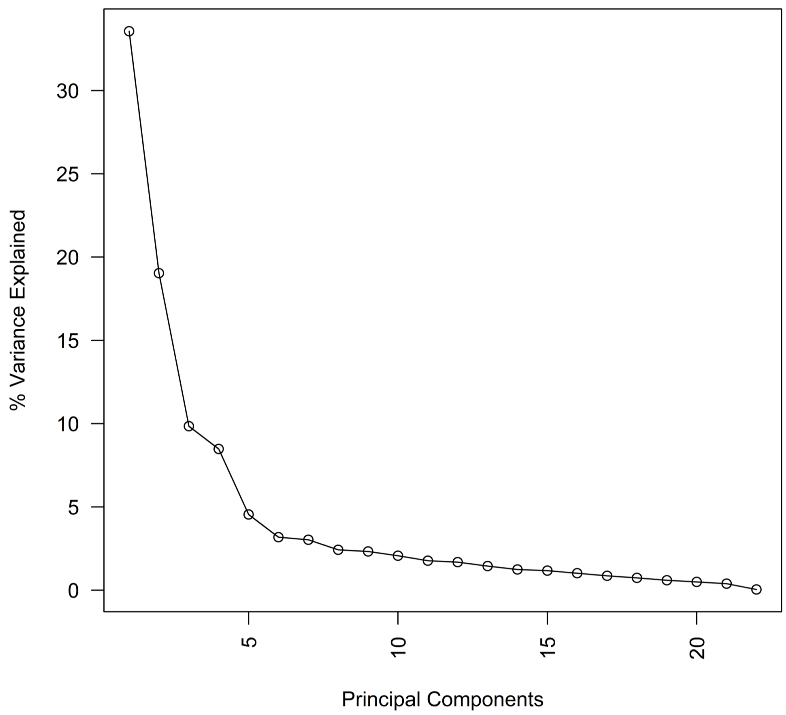

Figure S10. Contribution of principal components to point variability.

3. Model comparison

Table S9. Accuracy and Kappa statistic for each model (* indicates that the accuracy was significantly better than the no information rate at a significance level of <0.01; LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, RFA: random forest analysis, Cluster: cluster analysis).

	Accuracy (95% CI)	Kappa statistic
LDA	0.803 (0.687, 0.891)*	0.638
QDA	0.864 (0.757, 0.936)*	0.755
RFA	0.909 (0.813, 0.966)*	0.842
Cluster	0.812 (0.753, 0.862)*	0.667

Table S10. Predictive model measures averaged across the three classes for each model (LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, RFA: random forest analysis, Cluster: cluster

	Cluster	LDA	QDA	RFA
Sensitivity	0.75	0.74	0.85	0.92
Specificity	0.90	0.88	0.92	0.96
Positive Predictive Value	0.86	0.79	0.83	0.92
Negative Predictive Value	0.89	0.88	0.92	0.94
Balanced Accuracy	0.83	0.81	0.88	0.94

Table S11. Predictive model measures by class (LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, RF: random forest analysis, Cluster: cluster analysis).

	Foraging				**Resting**				**Transit**			
	Cluster	LDA	QDA	RFA	Cluster	LDA	QDA	RFA	Cluster	LDA	QDA	RFA
Sensitivity	0.811	0.872	0.923	0.872	0.480	0.625	0.875	0.875	0.951	0.737	0.737	1.000
Specificity	0.930	0.778	0.889	1.000	1.000	0.983	0.966	1.000	0.776	0.872	0.915	0.872
Positive Predictive Value	0.945	0.850	0.923	1.000	1.000	0.833	0.778	1.000	0.630	0.700	0.778	0.760
Negative Predictive Value	0.769	0.808	0.889	0.844	0.935	0.950	0.983	0.983	0.975	0.891	0.896	1.000
Balanced Accuracy	0.871	0.825	0.906	0.936	0.740	0.804	0.920	0.938	0.864	0.805	0.826	0.936

Table S12. AUC calculations by class for all models (LDA: linear discriminant analysis, QDA: quadratic discriminant analysis, RF: random forest analysis, Cluster: cluster analysis).

	Foraging	Transit	Resting	Mean
RFA	0.978	0.975	0.998	0.984
QDA	0.961	0.941	0.972	0.958
LDA	0.937	0.921	0.914	0.924
Cluster	0.882	0.935	0.661	0.826

4. Dive descriptors.

Table S13. Dive descriptors. + indicates that the descriptor was calculated for the corresponding portion of the dive. A=ascent, B=bottom, D=descent, F=first half, S=second half, W=whole). Descriptors are calculated based on 1 Hz data unless otherwise specified. All distances were estimated using corrected coordinates. Rate of change (ROC) is calculated per second unless otherwise indicated.

Descriptor (R)	**Descriptor full name**	**A**	**B**	**D**	**F**	**S**	**W**
Depthvar	Depth variance	+	+	+	+	+	+
Diveduration	Dive duration (min)						+
headingROC	Mean heading rate of change	+	+	+	+	+	+
Headingvar	Heading variance	+	+	+	+	+	+
Horzpathlinearity	Horizontal path linearity				+	+	+
Maxconsectimeatzero	Maximum consecutive number of seconds speed =0 m/sec						+
maxconsectimeunderpt3	Maximum consecutive number of seconds speed <0.3 m/sec						+
maxconsectimeunderpt5	Maximum consecutive number of seconds speed <0.5 m sec^{-1}						+
Maxdepth	Maximum depth (m)	+	+	+	+	+	+
Maxspeed	Maximum speed (m sec^{-1})	+	+	+	+	+	+
Maxspeedh	Maximum horizontal speed (m sec^{-1})	+	+	+	+	+	+
Maxstrokingrate	Maximum stroking rate (strokes sec^{-1})	+	+	+	+	+	+
Meanpitchangle	Mean pitch angle (°)	+		+			
Meanrollangle	Mean roll angle (°)	+		+			
Meanspeed	Mean speed (m sec^{-1})	+	+	+	+	+	+
Meanspeedh	Mean horizontal speed (m sec^{-1})	+	+	+	+	+	+
Meanstrokingrate	Mean stroking rate (strokes sec^{-1})	+	+	+	+	+	+
Meanvectorlength	Mean vector length (measure of angular dispersion; 0=uniform , 1=none)	+	+	+	+	+	+
Meanverticalspeed	Mean vertical speed (m sec^{-1})	+		+			
Minspeedh	Minimum horizontal speed (m sec^{-1})	+	+	+	+	+	+
numsecshighROCzaxis	Number of seconds normalized rate of change in z-axis accelerometer >0.2						+
numsecslowspeedpt3	Number of seconds speed <0.3 m sec^{-1}						+
numsecslowspeedpt5	Number of seconds speed <0.5 m sec^{-1}						+
numsecslowspeedzero	Number of seconds speed =0 m sec^{-1}						+
Pathlinearity	Path linearity (3D)				+	+	+
proportiondiveconsectimeatzero	maxconsectimeatzero $diveduration^{-1}$						+
proportiondiveconsectimeunderpt3	maxconsectimeunderpt3 $diveduration^{-1}$						+
proportiondiveconsectimeunderpt5	maxconsectimeunderpt5 $diveduration^{-1}$						+
proportiontimehighROCzaxis	numsecshighROCzaxis $diveduration^{-1}$						+
proportiontimelowspeedpt3	numsecslowspeedpt3 $diveduration^{-1}$						+
proportiontimelowspeedpt5	numsecslowspeedpt5 $diveduration^{-1}$						+
proportiontimelowspeedzero	numsecslowspeedzero $diveduration^{-1}$						+
speedcalcROC_4	Mean speed (4 Hz) rate of change	+	+	+	+	+	+
speedcalcvar_4	Speed (4 Hz) variance	+	+	+	+	+	+
Speedhvar	Horizontal speed variance	+	+	+	+	+	+
speedROC	Mean speed rate of change	+	+	+	+	+	+
Speedvar	Speed variance	+	+	+	+	+	+
Straightlinehorzdist	Straight-line distance (m)				+	+	+
strokingROC	Mean stroking rate rate of change	+	+	+	+	+	+
totaldist3D	Total 3D distance (m)				+	+	+
Totalhorzdist	Total 2D distance (m)				+	+	+
Varpitchangle	Pitch angle variance	+		+			
Varrollangle	Roll angle variance	+		+			
Varstrokingrate	Stroking rate variance	+	+	+	+	+	+
Varverticalspeed	Vertical speed variance	+		+			
Vertpathlinearity	Vertical path linearity				+	+	+
Xaccelvar	X-axis accelerometer variance	+	+	+	+	+	+
xaccelvar_16	X-axis accelerometer (16 Hz) variance	+	+	+	+	+	+
yaccelROC_16	Mean y-axis accelerometer (16 Hz) rate of change	+	+	+	+	+	+
Yaccelvar	Y-axis accelerometer variance	+	+	+	+	+	+
yaccelvar_16	Y-axis accelerometer (16 Hz) variance	+	+	+	+	+	+
zaccelROC	Mean z-axis accelerometer rate of change	+	+	+	+	+	+
zaccelROC_16	Mean z-axis accelerometer (16 Hz) rate of change	+	+	+	+	+	+
Zaccelvar	Z-axis accelerometer variance	+	+	+	+	+	+