Supporting Information for:

"m2b" package in R: deriving multiple variables from movement data to predict behavioural states with random forests.

Andréa Thiebault[1*], Laurent Dubroca[2], Ralf Mullers[3], Yann Tremblay[4], Pierre Pistorius[5]

1. Department of Zoology, Nelson Mandela University, South Campus, PO Box 77000, Port Elizabeth 6031, South Africa
2. Datacall Response Unit (CREDO), IFREMER, Avenue du Général de Gaulle, 14520, Port-en-Bessin-Huppain, France
3. Department of Biodiversity, University of Limpopo, Private Bag X1106, Sovenga 0787, South Africa
4. Institut de Recherche pour le Développement, UMR MARBEC 248: Marine Biodiversity, Exploitation and Conservation, Avenue Jean Monnet CS 30171, 34203 Sète cedex, France
5. DST/NRF Centre of Excellence at the Percy FitzPatrick Institute, Department of Zoology, Nelson Mandela University, South Campus, PO Box 77000, Port Elizabeth 6031, South Africa
* Corresponding author: andrea.thiebault@gmail.com

Supporting information 1. Data collection and preprocessing.

Cape gannets

Data collection

Thirty five breeding Cape Gannets were fitted with devices at Bird Island (33° 50' 26.6"S, 26° 17' 14.5"E, Algoa Bay, South Africa) during December 2010 and January 2011. Chick-rearing birds were captured during the chick guard phase when about to depart to sea after a change over with their partner. Only one adult per nest was equipped for one foraging trip, while the partner stayed at the nest with the chick. We deployed two types of devices that were attached together: a GPS logger (i-GotU GT-600, Mobile Action Technology Inc., Taipei, Taiwan, 43 x 40 x 12 mm, 36 g) to record the movement path and a video-camera (Camsports nano, CamsportsTM, Estrablin, France, 68 x 19 mm, 22 g) to observe the behaviour and surroundings of the animal while at sea. The GPS loggers were set to record a geographical position every five seconds when the animal moved faster than 10km.h^{-1} and every 10 or 30 seconds otherwise. The video-cameras recorded 736 x 480 pixels images at 25 frames per second with a 74° lens angle for a maximum of 90 minutes. The handling process lasted less than eight minutes and consisted of weighing the bird using a spring balance (pesolaTM, Baar, Switzerland with a precision of 50g) in addition to fastening the devices using waterproof adhesive tape (TesaTM, Hamburg, Germany). The total mass attached to a bird, including both devices and tape, was 70-75 g that corresponded to 2.3-3.0% of the birds' body mass (2400-3100g). The loggers were attached on the lower back of the bird in such a way that potential drag due to modification of the birds' body shape was minimized. Just before the bird was released, a hand-held GPS was placed in front of the camera eye so that the Greenwich Mean Time (GMT) was recorded in the first few images. The later video observations were hence accurately synchronized to movement data using the satellite derived time. The nests were then monitored every hour (from sunrise to sunset) and the study birds were recaptured and the devices retrieved soon after their return to the colony.

Data preprocessing

The raw tracking data were interpolated using a Bézier curve to obtain a track with regular step durations of 5 s (Tremblay *et al.* 2006). The whole methodology as explained in the main text was applied on both the raw and interpolated tracks, and both provided similar results. As a consequence, we only present the results from the interpolated data.

The video footage were observed frame by frame using a video reader in Matlab software and the events of interest were visually flagged using a purpose-built video event recorder. Video data provided information on the behaviour of equipped birds, including taking off, sea landing, and diving, from which we inferred three activities: flying, sitting on the water and diving. Because GPS and video data were not sampled at similar rates, the video observations were related to GPS locations as followed. Each of the GPS positions was assigned to a behavioural activity based on the observations made from 2s before the position time to 2s after (i.e. including 4s * 25 frames = 100 observations), according to the following rules:
- diving: at least one observation of diving activity during the 4 s interval time (can include some flying and/or sitting on the water)
- sitting on the water: study bird sitting on the water during ≥50% of the observations (can include some flying)
- flying: study bird flying during >50% of the observations (can include some sitting on the water)
The surroundings of the study bird were also directly observed, as was the presence of conspecifics. Each of these observations was assigned to the closest point in time on the tracking data.

Fishing vessels

The data was derived from the deployment of devices on small polyvalent fishing vessels (length < 12m) in the Mediterranean sea (Leblond *et al.* 2010). Movement data were collected using GPS loggers set to record a geographical position every two minutes. Observation data were collected by an onboard independent observer. Three main activities were recorded : fishing, cruising and stopped. Fishing activities include setting and removing the fishing gear from water, while stopped included stay in port or stopping at sea. The change between these three activities were recorded and associated with each position leading to 9516 observations collected on two vessels during July 2012.

References

Leblond, E., Lazure, P., Laurans, M., Rioual, C., Woerther, P., Quemener, L. & Berthou, P. (2010). The Recopesca Project: a new example of participative approach to collect fisheries and in situ environmental data. *Mercator Ocean - Quarterly Newsletter* (Mercator Ocean), **37**, 40-48.

Tremblay, Y., Shaffer, S.A., Fowler, S.L., Kuhn, C.E., McDonald, B.I., Weise, M.J., Bost, C.-A., Weimerskirch, H., Crocker, D.E., Goebel, M.E. & Costa, D.P. (2006). Interpolation of animal tracking data in a fluid environment. *The Journal of Experimental Biology*, **209**, 128–140.