Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

"M2B" package in R: Deriving multiple variables from movement data to predict behavioural states with random forests

Andréa Thiebault[1] | Laurent Dubroca[2] | Ralf H.E. Mullers[3] | Yann Tremblay[4] | Pierre A. Pistorius[5]

[1]Department of Zoology, Nelson Mandela University, Port Elizabeth, South Africa

[2]Datacall Response Unit (CREDO), IFREMER, Port-en-Bessin-Huppain, France

[3]Department of Biodiversity, University of Limpopo, Sovenga, South Africa

[4]Institut de Recherche pour le Développement, UMR MARBEC 248: Marine Biodiversity, Exploitation and Conservation, Sète, France

[5]DST/NRF Centre of Excellence at the Percy FitzPatrick Institute, Department of Zoology, Nelson Mandela University, Port Elizabeth, South Africa

Correspondence
Andréa Thiebault
Email: andrea.thiebault@gmail.com

Abstract

1. The behaviour of individuals affect their distributions and is therefore fundamental in determining ecological patterns. While, the direct observation of behaviour is often limited due to logistical constraints, collection of movement data has been greatly facilitated through the development of bio-logging. Movement data obtained through tracking instrumentation may potentially constitute a relevant proxy to infer behaviour.

2. To infer behaviour from movement data is a key focus within the "movement ecology" discipline. Statistical learning constitutes a number of methods that can be used to assess the link between given variables from a fully informed training dataset and then predict the values on a non-informed variable. We chose the random forest algorithm for its high prediction accuracy and its ease of implementation. The strength of random forest partly lies in its ability to handle a very large number of variables. Our methodology is accordingly based on the derivation of multiple predictor variables from movement data over various temporal scales, to capture as much information as possible from changes and variations in movement.

3. The methodology is described in four steps, using examples on foraging seabirds and fishing vessels for illustration. The models showed very high prediction accuracy (92%–97%), thereby confirming the influence of behaviour on movement decisions and demonstrating the ability to derive multiple variables from movement data to predict behaviour with random forests.

4. The codes developed for this methodology are published in the "M2B" (Movement to Behaviour) R package, available at https://CRAN.R-project.org/package=m2b. They can be used and adapted to datasets where movement was sampled from a wide range of taxa, sampling schemes or tracking devices. Observations are needed for a subset of the data, but once the model is trained, it can be used on any dataset with similar movement data.

KEYWORDS
Cape gannet, fisheries, GPS, local enhancement, machine learning, onboard observers, social interactions, video cameras

1 | INTRODUCTION

Behaviour significantly shapes movement patterns of individuals (Nathan et al., 2008; Tremblay, Thiebault, Mullers, & Pistorius, 2014). Because direct observations of behaviour can be challenging (difficult accessibility, restricted time), movement data is often easier to collect due to recent advances in bio-logging science (Boyd, Kato, & Ropert-Coudert, 2004). Movement may hence constitute a valuable proxy to infer behaviour, provided a robust link can be identified between the two (Rutz & Hays, 2009).

To build a link between categorical observations and local movement information, a wide range of methods is available. The field of statistical learning encompasses methodologies such as regressions, decision trees, support vector machines, neural networks or random forests (Hastie, Tibshirani, & Friedman, 2009). These methods extract information from the structure of a dataset (distribution of variables and relationship among variables) that may then be used for predictive purposes. Statistical learning algorithms can be used to (1) assess the link between behavioural and movement data from a dataset with observations and (2) predict behaviour in similar datasets where direct observations were not available.

Among the set of statistical learning algorithms, random forests seem particularly suitable for behavioural data. They perform well on unbalanced data, where some classes (or behaviours) are observed more often than others (Brown & Mues, 2011). Classification algorithms in general aim to minimize the overall error rate, and often perform badly for rare classes that can be of major ecological interest. The random forest algorithm provides procedures to deal with unbalanced data (Chen, Liaw, & Breiman, 2004). Another advantage is that the performance of the model is estimated intrinsically using the whole dataset, which saves the user from adding a procedure to define a training and a test dataset. The random forest classifier has furthermore been shown to have high prediction accuracy and outperforms other classifiers on numerous datasets (Cutler et al., 2007; Fernández-Delgado, Cernadas, Barro, & Amorim, 2014; Svetnik, Liaw, Tong, & Wang, 2004). This method requires only a few tuning parameters, no assumption on the distribution of predictor variables and it can handle a large number of variables (Breiman, 2003). Hence no pre-processing is needed for variable selection. A disadvantage of using random forests is that they do not provide an easy representation of the model, so interpretation of the relationship between the response and predictor variables may be difficult.

We developed a method to link the movement patterns of individuals with their behavioural state, using random forests. The behavioural state of an individual can refer here to any categorization of the observed behaviour. The specificity of this method relies on the derivation of multiple predictor variables from the movement data over a range of temporal windows, with no need for an initial knowledge of exact behaviour duration (appropriate window size). This procedure maximizes the ability to capture information on the changes of movement. This method is hence very generic and applicable to any dataset providing movement data together with an observation of behaviour. We present three examples of application, on foraging seabirds' activity and presence of conspecifics and on fishing vessels' activity.

The method, including all codes and functions, is available in R software (R Core Team, 2013) in the package "M2B"—Movement to Behaviour. The stable version is on the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=m2b, under a license GNU GPL 3. Its latest version is on the Github repository of the project, at https://github.com/ldbk/m2b.

2 | OVERVIEW OF THE "M2B" PACKAGE

The package aims to convert a complex procedure into a few simple steps. Those include the calculation of predictor variables from movement data, the run of the random forest algorithm to infer behaviour and the visual presentation of results. A new class named 'xytb' contains in a single object all information associated to a track: the tracking data themselves (spatial coordinates, time and behavioural states), the predictor variables, the resulting model (to be used on other datasets) and the predictions (on the given dataset). This class was created to address practical issues regarding the exchange of data, model, prediction and results among users. Functions to export a 'xytb' object to formats used in other R packages dealing with movement analysis were also included. The results from a home range analysis (package "ADEHABITATLT" (Calenge, Dray, & Royer-Carenzi, 2009)) or a hidden markov model (package "MOVEHMM" (Michelot, Langrock, & Patterson, 2016)) can be compared to the ones obtained with "M2B". For more details about the 'xytb' class or all the functions constituting the package, please refer to its vignette and help documents (Dubroca & Thiebault, 2017). This article describes the main steps involved in the methodology.

2.1 | Step 1: Data collection

The data must provide (a) a record of the movement and (b) a record and categorization of the behaviour for at least a part of the dataset. In our examples, the movement data were collected using GPS devices on foraging Cape gannets *Morus capensis* Lichtenstein 1823 and on small polyvalent fishing vessels (Table 1, Supporting Information 1).

Observation data of Cape gannets at sea were collected using bird-borne video cameras. Two types of behavioural categorization were available, from which two models were developed. The first model refers to the activity of the equipped birds: flying, sitting on the water or diving. The second model refers to their interactions with conspecifics: associated with conspecifics or alone. Observation data of fishing vessel activities were recorded by an on board observer. Three main activities were recorded: fishing, cruising and stopped.

The collected data should be explored at this stage for the observation of common movement patterns related to specific behaviours. The application of our methodology on a dataset where such a pattern is week, non-existent or subject to too much variability (in the

TABLE 1 Datasets used for illustration together with details on implementation of the M2B package and of the random forest algorithm. CG = Cape gannets; FV = fishing vessels; DS = down-sampling; Sp. = species; Mov. = movement; Obs. = observations; Cam. = camera; Behav. = behavioural; Consp. = "with conspecifics"; Turn. = turning; P. = position; Redund. = redundancy; Num. = number. Default operators are: Mean, Standard deviation, Median absolute deviation, Quantiles (0%, 25%, 50%, 75%, 100%). Default window length is: 3 to 13 positions, every 2nd position

Data				«M2B» package						Random forest		
Sp.	Mov.	Obs.	Behav. states	Primary variables	Operators	Window length	Backward shift	Redund.	Num. of predictors	ntree	mtry	Balanced procedure
CG	GPS 5 s	Video Cam.	Flying, Sitting, Diving	Speed, Turn. angle	Default	Default	No	Tested	31	500	10	DS
CG	GPS 5 s	Video Cam.	Consp., Alone	Speed, Turn. angle	Default	Default	5-25 p. every 5th p.	Tested	1,282	500	768	DS
FV	GPS 2 min	On Board	Fishing, Cruising, Stopped	Distance, Turn. angle	Default	Default	No	Tested	147	500	30	DS

case of high individual variability for example) will lead to inconsistent predictions.

2.2 | Step 2: Predictor variables

The predictor variables are calculated from the movement data. They all derive from three primary variables: speed, distance and turning angle. The distance is only provided as an option to be used instead of speed in case the time data are not reliable. Speed is calculated between successive positions by dividing the distance measured between consecutive positions by the step duration. The turning angle is measured as the change in direction between successive steps, where a step is a vector connecting two successive positions. These variables characterize the movement of the individual at a given time.

A set of mathematical operators is exhaustively calculated to quantify the variations of speed (or distance) and turning angle over a range of interval times to cover the temporal scales at which the behaviours may occur. Besides the fact that we might not know at which time scales the behaviours occurred, the temporal scale of the recorded data might not match with the behaviour. For example, GPS positions recorded at one-second time intervals do not reflect a feeding or fishing activity that will last for minutes, or a feeding attempt that will last for seconds. Computationally, on a moving window (of length defined by the user, default being 3–13 positions every 2nd position), the mean, standard deviation, median absolute deviation and quantiles (defined by the user, default being 0%, 25%, 50%, 75% and 100%) of the primary variables are computed.

In addition and if relevant to the study, the values of these variables can be shifted backwards in time in order to consider a lag-effect between the reaction of the individual captured in the movement data and the appearance of the cause of movement in the observed data. For example, an individual may react to something detected from a distance (like a possible feeding area), a change immediately captured in the movement data, but the reason for its movement (starting to feed) will only appear later in the observation data. In our example on foraging Cape gannets, the detection of conspecifics often occurs before grouping with them (Thiebault, Mullers, Pistorius, & Tremblay, 2014). For this model, the predictors were shifted 5–250 positions (25 s to 20 min) every 5th position, as most of the distances between reaction and grouping were shown to be under 15 km (corresponding to a time delay of 20 min at an average flying speed of 45 km/hr) for this species (Thiebault et al., 2014).

The set of predictor variables can then be tested to reduce redundancy and non-informative parameters (optional). The predictors can be tested for "near zero-variance" and multicollinearity (Kuhn, 2008). Finally, the primary variables can be forced to be kept in the final set of predictors, for ecological interpretations.

2.3 | Step 3: Random forest

A random forest is based on the classification and regression tree algorithm, on which two levels of randomness were added
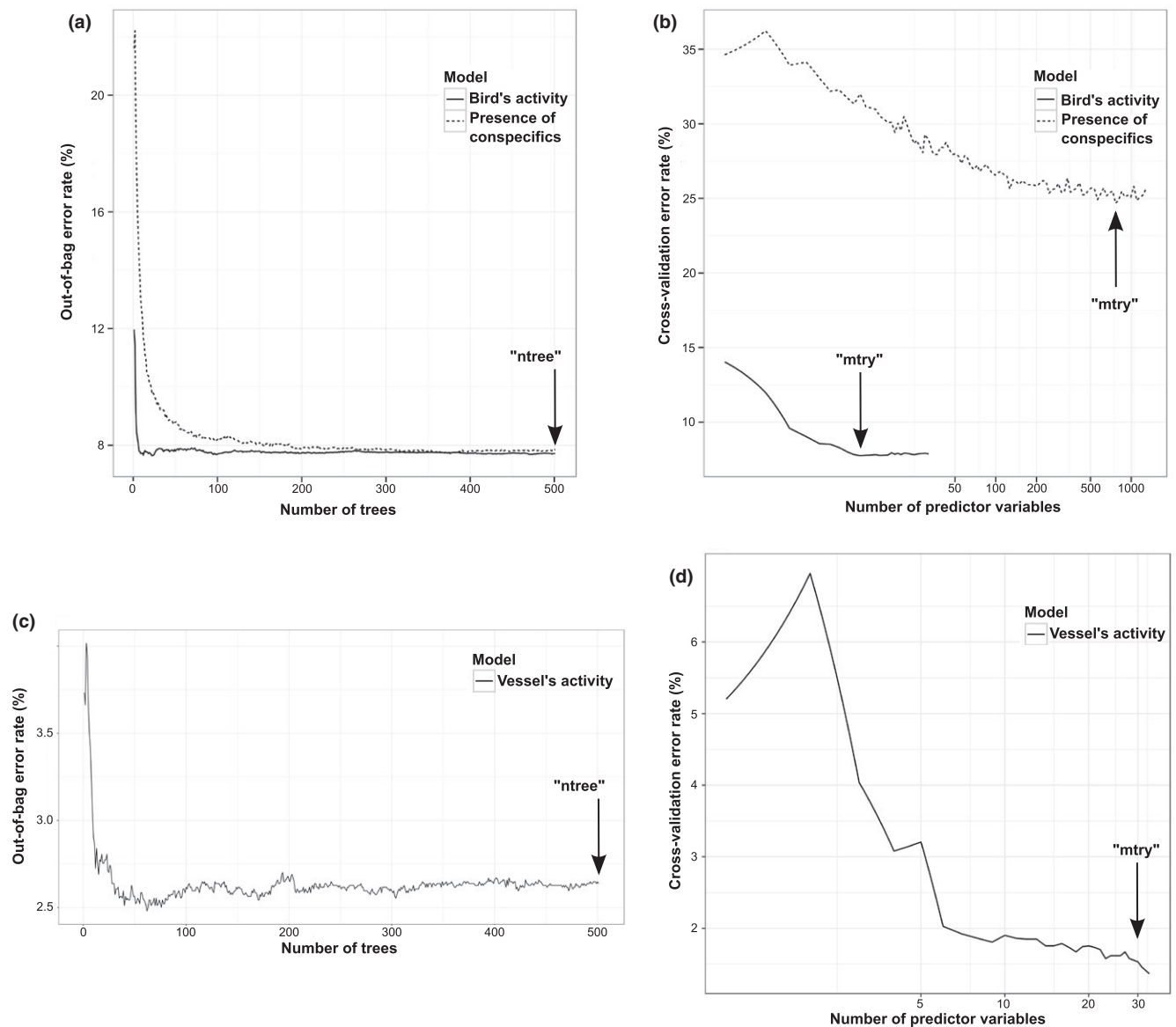
FIGURE 1 Selection of parameters for the random forest model on the Cape gannets data (a and b) and on the fishing vessels data (c and d). (a, c) Out-of-bag error rate (%) as a function of the number of trees grown (parameter "ntree"). (b, d) Cross-validation error rate (%) as a function of the number of predictor variables sampled at each node (parameter "mtry")

(Breiman, 2001, 2003; Supporting Information 2): at each node a subset of the predictor variables is randomly selected to split the data (defined by the parameter "mtry"), and several trees are grown on bootstrap samples of the original dataset (parameter "ntree"). To set the number of trees to be grown, the convergence of the model can be observed as a function of the number of trees used. In our examples, the global error rate stabilized at about 100–200 trees depending on the model (Figure 1a,c). To ensure the stability of the results, we chose 500 trees to be grown. To set the number of variables to be selected at each node from the pool of predictors, a nested cross-validation procedure can be used (Svetnik et al., 2004). In our examples, this parameter was set at 10, 768 and 30 for the different models as they provided the lowest cross-validation error rate (Figure 1b,d).

Observational data were unbalanced, with the diving behaviour of Cape gannets associated with c. 10% of the observations, the presence of conspecifics with c. 14%, and the cruising fishing vessels activity with <10%. To increase the prediction accuracy for the rare classes, we used the balanced random forest algorithm by downsampling the major observational classes according to the occurrence of the minor class (Chen et al., 2004). Sensitivity analyses show that the random forest algorithm can achieve high prediction accuracy in those circumstances (Brown & Mues, 2011; Huang & Boutros, 2016).

The random forest algorithm was implemented in R software in the package "RANDOMFOREST," which provides an interface with the algorithm written by Leo Breiman in Fortran language (Liaw & Wiener, 2002). The "M2B" package calls on this package for the computation of the random forest algorithm.

TABLE 2 Accuracy of prediction for the model predicting the activity of Cape gannets

Confusion matrix	Diving (observed)	Sitting (observed)	Flyling (observed)	Total
Diving (predicted)	275	831	925	2,031
Sitting (predicted)	10	13,316	219	13,545
Flying (predicted)	5	363	14,549	14,917
Total	290	14,510	15,693	
Indicators per class	Diving	Sitting	Flying	
Sensitivity	0.9483	0.9177	0.9271	
Precision	0.1354	0.9831	0.9753	
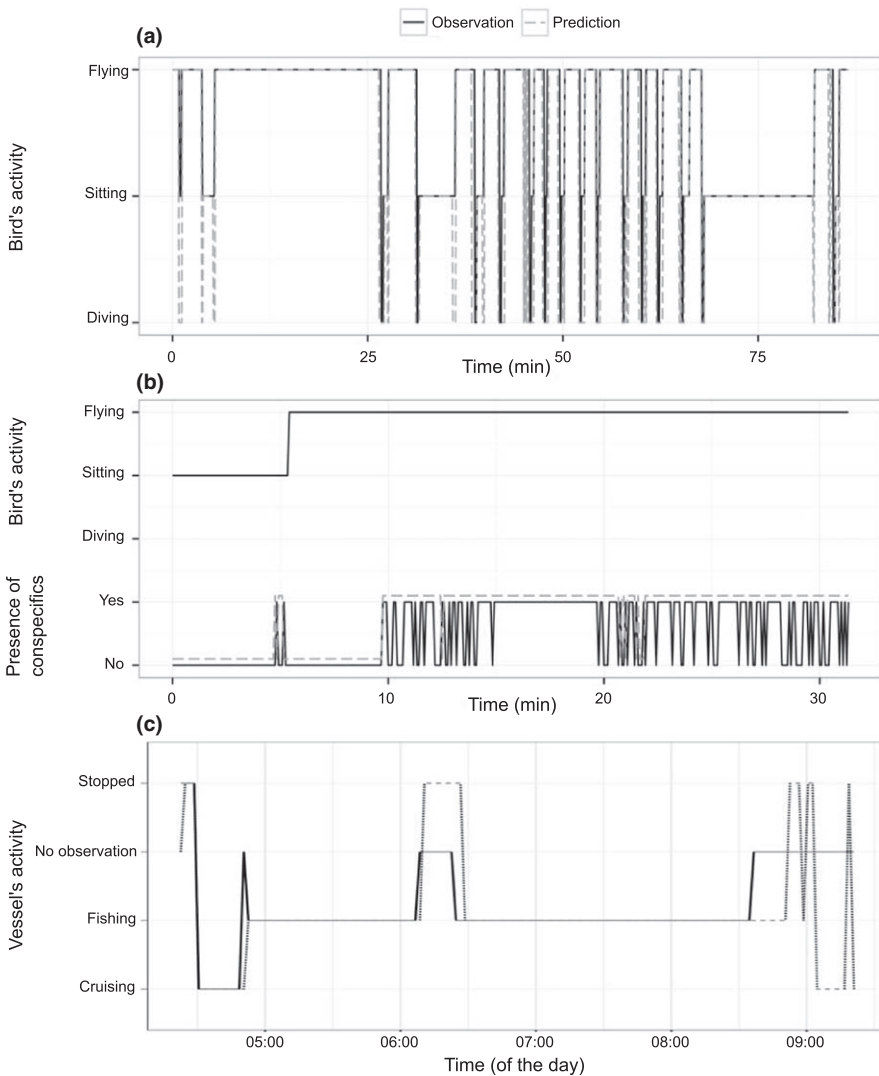
FIGURE 2 (a) Observation and prediction of the activity of a Cape gannet over time. (b) Similar observations from another individual, together with the observation and prediction of the presence of conspecifics in its surroundings ("Yes" for presence and "No" for absence). (c) Observation and prediction of the activity of a fishing vessel over time

2.4 | Step 4: Visualization of results

2.4.1 | Accuracy

The random forest algorithm provides an internal measure of error (Breiman, 2003). This informs on the percentage of mis-classification from which the global accuracy of the model can be calculated. In addition to these global estimations, indicators can evaluate the performance of the model for each class, based on the confusion matrix. Here, we used the indicator "sensitivity" (Altman & Bland, 1994a): the number of correct predictions for a given class divided by the total number of actual occurrences of the class (including all the cases

TABLE 3 Accuracy of prediction for the model predicting the presence of conspecifics in the surroundings of Cape gannets

Confusion matrix	Alone (observed)	Conspecifics (observed)	Total
Alone (predicted)	17,638	390	18,028
Conspecifics (predicted)	1,347	2,713	4,060
Total	18,985	3,103	
Indicators per class	**Alone**	**Conspecifics**	
Sensitivity	0.9290	0.8743	
Precision	0.9784	0.6682	

where this class was incorrectly predicted as another one). We also used the indicator "precision" (Altman & Bland, 1994b): the number of correct predictions for a given class divided by the total number of predictions of the class (including the incorrect prediction of this class).

The global accuracy of prediction was 92.28%, 92.14% and 97.40% for our examples on bird's activity, presence of conspecifics, and fishing vessel's activity respectively. The model predicted very precisely when the birds were sitting on the water or flying and missed few of these behaviours, as shown by the high values of indicators for these classes (Table 2). A very high percentage of the dives were also correctly predicted. The false predictions of diving were either situated around the actual dives, or at the transitions between flying and sitting (Figure 2a). Concerning the presence of conspecifics in the surroundings of Cape gannets, the model performed well at predicting when conspecifics were observed (92.90%) or not observed (87.43%) in the surroundings of the equipped bird (Table 3). Interestingly, the false predictions of presence were situated around the actual observations. The field of view given by the camera is limited, so the observation of conspecifics in the surroundings of an equipped bird can be intermittent (the conspecific is observed in one frame but disappears in the following one because it moved away from the field of view, and comes back a few frames later, etc.). The predictions of the presence of conspecifics seem to smooth those intermittent observations, by predicting the presence for longer continuous periods of

time (Figure 2b). The model predicting the activity of fishing vessels showed exceptionally good accuracy and sensitivity (Table 4). The predictions followed the observations closely and the model predicted activities for time ranges where no observations were recorded (Figure 2c).

Further visualizations of the results are available in the "M2B" package (Dubroca & Thiebault, 2017).

3 | CONCLUSION

Classifiers are powerful tools to infer variables that are difficult to measure directly. Combined with bio-logging techniques (Boyd et al., 2004), they allow us to go beyond the limits of our capacity to observe animals in their natural environment. Our method performed exceptionally well (global accuracy 92%–97%, sensitivity 87%–97%) compared to other studies: e.g. global accuracy of 78% and sensitivity 65–80% using support vector machine on cow's behaviour (Martiskainen et al., 2009), global accuracy of 80% and sensitivity of 87% using hidden semi-markov models on fishing boats' activity (Joo, Bertrand, Tam, & Fablet, 2013) or sensitivity of 83% using an acceleration signal from penguin head movements (Watanabe & Takahashi, 2013).

The goal of our methodology is to provide the user with a classification procedure that is robust (no assumptions on the data, good accuracy), easy to implement (no training or prior knowledge required on how the behaviour influences the movement, all codes available in a R package), easy to compare with other methods for movement analysis in R (using exporting functions provided) and easy to exchange and compare with other R users (using the new created class).

By deriving multiple variables over a number of temporal scales, the changes of movement are quantified using all metrics describing the distribution of their speed and turning angle. The error of prediction reveals how much of the process measured by the response variable (the behaviour) is integrated in the process measured by the predictor variables (the movement).

The package "M2B" provides an interface to easily apply the random forest algorithm to movement data (from the calculation of predictor variables to the visualization of results). The

TABLE 4 Accuracy of prediction for the model predicting the activity of fishing vessels

Confusion matrix	Cruising (observed)	Fishing (observed)	Stopped (observed)	Total
Cruising (predicted)	1,326	19	22	1,367
Fishing (predicted)	10	6,964	9	6,983
Stopped (predicted)	27	160	977	1,164
Total	1,363	7,143	1,008	
Indicators per class	**Cruising**	**Fishing**	**Stopped**	
Sensitivity	0.9749	0.9729	0.9692	
Precision	0.9973	0.9700	0.8393	

methodology can be used on datasets where movement was sampled from different species (including subgroups within a species, from individuals to populations), on different sampling schemes or with different devices recording movement. Observations (direct or remote) are needed for part of the data, but once the model is trained it can be used on any dataset providing similar movement data. This assumes that the link between movement and behaviour is similar in the new dataset compared to the training dataset (caution should be taken in case of high individual variability in which case a model should be run per individual). Besides the predictive model, the percentage of accuracy informs on the significance of the link between predictors and response variables.

AUTHORS' CONTRIBUTIONS

L.D. and A.T. developed the method. L.D. wrote the codes. Y.T. and P.A.P. designed the experiment for the example of application on Cape gannets; A.T. and R.H.E.M. collected the data; A.T. and L.D. analysed the data. A.T. and L.D. wrote the manuscript. All authors reviewed the manuscript.

DATA ACCESSIBILITY

The "м2в" package is available at https://cran.r-project.org/package=m2b. The data on foraging seabirds are available at: https://doi.org/10.5281/zenodo.1168918. The data on fishing vessels' activity are available at: https://doi.org/10.5281/zenodo.1168832.

ORCID

Andréa Thiebault (iD) http://orcid.org/0000-0001-9585-1083

Laurent Dubroca (iD) http://orcid.org/0000-0002-1861-0507

Ralf H.E. Mullers (iD) http://orcid.org/0000-0002-8323-4122

Pierre A. Pistorius (iD) http://orcid.org/0000-0001-6561-7069

REFERENCES

Altman, D. G., & Bland, J. M. (1994a). Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal*, *308*, 1552. https://doi.org/10.1136/bmj.308.6943.1552

Altman, D. G., & Bland, J. M. (1994b). Diagnostic tests 2: Predictive values. *British Medical Journal*, *309*, 102. https://doi.org/10.1136/bmj.309.6947.102

Boyd, I. L., Kato, A., & Ropert-Coudert, Y. (2004). Bio-logging science: Sensing beyond the boundaries. *Memoirs of National Institute of Polar Research*, *58*, 1–14.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2003). *Manual – Setting up, using, and understanding random forests V4.0*. Berkeley: University of California. Retrieved from https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf

Brown, I., & Mues, C. (2011). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, *39*, 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

Calenge, C., Dray, S., & Royer-Carenzi, M. (2009). The concept of animals' trajectories from a data analysis perspective. *Ecological Informatics*, *4*, 34–41. https://doi.org/10.1016/j.ecoinf.2008.10.002

Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data*. Berkeley, CA: University of California.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783–2792. https://doi.org/10.1890/07-0539.1

Dubroca, L., & Thiebault, A. (2017). *Movement to behaviour: 'м2в' r package*. Comprehensive R Archive Network (CRAN). Retrieved from https://cran.r-project.org/web/packages/m2b/vignettes/tutorial.html.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (2nd ed.). New York, NY: Springer. https://doi.org/10.1007/978-0-387-84858-7

Huang, B., & Boutros, P. (2016). The parameter sensitivity of random forests. *BMC Bioinformatics*, *17*, 331–344. https://doi.org/10.1186/s12859-016-1228-x

Joo, R., Bertrand, S., Tam, J., & Fablet, R. (2013). Hidden Markov models: The best models for forager movements? *PLoS ONE*, *8*, e71246. https://doi.org/10.1371/journal.pone.0071246

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M., & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, *119*, 32–38. https://doi.org/10.1016/j.applanim.2009.03.005

Michelot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: An r package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, *7*, 1308–1315. https://doi.org/10.1111/2041-210X.12578

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*, 19052–19059. https://doi.org/10.1073/pnas.0800375105

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rutz, C., & Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, *5*, 289–292. https://doi.org/10.1098/rsbl.2009.0089

Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In F. Roli, J. Kittler, & T. Windeatt (Eds.), *Multiple classifier systems* (pp. 334–343). Lecture Notes in Computer Science. Berlin Heidelberg: Springer. https://doi.org/10.1007/b98227

Thiebault, A., Mullers, R. H. E., Pistorius, P. A., & Tremblay, Y. (2014). Local enhancement in a seabird: Reaction distances and foraging consequence of predator aggregations. *Behavioral Ecology*, *25*, 1302–1310. https://doi.org/10.1093/beheco/aru132

Tremblay, Y., Thiebault, A., Mullers, R., & Pistorius, P. (2014). Bird-borne video-cameras show that seabird movement patterns relate to previously unrevealed proximate environment, not prey. *PLoS ONE, 9,* e88424. https://doi.org/10.1371/journal.pone.0088424

Watanabe, Y. Y., & Takahashi, A. (2013). Linking animal-borne video to accelerometers reveals prey capture variability. *Proceedings of the National Academy of Sciences, 110,* 2199–2204. https://doi.org/10.1073/pnas.1216244110

How to cite this article: Thiebault A, Dubroca L, Mullers RHE, Tremblay Y, Pistorius PA. "м2в" package in ʀ: Deriving multiple variables from movement data to predict behavioural states with random forests. *Methods Ecol Evol.* 2018;9:1548–1555. https://doi.org/10.1111/2041-210X.12989

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.