



# Distribution and habitat use modelling from satellite tracking data of humpback whales in Brazil agrees with shipboard survey data modelling

Guilherme A. Bortolotto<sup>1,2,3,\*</sup>, Alexandre N. Zerbini<sup>3,4,5,6</sup>, Len Thomas<sup>2</sup>,  
Artur Andriolo<sup>3,6</sup>, Philip S. Hammond<sup>1,2</sup>

<sup>1</sup>Sea Mammal Research Unit & Scottish Oceans Institute, University of St Andrews, St Andrews KY16 8LB, UK

<sup>2</sup>Centre for Research into Ecological and Environment Modelling, University of St Andrews, St Andrews KY16 9LZ, UK

<sup>3</sup>Instituto Aqualie, Juiz de Fora 36036-330, MG, Brazil

<sup>4</sup>Cooperative Institute for Climate, Ocean and Ecosystem Studies, University of Washington & Marine Mammal Laboratory, Alaska Fisheries Science Center, Seattle, WA 98115, USA

<sup>5</sup>Marine Ecology and Telemetry Research, Seabeck, WA 98380, USA

<sup>6</sup>Laboratório de Ecologia Comportamental e Bioacústica, Programa de Pós-graduação em Biodiversidade e Conservação da Natureza, Universidade Federal de Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil

**ABSTRACT:** Statistical modelling of animal distributions has been widely applied to explain how mobile species use their habitats. The distribution of and habitat use by humpback whales *Megaptera novaeangliae* off the eastern coast of Brazil have previously been investigated by modelling visual survey data. Here, we modelled distribution in their breeding range using individual tracking data to compare ecological inferences with those from previous models from line transect data. A generalised estimating equation framework was used to model the tracking data and pseudo-absences as functions of spatial covariates. Covariates considered were latitude and longitude, sea surface temperature (SST), current and wind speeds near the surface, distances to shelf-break and the coast, sea bottom depth and slope, and a factor variable representing 'shelter'. Two modelling exercises were developed: a habitat use model (HUM) and a distribution model (DIM). Covariates retained in the selected HUM were SST, distance to coast and shelf-break, current and wind speeds and shelter. Covariates retained in the selected DIM were latitude/longitude, current speed and distances to shelf-break and coast. The modelled relationships between whale occurrence and environmental covariates using tracking data were similar to those using line transect data. Distribution maps were also similar, supporting higher densities around the Abrolhos Archipelago and to its south. We showed that habitat use and distribution of this population in the area could be similarly inferred by modelling either line transect or tracking data. Using these 2 approaches in conjunction can strengthen the understanding of important ecological aspects of animal populations.

**KEY WORDS:** *Megaptera novaeangliae* · Ecology · Conservation · Marine mammals · Population recovery

Resale or republication not permitted without written consent of the publisher

## 1. INTRODUCTION

Animal distributions and habitat use result from animals exploiting resources to meet their ecological

requirements, from their physiological constraints and the effects of the environment on them (Matthiopoulos & Aarts 2010). Data on where animals are distributed and how they use the environ-

ment can inform the management of human activities to avoid conflicts. In that context, distribution models are useful tools to support the identification of areas that require management (Runge et al. 2015). Because different types of data for studying species' spatial ecology may be available, it is important to explore whether similar inferences about distribution and habitat use are obtained. Animal distribution data obtained through different methods within an area of interest could, for example, complement each other to improve the spatial extent of information available.

Population distribution and habitat use can be explored using a range of data types and analytical methods, with outputs partially depending on the nature of the data used (Redfern et al. 2006, Aarts et al. 2008). Line transect sampling, for example, is designed to estimate density and abundance, and the data can be used to infer distribution and habitat use through spatial modelling methods (Miller et al. 2013, Roberts et al. 2016). In line transect surveys, animals are detected along transects, and the effective search areas of those transects are estimated from detection distances using distance sampling (Buckland et al. 2001). Line transect data are therefore counts of animals detected in units of sampled space.

Animal tracking data, commonly acquired using animal-borne tags, can also be used to investigate distributions and habitat use. For marine mammals, developments to improve equipment and the increasing availability of analysis tools have facilitated their application (Aarts et al. 2008, Jonsen 2016, Russell et al. 2016, Trudelle et al. 2016, Elith et al. 2020, Carter et al. 2022). In contrast to line transect surveys, for which the data are both locations where animals were detected and, crucially, where they were not, the sampling units in tracking studies are individual animals, and the data are formed by a series of observed locations of the tracked individuals. This means that the portion of space that is sampled is only that which is visited by those animals, meaning that there is no information in the data about places that are not visited. Another important difference between the 2 data types is that line transect data are a snapshot of the location of animals at the time of the survey (i.e. cross-sectional data), while tracking data consist of a set of repeated measurements of the location of a sample of animals that have been telemetered (i.e. longitudinal data, on multiple individuals). Despite the above differences, the analysis of contrasting data types can provide different views of the underlying distribution of a species (Matthiopoulos et al. 2022).

For telemetry tracking data to be useful for habitat use and distribution modelling, the design of analysis and interpretation of results must consider many potential biases (Elith et al. 2020). For example, the distribution of locations can be greatly influenced by the distribution of tag deployments, resulting in tagging location bias (Block et al. 2011). Also important is the serial autocorrelation in locations along tracks because observations are naturally space–time series (Matthiopoulos & Aarts 2010). It is crucial to consider whether tracked individuals represent a small portion of the population for which inferences are intended, meaning that although the sample size of locations may appear large (i.e. many locations), sample sizes for individuals can be small (i.e. few animals). This also raises concerns regarding whether tracked animals realistically represent their population with respect to distribution and habitat use (Sequeira et al. 2019).

Animal distribution and habitat use can be inferred using a presence–absence approach. However, such an approach requires information on both where animals were present (i.e. places surveyed where animals were detected) and where animals were absent (i.e. places surveyed with no animals detected). Because of the lack of information on real absences in tracking data (e.g. Pirodda et al. 2011), 'pseudo-absence' locations, which represent the available habitat that was not visited, are needed for presence–absence distribution modelling of tracking data (Elith et al. 2020). When using pseudo-absences, the number of absences to include in the models is controlled by the user, so the ratio of pseudo-absence per presence to be used must be decided. That decision should be guided by model 'coefficient stability analysis' (when alternative models with an increasing number of pseudo-absences per presence are run to visually investigate the stability of the output coefficients; see Supporting Information S4 in Ventura et al. 2019), while ensuring that pseudo-absences have the corresponding fraction of the weight of a presence in the models (i.e. the weighted sum of presences equals the weighted sum of pseudo-absences; Barbet-Massin et al. 2012).

A decision about how to define the area of inference is also required. Some studies have used a percentage of kernel density estimates to represent habitats frequently used by the animals (e.g. Pendoley et al. 2014, Mei et al. 2017, Thorne et al. 2017); that percentage is, again, user-controlled. Adopting a pre-defined area, where the distribution of and habitat use by a population have already been studied using different data, can contribute to a better

understanding of how those animals relate to their environment.

Most humpback whale *Megaptera novaeangliae* populations are recovering after the end of global commercial whaling (Zerbini et al. 2019), and the population breeding in coastal waters of Brazil during winter and spring is no exception (Bortolotto et al. 2016, Wedekin et al. 2017). However, the coincident increase in human activities in Brazilian coastal waters, especially those related to oil and gas production (Bortolotto et al. 2017), means that there is a need to understand how these activities may affect distribution and habitat use to inform whether management actions may be necessary to avoid impact on the population. Distribution of this population in the area was investigated in the past using descriptive analysis (Martins et al. 2001, 2013, Zerbini et al. 2006, Castro et al. 2014, Gonçalves et al. 2018), from a difference in densities within line transect survey blocks (Andriolo et al. 2010) or for very restricted areas (Martins et al. 2001, Gonçalves et al. 2018). Analysis of line transect data to investigate the distribution of humpback whales off the coast of Brazil has shown that density was strongly related to sea temperature and bathymetric features (Bortolotto et al. 2017, Pavanato et al. 2018). Telemetry data for the species in the area has been used to investigate the relative usage of protected areas (Castro et al. 2014) and movements (Zerbini et al. 2006), but not distribution.

Here, we present distribution and habitat use models based on telemetry tracking data of humpback whales in Brazil and compare model outputs with those from a study using line-transect survey data (i.e. Bortolotto et al. 2017). Comparing the results on distribution and habitat from different approaches can strengthen the understanding of important ecological aspects of animal populations, especially when the results support each other. Therefore, management actions, such as spatial planning of natural resources exploration (e.g. oil exploration, fishing), and the need for conservation strategies for populations may be better informed.

## 2. MATERIALS AND METHODS

The area of inference for modelling humpback whale distribution and habitat use in this study was determined by adapting the survey area from Bortolotto et al. (2017), originally comprising the continental shelf from 5 to 23° S, to allow a direct comparison. The original area was restricted considering both the extent of spatial covariates available and the distribution of

tagging locations. Since in a presence/pseudo-absence approach for spatial modelling the area of inference is assumed to be the habitat available to the animals (Aarts et al. 2008), and because the tagging locations were not distributed evenly within the original study area (Fig. 1), the northernmost portion of the original area, originally extending north to Natal (Fig. 4 in Bortolotto et al. 2017), was excluded to minimise the chances of failing to meet that assumption.

### 2.1. Data acquisition: whale tracking

From 2003 to 2012, satellite-linked telemetry tags were attached to adult humpback whales along the

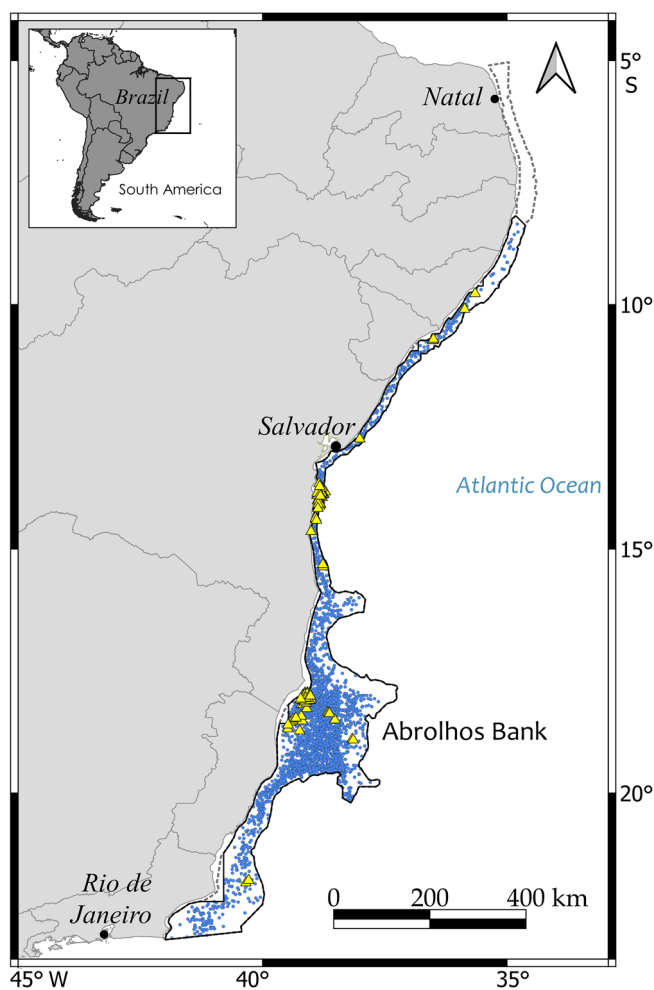


Fig. 1. Humpback whale locations used in the distribution and habitat use analysis (blue dots) and tagging locations (yellow triangles; see Fig. S1 in the Supplement, [www.int-res.com/articles/suppl/m720p161\\_supp.pdf](http://www.int-res.com/articles/suppl/m720p161_supp.pdf), for further details). The area of inference (black solid line) was adapted from Bortolotto et al. (2017) (grey dashed line), considering the uneven tagging locations and extent of environmental covariates

coast of Brazil. Tagging operations occurred every year between August and December except for 2004, when no tagging took place (see Table S1 in the Supplement at [www.int-res.com/articles/suppl/m720p161\\_supp.pdf](http://www.int-res.com/articles/suppl/m720p161_supp.pdf)). Animals were tagged only in good weather conditions, with calm seas and light to moderate winds (i.e. sea state less than 4 on the Beaufort scale).

Implantable ( $n = 108$ ) and Low Impact Minimum Percutaneous Electronic Transmitter (LIMPET;  $n = 5$ ) tags from Wildlife Computers were used. Implantable tags were attached to the animals using a carbon fibre pole or with an Air Rocket Transmitter System (ARTS; Heide-Jørgensen et al. 2001); LIMPET tags were attached using a crossbow (Andrews et al. 2008). The targeted tag location on the animal was the base of the dorsal fin because this area is most frequently exposed when animals surface to breathe, maximising the chances of transmitter–satellite signal linking. The tagging boat approached target animals to distances between 3 and 10 m. When the pole was used, a maximum distance of 5 m was required. Photographs of the tagged animal’s tail fluke, dorsal fin and attached tag were taken for individual identification and quality control of tagging methods. Tracking data from tagged whales were obtained via the Argos satellite system ([www.argos-system.org](http://www.argos-system.org), Collecte Localisation Satellites), and tags were programmed to transmit in various duty cycles (e.g. transmitting every second day or every 4 d; Table S1), to maximise the longevity of animal tracking. Sexes were genetically identified from skin biopsies collected with a crossbow and a modified dart (Dalla Rosa et al. 2008).

## 2.2. Tracking data processing

Because the present whale tracks were obtained via the Argos satellite system, which uses Doppler-based positioning (Lopez et al. 2015), each location was estimated with an associated uncertainty. To minimise the number of unrealistic locations, a speed filter was applied to the whale tracking data, implemented with the R package ‘trip’ (version 1.8.7; Sumner et al. 2009) and assuming a maximum swimming speed of  $12 \text{ km h}^{-1}$  for humpback whales (Garrigue et al. 2010). To reduce the occurrence of long gaps between locations, for which no informa-

tion was available, tracks were split into 2 or more if time gaps were longer than 10 d (Table S2). To account for irregular time intervals and uncertainty in estimated locations, humpback whale tracks were re-estimated using a hierarchical form of a first-difference-correlated-random walk model (DCRW; Jonsen 2016). That model was fitted with the R package ‘bsam’ (version 1.1.2; Jonsen 2016) and used to predict 2 locations per day (i.e. time step of 12 h) for each track. Default ‘bsam’ package model diagnostics and plots of predicted locations versus original locations were inspected to check for both model convergence and whether derived locations were aligned with observed locations. Data from individuals for which the model did not converge were discarded from the analysis. A summary of the data used in the analysis is presented in Table 1 (see Table S1 for further details). The uncertainty associated with the estimated locations was not considered in the distribution and habitat use models.

Because the objectives here were related to investigating distribution in the breeding area and to compare results to a previous study, derived locations beyond the limits of the survey area, e.g. in offshore waters or during migration, were censored from the dataset (Fig. 1) before the distribution models were built.

## 2.3. Covariates

Environmental covariate values used in presence/pseudo-absence spatial models were extracted from published datasets for the post-processed track locations. Candidate covariates considered in the models

Table 1. Summary of the data set (presences only) used for modelling the distribution of humpback whales off the coast of Brazil. Max locations = maximum number of locations per track within a year

Year	Tracks	Locations	Initial date	Ending date	Max locations
2003	8	340	18 Oct	25 Dec	118
2005	11	359	11 Oct	27 Nov	83
2006	3	75	12 Oct	10 Nov	37
2007	7	281	08 Sep	18 Oct	78
2008	17	598	27 Aug	03 Nov	97
2009	9	399	14 Sep	09 Nov	98
2010	10	282	18 Sep	06 Nov	56
2011	8	314	29 Sep	20 Nov	63
2012	13	407	08 Aug	17 Dec	85
2013 <sup>a</sup>	1	16	20 Aug	27 Aug	16
Total/overall	87	3071	—	—	—

<sup>a</sup>Locations from a whale tagged in 2012, which returned to the breeding ground in the following year

were chosen to permit comparison to a previous modelling study using line transect data from humpback whales in the same region (Bortolotto et al. 2017). Current speed close to the surface (Curr.sp) values were extracted from the OSCAR Third Degree Sea Surface Velocity dataset (ESR 2009), with a resolution of  $0.33 \times 0.33^\circ$  (latitude  $\times$  longitude), in 5 d intervals. Daily values for sea surface temperature (SST), with a resolution of  $0.01 \times 0.01^\circ$ , were extracted from the JPL MUR SST project dataset (JPL MUR MEaSURES Project 2010). Daily wind speed at the surface of the sea (Wind.sp) values were extracted from the Era-Interim dataset (Dee et al. 2011), with a horizontal resolution of  $0.125 \times 0.125^\circ$ . Depth values were extracted from ETOPO1 (Amante & Eakins 2009), for which the resolution is  $0.1 \times 0.1^\circ$ . Slope was also derived from ETOPO1. Distances from coast (Dist.coast) and from the shelf-break (Dist.shelf), represented here by the 500 m depth contour, were measured with the 'gDistance' function in the 'rgeos' R package (version 0.3-26; Bivand et al. 2013). To represent regions within the area where animals could be sheltered from rougher weather and colder waters, a factor covariate for 'shelter' was created by combining values of SST and wind speed in 6 classes, defined by quantiles of these covariates, sensu Bortolotto et al. (2017). The logarithm of depth was used in the models.

Values for the dynamic covariates (i.e. Curr.sp, SST and Wind.sp) were obtained by matching the timestamps of each location with the nearest (in time and horizontal space) covariate value.

#### 2.4. Data analysis and modelling

Processed track locations were considered as presences in spatial models. To provide contrasting data, pseudo-absences were generated randomly within the survey area to represent where animals could have been, assuming that the survey area was accessible to all tagged animals. To select the most appropriate ratio of presence to pseudo-absences, different scenarios (i.e. 1:1, 1:3, 1:5 and 1:10) were explored to verify model 'coefficient stability' (see Supporting Information S4 in Ventura et al. 2019). In every model scenario, weighting was applied so that the weight of pseudo-absences followed the corresponding fraction, to ensure that the number of presences was equal to the sum of the weights of the pseudo-absences. The model fitted to the data with 5 pseudo-absences per presence produced the same fitted coefficients as with 10, but different from smaller

ratios (1:1 and 1:3), indicating that 5 pseudo-absences per presence was sufficient to represent the underlying background of the inference area. To match a set of 5 pseudo-absences to 1 specific presence, all locations within a set of 5 pseudo-absences were given the same timestamp as the corresponding presence.

Covariate pairs that were strongly correlated ( $>0.7$ ), or that had high ( $>10$ ; Hair et al. 2014) variance inflation factor scores ('vif' function, 'car' R package, version 1.2-7; Fox & Weisberg 2019) when together in a model, were not included simultaneously in the same model.

Two objectives guided the inclusion of candidate covariates in the spatial models. For the distribution model (DIM), the objective was to create the best distribution map possible, so that all available covariates were considered. For the habitat use model (HUM), the objective was to investigate the relationship of whale occurrence and important ecological characteristics, and all covariates but geographic position (latitude and longitude) were considered.

To compare the present models with those of Bortolotto et al. (2017), latitude and longitude were not considered in the HUM. This procedure was adopted because these covariates have no logical biological interpretation for habitat use, and because they were both strongly correlated with SST. The DIM included those covariates as potential explanatory terms because its objective was to identify those features that describe most variability in whale occurrence and to generate the best predictive map for potentially informing management. The present DIM map was created to be compared to the abundance estimation model (AEM) map of Bortolotto et al. (2017).

Presence vs. pseudo-absence was modelled as a binomial random variable, with probability of presence a smooth function of the continuous covariates, first within a generalised additive model (GAM; Wood 2017) framework and then fitting the GAM using a generalised estimating equation (GEE; Hardin & Hilbe 2013) approach to account for possible autocorrelation, as detailed below. Continuous covariates were included in the models using b-splines, and a logit link function was used. Weighted regression was used, where presences were assigned a weight of 1 and pseudo-absences a weight of 0.2. Model fitting was done using software R (R Core Team 2017), with the functions specified below.

For both the DIM and HUM, covariate selection proceeded in 3 stages. First, the full model (i.e. model with all candidate covariates) was fitted, with the number and location of knots in each b-spline selected using the Spatially Adaptive Local Smoothing Algo-

rithm (SALSA; Walker et al. 2011), implemented with the ‘MRSea’ R package (version 1.0.beta; Scott-Hayward et al. 2017). The maximum number of knots was restricted to 8 for 1-dimensional smoothers and to 20 for 2-dimensional smoothers (i.e. for geographic position), to prevent overfitting of smooth terms (Wood 2017). Continuous covariates for which no knot was indicated as significant, at  $\alpha = 0.05$ , were not considered in subsequent steps.

In the second stage, a backwards covariate selection process was followed. The model from the first step was refitted using a GEE to accommodate residual autocorrelation. For this, the data were organised in correlation panels, with a panel for each set of presences within a track (1 panel per track) and a different panel for each pseudo-absence (1 panel per pseudo-absence). Using this panel structure was congruent with the assumption that model residuals within a track were correlated, but that residuals in different tracks were not, and that residuals for pseudo-absences were mutually independent. Model fitting used the ‘geeglm’ function within the ‘geepack’ R package (version 1.2-1; Højsgaard et al. 2016). The quasi-likelihood under the independence model criterion (QICu) score (Pan 2001) of the fitted model was calculated. A series of models were then fitted, leaving one covariate out at a time, and QICu scores were calculated for each of these models. The model with the lowest QICu score was retained. If this model had fewer covariates than the full model, then the process was repeated to drop another covariate. This was repeated until dropping covariates did not result in any further decrease in QICu.

In the third stage, the statistical significance of the remaining covariates was assessed using marginal p-values, via the function ‘getPvalues’ in the ‘MRSea’ R package (Scott-Hayward et al. 2017). Any covariates that were non-significant using an  $\alpha$ -level of 0.05 were dropped and the GEE re-fitted.

For the selected DIM and HUM, model performance was verified with receiver operating characteristic (ROC) curves and confusion matrices, as per Pirotta et al. (2011), using the R package ‘ROCR’ (version 1.0-7; Sing et al. 2009). The ROC and confusion matrix can be used to calculate percentages of false positives and false negatives expected for the model, by comparing the predicted values to the observed values.

For comparison to the distribution maps presented in Bortolotto et al. (2017), prediction grids ( $8 \times 8$  km cells) containing covariate values from 2008 and 2012 (the years for which distribution models from sighting data were developed in that study) were

used for predictions using the selected DIM. The contribution of each covariate in the final HUM was visualised with partial plots, with confidence intervals based on the GEE estimated uncertainty, adapting custom R code from Pirotta et al. (2011).

### 3. RESULTS

Data from 113 tags, deployed from 2003 to 2012, were available for the analysis. Model outputs from the R package ‘bsam’ (e.g. posterior densities and model convergence plots) indicated poor model fit and/or convergence for whale tracks with fewer than 12 locations or fewer than 5 days of tracking. Those tracks were therefore excluded from the analysis. After filtering, interpolated locations from 87 tracks were available for investigating distribution and habitat use (Fig. 1), comprising 62 tracks from females (mothers with calves), 19 from males and 6 from animals of unidentified sex (Table 1). Because 5 pseudo-absences were created for each presence (i.e. interpolated locations within the survey area;  $n = 3071$ ), 15 355 pseudo-absences were used, totaling 18 426 locations to be modelled.

The final HUM included smooth terms for current speed, SST, wind speed, distance to the coast and to the shelf-break and the factor variable shelter (Table 2). As a result of using GEEs for dealing with autocorrelation in the data, confidence intervals for fitted relationships between the response variable and the covariates were very wide (i.e. high uncertainty), except for SST (Fig. 2). Depth had no statistically significant knots

Table 2. Covariates retained in models for both present tracking data and for line transect data. HUM: habitat use model; DIM: distribution model; AEM: abundance estimation model from Bortolotto et al. (2017); SST: sea surface temperature. Asterisks (\*) indicate covariates retained in the final model; blank cells indicate covariates not retained; and long dashes (—) indicate covariates not considered as candidates

Covariate	Tracking data (present)		Line transect data	
	HUM	DIM	HUM	AEM
Current speed ( $m s^{-1}$ )	*	*	*	*
Depth (m)				
Distance to shelf-break (m)	*	*	*	*
Distance to coast (m)	*	*	*	*
Shelter	*		*	*
Slope (°)				
SST (°C)	*		*	
Wind speed (kt)	*			*
Latitude and longitude (°)	—	*	—	*

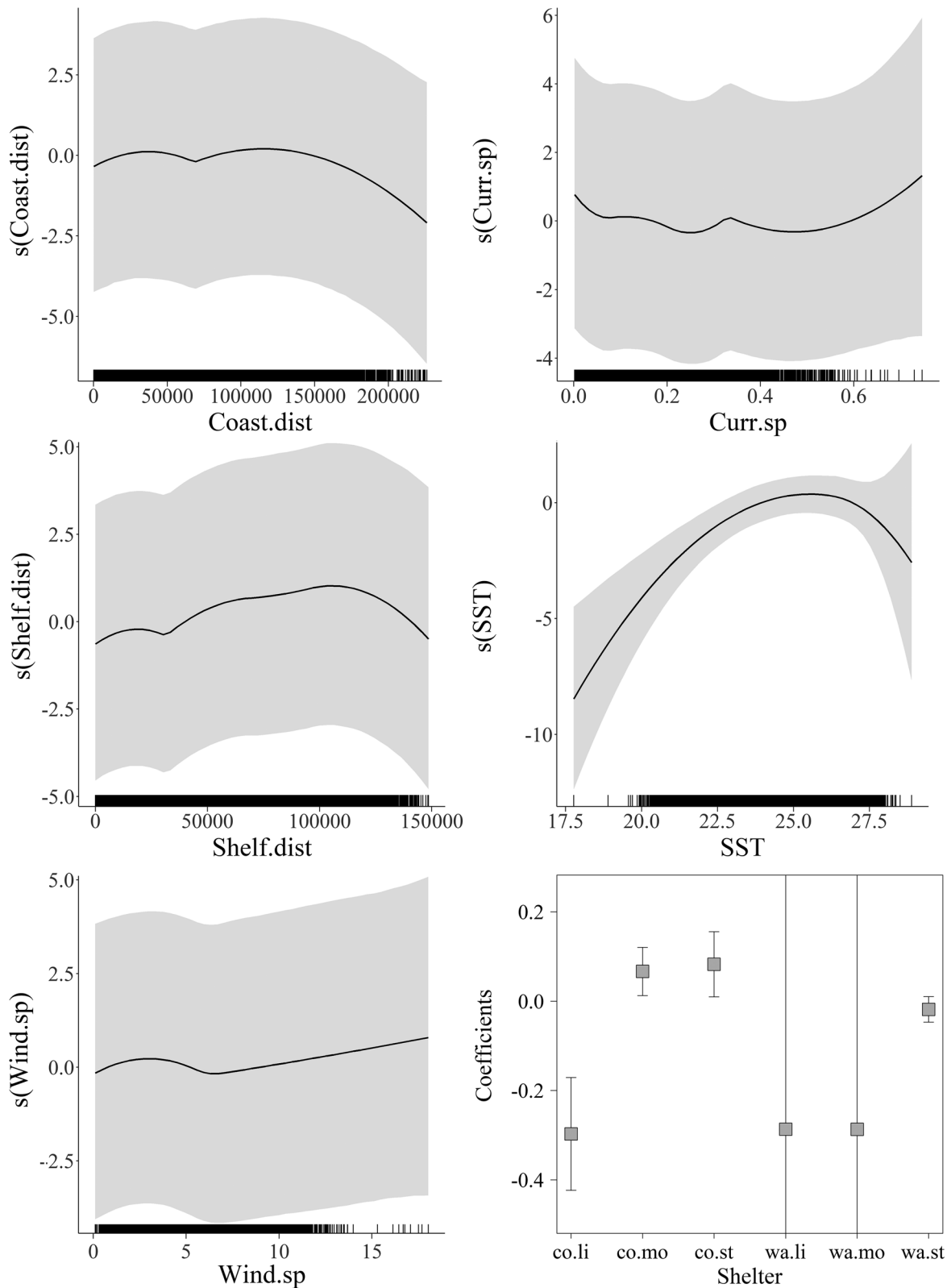


Fig. 2. Fitted relationships (black solid lines) and 95% confidence intervals (grey areas around fitted lines) for smooth functions of covariates in the final habitat use model (HUM). Error bars in the ‘Shelter’ plot represent 95% normal confidence intervals. co: cold; wa: warm; li: light; mo: moderate; st: strong; Curr.sp: current speed; SST: sea surface temperature; Wind.sp: wind speed; Coast.dist: distance to the coast; Shelf.dist: distance to the shelf-break

when the full model was fitted with SALSA and was not considered further. The fitted relationship for the SST covariate showed a clear peak around 24–25°C (Fig. 2), similar to Bortolotto et al. (2017).

The final DIM included latitude and longitude, current speed and distances to coast and shelf-break (Table S4). Shelter and wind had no significant knots at the first step of covariate selection, and depth had a non-significant marginal p-value in the last step of covariate selection for the DIM (Table 2). Overall, higher probabilities of encountering whales were predicted for the region around the Abrolhos Archipelago and to the south, supporting findings from Bortolotto et al. (2017) (Fig. 3). For 2008, 74.7% of the tracking data model predictions that were above the median (median = 0.44) spatially overlapped with line transect model predictions above the median (median = 4.01). For 2012, 73.7% of the tracking data model predictions that were above the median (median = 0.41) overlapped with line transect model predictions above the median (median = 5.11). However, clear differences were observed at a smaller scale: in the Abrolhos Bank region, a patch of predicted lower probability of occurrence for both 2008 and 2012 overlapped with the area presenting the highest densities from Bortolotto et al. (2017).

Models performed better than random, with confusion matrices indicating 66.2% of correct predictions for the HUM, and 63.1% for the DIM, and with areas under the ROC curve (AUC) of 0.691 and 0.732, respectively (Table 3).

#### 4. DISCUSSION

We investigated humpback whale distribution and habitat use off Brazil with spatial models applied to locations obtained through animal tracking and to spatial covariates. To generalise our results to other whale populations, they should be interpreted in light of several important specificities of the population studied here. For instance, the increasing abundance of this recovering population (Zerbini et al. 2019, Bortolotto et al. 2021) may have an effect on their distribution and on how they use their habitats in the area, which was not possible to evaluate here because of the nature of the individual tracking data (i.e. data from multiple years pooled together). Also, because here we considered whales on a breeding ground, the imbalance in the reproductive stage (or animal sex) in our tracked animals (see Section 3; Table S1) very likely prevents them from accurately representing the population in general, because ani-

mals in different reproductive stages are known to use habitat differently (Cartwright et al. 2012, Derville et al. 2018). The scale of the dynamic environmental predictors (e.g. current speed, SST) and the effect of environmental variability in other systems, such as their feeding grounds, must also be considered. However, here we focus on describing mainly the differences in the implementation and interpretation of modelling of either data from individual tracking or from line transect sampling.

##### 4.1. Model considerations and constraints

Spatiotemporal correlation in track locations and other common issues in applying presence/absence modelling for distribution and habitat use (Aarts et al. 2008) were carefully considered in the analysis. Despite differences in the sampling unit of the tracking and survey data (individual vs. space) and statistical tools used, some of the outputs from modelling track locations in a presence/pseudo-absence modelling approach support findings from distribution models fitted to line transect data (Bortolotto et al. 2017). Although the uncertainty in most of the covariate-response (telemetry data) relations fitted here was much higher (Fig. 2), SST showed a clear peak around the same range of temperatures in both the present study and in the line transect modelling (Fig. S3), which probably reflects how strongly habitat use for these animals is related to temperature in their breeding grounds (see Bortolotto et al. 2017 for a discussion on the role of SST in habitat use by humpback whales). The apparent agreement between the 2 methods is not only reassuring for what is known about habitat use for this population, but also shows that similar, possibly complementary, conclusions can be drawn regardless of which of the 2 methods is used, given a sufficient sample size and careful consideration of potential sources of bias. However, differences, such as the patch of present lower probability of whale occurrence that overlapped with the area presenting the highest densities from Bortolotto et al. (2017), need to be considered at smaller scales.

It is important to note that by creating pseudo-absences from random locations in the entire area, it was assumed that the entire survey area was available to the animals at any time (Soberón & Peterson 2005). This is an arbitrary decision and the available area from the animals' perspective could be different (Aarts et al. 2008, Hazen et al. 2021). For example, an animal tagged in the vicinity of the Abrolhos Archipelago (Fig. 1) would take at least 2 d to reach the



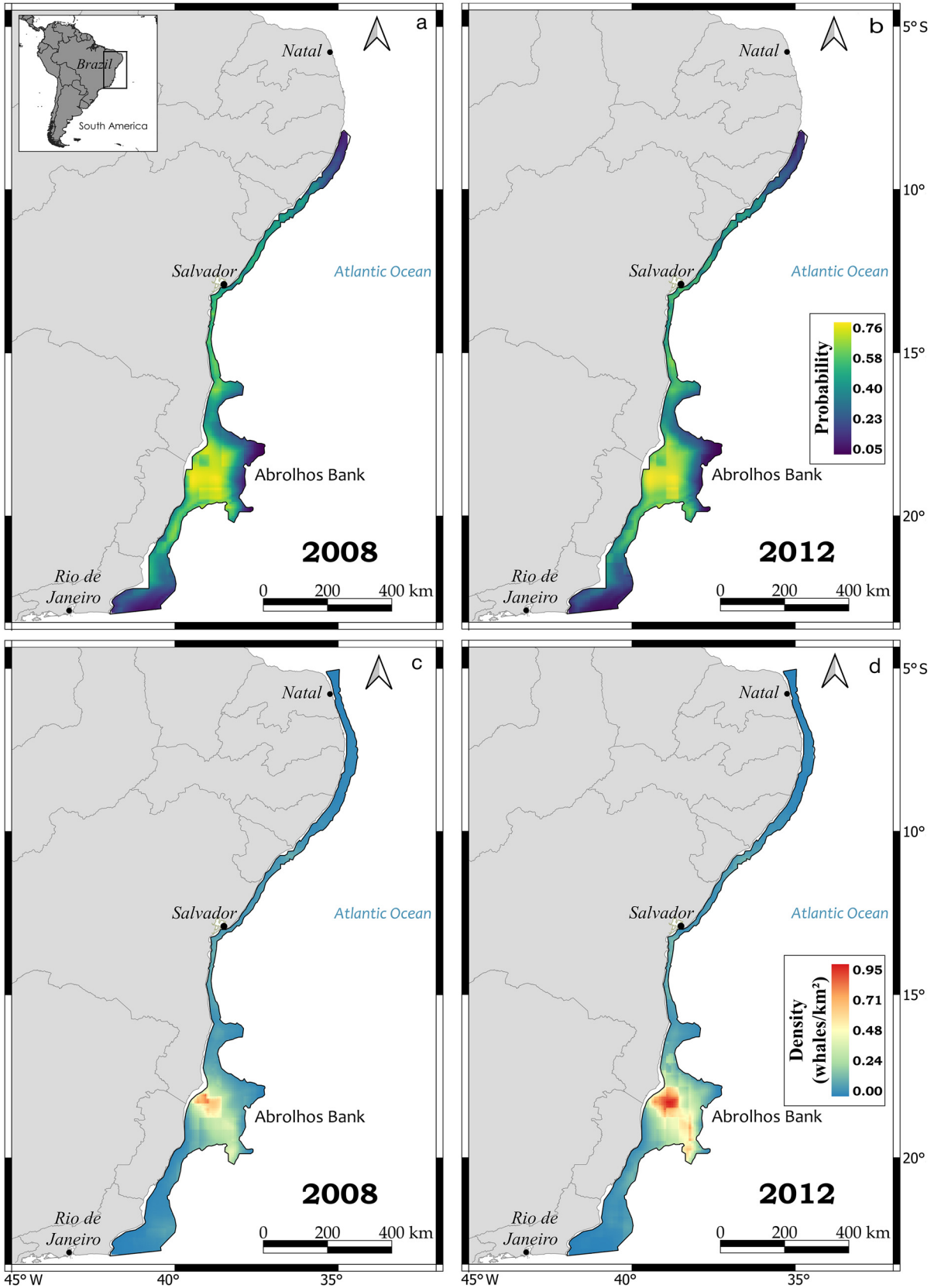


Fig. 3. (a,b) Occurrence probability surfaces, predicted using the present distribution model (DIM) of tracking data. (c,d) Abundance estimation model (AEM) maps adapted from Bortolotto et al. (2017), showing humpback whale density surfaces, for comparison. See Fig. S2 for uncertainty maps of present DIMs

Table 3. Performance of models fitted to telemetry tracking data of humpback whales. HUM: habitat use model; DIM: distribution model; AUC: area under the receiver operating curve

Model	HUM		DIM		
Correctly predicted	66.2%		63.1%		
AUC	0.691		0.732		
Confusion matrices	Predicted	Observed		Observed	
		1	0	1	0
	1	1871	5022	2330	6063
	0	1200	10333	741	9292
Total observed		3071	15355	3071	15355
Percent of observed	1	60.9%	32.7%	75.9%	39.5%
	0	39.1%	67.3%	24.1%	60.5%

southernmost portion of the survey area, swimming in a straight line at 12 km h<sup>-1</sup> (i.e. the maximum travel speed [Garrigue et al. 2010] used as a criterion in the data pre-processing/filtering).

The restriction of the area of inference (survey area) in comparison to that of Bortolotto et al. (2017) (compare north extents of the upper and the lower panels of Fig. 3) was adopted to reduce the effects of violating the area availability assumption (Hazen et al. 2021). Further restrictions could be investigated at the expense of inferring over a smaller region and considering fewer data. Another possibility to reduce issues from non-uniformly distributed tagging locations could be to truncate the first days from each track. This was not adopted here because many tracks presented relatively short periods of locations within the survey area, and such truncation would eliminate a substantial amount of information from the dataset. The extent of the area of inference also has a major influence on where pseudo-absences are randomly placed to represent the background environment where animals were not observed. The restriction of the area considered in this study potentially also minimised the effects related to that issue, because areas further north from the limits of the survey area (Fig. 1) were visited by only 2 animals (out of 113). Also, because of the imbalance in the tagging locations and number of animals tagged across years (Table 1), data from all years were pooled to model distribution. Therefore, it was impossible to evaluate temporal variation in distribution and habitat use patterns. Data were assumed to be representative for the population as a whole and for the period between 2003 to 2013. Year was not considered as a covariate in the analysis, and inferences presented here must be interpreted as the overall distribution pattern for that population. Investigating temporal variation in distribution could help us understand the potential

expansion of population range in the breeding area (Pavanato et al. 2018), but spatiotemporally balanced data within the area of inference are needed for that. One possible option is to restrict the area of inference in ways that allow meeting the above criteria, at the expense, again, of reducing the amount of data to be analysed. There is evidence that this population was increasing near the upper limit of the rate expected during the period considered here (Wedekin et al. 2017), and we suggest that the potentially resulting distributional shifts would be

better assessed with line transect data modelling (Table S5). To investigate the potential expansion of the breeding range, however, tracking data could indicate areas used by animals that were not designed to be surveyed.

The first step of covariate selection (i.e. observing ‘robust standard errors’ from model fitting with SALSA) led to discarding covariates that were not estimated precisely enough to be considered significant, even before the residual autocorrelation was accounted for. Because serial autocorrelation may cause non-important covariates to appear more significant (Aarts et al. 2008), GEEs permit realistic estimation of uncertainty for covariate–response relations. However, such fitted relations can be difficult to interpret (Pirotta et al. 2011), especially for those covariates with wide confidence intervals in their fitted relationships (Fig. 2). The second step of covariate selection (observing the QICu scores) has been used in previous studies analysing similar data to the present and is a reasonable way to account for the residual auto-correlation issue (Pirotta et al. 2011, Jones et al. 2017). The last step of covariate selection (marginal p-values) was the final check for contribution of covariates to the models. The combination of the 3 criteria adopted improves the chances of only retaining important covariates related to animal distribution. Despite the above criteria, all covariates but SST presented large confidence intervals in their fitted relationships (Fig. 2), precluding detailed interpretation of those results.

#### 4.2. Habitat use from humpback whale tracking data

Like the model used by Bortolotto et al. (2017) to investigate habitat use, latitude and longitude were

not considered in the HUM. The present final HUM adds wind speed to the already identified important covariates related to habitat use by humpback whales in the area (Martins et al. 2001, Bortolotto et al. 2017, Pavanato et al. 2018). However, the combination of SST and wind speed, represented by shelter, had been already identified as important in the line transect models. Using autoregressive models applied to line transect data, Pavanato et al. (2018) identified bathymetry and distance to the shore to be related to the occurrence of humpback whale groups in the area. Wind speed was also considered as a candidate covariate in that study, but it was not retained in that final model. In the present study, it is unclear what is the relation between wind and whale occurrence from the partial plot for that covariate (Fig. 2) or from its estimated coefficients (Table S3). Looking at the fitted coefficients for shelter (Fig. 2), wind seems to be important when the water is relatively colder. However, present results for shelter suggest a contradicting interpretation of its importance than from the line transect data models of Bortolotto et al. (2017), which is likely due to the high uncertainty in the shelter coefficients (Fig. 2). Other covariates in the final HUM include current speed and distance from the shelf-break and to the shore, which may be related to calf survival probabilities, protection against predators or habitat selection for specific reproduction-related groups (Corkeron & Connor 1999, Félix & Botero-Acosta 2011). Here, tracking data were used to investigate the population distribution overall; therefore, sex and other individual characteristics were not incorporated in the models. The reason for not using this information was that there is no clear way to allocate sex and other individual characteristics to pseudo-absences within the GEE framework. One option would be to fit models separately for males and females because this information is available (Table S1). However, sex was unbalanced in the present data, with most animals being females. Distribution patterns found by Bortolotto et al. (2017), and now supported here, are consistent with what could be expected for population distribution driven by female habitat selection.

Alternative approaches to investigate habitat use from tracking data include inferring animal behaviour, which can be related to specific biological and environmental features (McClintock et al. 2015, Jonsen 2016, Roncon et al. 2018, Jonsen et al. 2019). There are also methods to study spatial distribution from tagging data which do not require creating pseudo-absences, such as point process models (Johnson et al. 2013) or by considering animal tracks in a

grid over the study area as a result of a Markov process (Whitehead & Jonsen 2013). Movement models can be used to quantify the relationship between covariates and the probability of animals being in specified behavioural states (Roncon et al. 2018, Jonsen et al. 2019), and therefore, to investigate habitat use, not necessarily from a spatial perspective. In that sense, direct comparisons to line transect models (Table S5) as presented here would be more challenging. For instance, that type of model would be very useful to investigate habitat use in more dynamic systems, such as humpback whale feeding grounds where whales are constantly tracking suitable foraging conditions (Bamford et al. 2022) in relation to highly dynamic ocean features (i.e. upwellings, eddies, fronts).

### 4.3. Management implications

It is certainly important to investigate the potential effect of human activities on the occurrence of whales in the area (Pavanato et al. 2018). To evaluate this, data from before the presence of such activities are needed, or drastic enough changes in their distribution or intensity must happen before being able to identify a redistribution, coinciding with the change in activities. However, the fact that the 2 methods considered here agreed in terms of distribution and habitat use inferences means that either could be used to inform conservation or management actions. Alternatively, a simpler analysis of overlap between whale distribution and the distribution of potential harmful human activities (Martins et al. 2013) could also contribute to such an investigation.

Modelling animal distribution using line transect data may have some important advantages in data collection, survey design and modelling techniques, such as being able to control the places visited in the survey design (Table S5). Line transect surveys also allow abundance estimates to be derived when the assumptions of distance sampling are dealt with (Buckland et al. 2015). In contrast, satellite tagging of whales allows data to be collected remotely, meaning that the field work necessary for the method is restricted to that of tagging the animals. Also, the data may allow different approaches for investigating habitat use and distribution, because they include information about movement of individuals over time (Matthiopoulos & Aarts 2010). Because close proximity to the animal is required for tagging, other useful detailed information may also be simultaneously collected. For example, it is common to col-

lect skin and blubber biopsies during the tagging procedure. Important information such as sex, hormone levels and contaminants can be determined from those samples (Heide-Jørgensen et al. 2006, Reisinger et al. 2014). Tagging whales is a more invasive field procedure (Alves et al. 2010, Andrews et al. 2019), but provides data that are impossible to obtain from line transect surveys only.

The 2 methods compared here are complementary and may be used in conjunction to expand the spatiotemporal coverage of studies on distribution or habitat use, therefore providing better information to evaluate the need for, and to implement, conservation and management actions when needed. Survey efforts to investigate distribution of a population can be split between tagging some animals in one portion of the area of interest and surveying another portion, enhancing data collection while reducing logistical costs. Therefore, distribution and habitat use investigations aiming at informing conservation of large whales can be more easily and realistically implemented.

**Acknowledgements.** The Monitoring Whales by Satellite Project (Projeto Monitoramento de Baleias por Satélite, PMBS) research cruises were sponsored by Shell Brasil. The Federal University of Rio Grande (FURG) and the crew of the RV 'Atlântico Sul' provided essential support during fieldwork. The PhD work of G.A.B. was funded by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq; Science Without Borders, scholarship number 208203/2014-1). We are thankful to Luiz Cláudio Alves, Claudia Bethlem, Franciele Castro, Marco Aurélio Crespo, Doug DeMaster, Ygor Geyer, Daniela Godoy, Natália Mamede, Gustavo Miranda, Igor Morais, Jose Luis Pizzorno, Paulo Simões-Lopes, Sarah Stutz, Suzana Stutz, Federico Sucunza and Glenn VanBlaricom for their essential assistance in data collection and fieldwork. Debbie Russell and Vincent Ridoux reviewed an early version of this manuscript. The views expressed here are those of the authors and do not necessarily reflect the views of the US National Marine Fisheries Service–NOAA Fisheries. Tagging of animals was conducted under permits issued by the Brazilian Institute of the Environment and Renewable Natural Resources (IBAMA, permit #009/02/CMA/IBAMA; #02001.000085/02-27; ICMBio #11523–1). This project was approved by the School of Biology Ethics Committee, University of St Andrews (ref SEC18010).

#### LITERATURE CITED

- Amante C, Eakins BW (2009) ETOPO1 1 Arc-Minute Global Relief Model: procedures, data sources and analysis. NOAA Tech Memo NESDIS NGDC-24. National Geophysical Data Center, NOAA, Boulder, CO
- Andrews RD, Pitman RL, Ballance LT (2008) Satellite tracking reveals distinct movement patterns for Type B and Type C killer whales in the southern Ross Sea, Antarctica. *Polar Biol* 31:1461–1468
- Andrews RD, Baird RW, Calambokidis J, Goertz CEC and others (2019) Best practice guidelines for cetacean tagging. *J Cetacean Res Manag* 20:27–66
- Andriolo A, Kinas PG, Engel MH, Albuquerque Martins CC, Rufino AM (2010) Humpback whales within the Brazilian breeding ground: distribution and population size estimate. *Endang Species Res* 11:233–243
- Bamford CCG, Jackson JA, Kennedy AK, Trathan PN and others (2022) Humpback whale (*Megaptera novaeangliae*) distribution and movements in the vicinity of South Georgia and the South Sandwich Islands Marine Protected Area. *Deep Sea Res II* 198:105074
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol* 3: 327–338
- Bivand RS, Pebesma E, Gomez-Rubio V (2013) Applied spatial data analysis with R, 2nd edn. Springer, New York, NY
- Block BA, Jonsen ID, Jørgensen SJ, Winship AJ and others (2011) Tracking apex marine predator movements in a dynamic ocean. *Nature* 475:86–90
- Bortolotto GA, Danilewicz D, Andriolo A, Secchi ER, Zerbini AN (2016) Whale, whale, everywhere: increasing abundance of western South Atlantic humpback whales (*Megaptera novaeangliae*) in their wintering grounds. *PLOS ONE* 11:e0164596
- Bortolotto GA, Danilewicz D, Hammond PS, Thomas L, Zerbini AN (2017) Whale distribution in a breeding area: spatial models of habitat use and abundance of western South Atlantic humpback whales. *Mar Ecol Prog Ser* 585: 213–227
- Bortolotto GA, Thomas L, Hammond P, Zerbini AN (2021) Alternative method for assessment of southwestern Atlantic humpback whale population status. *PLOS ONE* 16: e0259541
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling. Oxford University Press, Oxford
- Buckland ST, Rexstad EA, Marques TA, Oedekoven CS (2015) Distance sampling: methods and applications. In: Robinson AP, Buckland ST, Reich P, McCarthy M (eds) *Methods in statistical ecology*. Springer International Publishing, London
- Carter MI, Boehme L, Cronin MA, Duck CD and others (2022) Sympatric seals, satellite tracking and protected areas: habitat-based distribution estimates for conservation and management. *Front Mar Sci* 9:875869
- Cartwright R, Gillespie B, Labonte K, Mangold T, Venema A, Eden K, Sullivan M (2012) Between a rock and a hard place: habitat selection in female–calf humpback whale (*Megaptera novaeangliae*) pairs on the Hawaiian breeding grounds. *PLOS ONE* 7:e38004
- Castro FR, Mamede N, Danilewicz D, Geyer Y, Pizzorno JLA, Zerbini AN, Andriolo A (2014) Are marine protected areas and priority areas for conservation repre-
- Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J (2008) Estimating space-use and habitat preference from wildlife telemetry data. *Ecography* 31: 140–160
- Alves LCPDS, Moreira S, Simões-Lopes PC, Andriolo A (2010) Behavioral responses of humpback whales, *Megaptera*

- sentative of humpback whale breeding habitats in the western South Atlantic? *Biol Conserv* 179:106–114
- ✦ Corkeron PJ, Connor RC (1999) Why do baleen whales migrate? *Mar Mamm Sci* 15:1228–1245
- ✦ Dalla Rosa L, Secchi ER, Maia YG, Zerbini AN, Heide-Jørgensen MP (2008) Movements of satellite-monitored humpback whales on their feeding ground along the Antarctic Peninsula. *Polar Biol* 31:771–781
- ✦ Dee DP, Uppala SM, Simmons AJ, Berrisford P and others (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597
- ✦ Derville S, Torres LG, Garrigue C (2018) Social segregation of humpback whales in contrasted coastal and oceanic breeding habitats. *J Mammal* 99:41–54
- ✦ Elith J, Graham C, Valavi R, Abegg M and others (2020) Presence-only and presence-absence data for comparing species distribution modeling methods. *Biodivers Inform* 15:69–80
- ✦ ESR (Earth Space Research) (2009) OSCAR third degree resolution ocean surface currents, PO.DAAC., version 1. [podaac.jpl.nasa.gov/dataset/OSCAR\\_L4\\_OC\\_third-deg](https://podaac.jpl.nasa.gov/dataset/OSCAR_L4_OC_third-deg) (accessed 4 Apr 2016)
- ✦ Félix F, Botero-Acosta N (2011) Distribution and behaviour of humpback whale mother–calf pairs during the breeding season off Ecuador. *Mar Ecol Prog Ser* 426:277–287
- Fox J, Weisberg S (2019) *An R companion to applied regression*, 3rd edn. Sage, Thousand Oaks, CA
- ✦ Garrigue C, Zerbini AN, Geyer Y, Heide-Jørgensen MP, Hanaoka W, Clapham P (2010) Movements of satellite-monitored humpback whales from New Caledonia. *J Mammal* 91:109–115
- ✦ Gonçalves MIC, Sousa-Lima RS, Teixeira NN, Morete ME, Carvalho GH, Ferreira HM, Baumgarten JE (2018) Low latitude habitat use patterns of a recovering population of humpback whales. *J Mar Biol Assoc UK* 98(Spec Issue 5): 1–10
- Hair JF Jr, Black WC, Babin BJ, Anderson RE (2014) *Multivariate data analysis*, 7th edn. Pearson, Essex
- Hardin JW, Hilbe JM (2013) *Generalized estimating equations*, 2nd edn. Chapman Hall/CRC, Boca Raton, FL
- ✦ Hazen EL, Abrahms B, Brodie S, Carroll G, Welch H, Bograd SJ (2021) Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. *Mov Ecol* 9:5
- ✦ Heide-Jørgensen MP, Nordøy ES, Øien N, Folkow LP and others (2001) Satellite tracking of minke whales (*Balaenoptera acutorostrata*) off the coast of northern Norway. *J Cetacean Res Manag* 3:175–178
- ✦ Heide-Jørgensen MP, Laidre KL, Jensen MV, Dueck L, Postma LD (2006) Dissolving stock discreteness with satellite tracking: bowhead whales in Baffin Bay. *Mar Mamm Sci* 22:34–45
- ✦ Højsgaard S, Halekoh U, Yan J (2016) Package ‘geepack’, version 1.2-1. <https://cran.r-hub.io/web/packages/geepack/>
- ✦ Johnson DS, Hooten MB, Kuhn CE (2013) Estimating animal resource selection from telemetry data using point process models. *J Anim Ecol* 82:1155–1164
- ✦ Jones EL, Sparling CE, McConnell BJ, Morris CD, Smout S (2017) Fine-scale harbour seal usage for informed marine spatial planning. *Sci Rep* 7:11581
- ✦ Jonsen I (2016) Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Sci Rep* 6:20625
- ✦ Jonsen ID, McMahon CR, Patterson TA, Auger-Méthé M, Harcourt R, Hindell MA, Bestley S (2019) Movement responses to environment: fast inference of variation among southern elephant seals with a mixed effects model. *Ecology* 100:e02566
- ✦ JPL MUR MEaSURES Project (2010) GHRSSST level 4 MUR global foundation sea surface temperature analysis, PO.DAAC, version 2. <https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR> (accessed 14 Jun 2016)
- ✦ Lopez R, Malardé JP, Danès P, Gaspar P (2015) Improving Argos Doppler location using multiple-model smoothing. *Anim Biotelem* 3:32
- Martins CCA, Morete ME, Engel MH, Freitas AC, Secchi ER, Kinan PG (2001) Aspects of habitat use patterns of humpback whales in the Abrolhos Bank, Brazil, breeding ground. *Mem Queensl Mus* 47:563–570
- ✦ Martins CCA, Andriolo A, Engel MH, Kinan PG, Saito CH (2013) Identifying priority areas for humpback whale conservation at Eastern Brazilian Coast. *Ocean Coast Manag* 75:63–71
- Matthiopoulos J, Aarts G (2010) The spatial analysis of marine mammal abundance. In: Boyd IL, Bowen WD, Iverson SJ (eds) *Marine mammal ecology and conservation: a handbook of techniques*. Oxford University Press, Oxford, p 68–97
- ✦ Matthiopoulos J, Wakefield E, Jeglinski JWE, Furness RW and others (2022) Integrated modelling of seabird–habitat associations from multi-platform data: a review. *J Appl Ecol* 59:909–920
- ✦ McClintock BT, London JM, Cameron MF, Boveng PL (2015) Modelling animal movement using the Argos satellite telemetry location error ellipse. *Methods Ecol Evol* 6:266–277
- ✦ Mei Z, Chen M, Li Y, Huang S and others (2017) Habitat preference of the Yangtze finless porpoise in a minimally disturbed environment. *Ecol Model* 353:47–53
- ✦ Miller DL, Burt ML, Rexstad EA, Thomas L (2013) Spatial models for distance sampling data: recent developments and future directions. *Methods Ecol Evol* 4:1001–1010
- ✦ Pan W (2001) Akaike’s information criterion in generalized estimating equations. *Biometrics* 57:120–125
- ✦ Pavanato HJ, Mayer FP, Wedekin LL, Engel MH, Kinan PG (2018) Prediction of humpback whale group densities along the Brazilian coast using spatial autoregressive models. *Mar Mamm Sci* 34:734–754
- ✦ Pendoley KL, Schofield G, Whittock PA, Ierodiaconou D, Hays GC (2014) Protected species use of a coastal marine migratory corridor connecting marine protected areas. *Mar Biol* 161:1455–1466
- ✦ Pirota E, Matthiopoulos J, MacKenzie M, Scott-Hayward L, Rendell L (2011) Modelling sperm whale habitat preference: a novel approach combining transect and follow data. *Mar Ecol Prog Ser* 436:257–272
- R Core Team (2017) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna
- ✦ Redfern JV, Ferguson MC, Becker EA, Hyrenbach KD and others (2006) Techniques for cetacean–habitat modeling. *Mar Ecol Prog Ser* 310:271–295
- ✦ Reisinger RR, Oosthuizen WC, Péron G, Toussaint DC, Andrews RD, De Bruyn PJN (2014) Satellite tagging and biopsy sampling of killer whales at subantarctic Marion Island: effectiveness, immediate reactions and long-term responses. *PLOS ONE* 9:e111835
- ✦ Roberts JJ, Best BD, Mannocci L, Fujioka EI and others (2016) Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. *Sci Rep* 6:22615

- ✦ Roncon G, Bestley S, McMahon CR, Wienecke B, Hindell MA (2018) View from below: inferring behavior and physiology of Southern Ocean marine predators from dive telemetry. *Front Mar Sci* 5:464
- ✦ Runge CA, Watson JEM, Butchart SHM, Hanson JO, Possingham HP, Fuller RA (2015) Protected areas and global conservation of migratory birds. *Science* 350:1255–1258
- ✦ Russell DJF, Hastie GD, Thompson D, Janik V and others (2016) Avoidance of windfarms by harbour seals is limited to pile driving activities. *J Appl Ecol* 53:1642–1652
- ✦ Scott-Hayward L, Oedekoven C, Mackenzie M, Walker C, Rexstad E (2017) MRSea Package (version 1.0-beta): statistical modelling of bird and cetacean distributions in offshore renewables development areas. Contract with Marine Scotland: SB9 (CR/2012/05). University of St. Andrews. <https://tethys.pnnl.gov/publications/mrsea-package-version-001-statistical-modelling-bird-cetacean-distributions-offshore>
- ✦ Sequeira AMM, Heupel MR, Lea MA, Eguíluz VM and others (2019) The importance of sample size in marine megafauna tagging studies. *Ecol Appl* 29:e01947
- ✦ Sing T, Sander O, Beerenwinkel N, Lengauer T (2009) ROCR: visualizing classifier performance in R, version 1.0-4. <http://ipa-tys.github.io/ROCR/>
- ✦ Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inform* 2:1–10
- ✦ Sumner MD, Wotherspoon SJ, Hindell MA (2009) Bayesian estimation of animal movement from archival and satellite tags. *PLOS ONE* 4:e7324
- ✦ Thorne LH, Foley HJ, Baird RW, Webster DL, Swaim ZT, Read AJ (2017) Movement and foraging behavior of short-finned pilot whales in the Mid-Atlantic Bight: importance of bathymetric features and implications for management. *Mar Ecol Prog Ser* 584:245–257
- ✦ Trudelle L, Cerchio S, Zerbini AN, Geyer Y and others (2016) Influence of environmental parameters on movements and habitat utilization of humpback whales (*Megaptera novaeangliae*) in the Madagascar breeding ground. *R Soc Open Sci* 3:160616
- ✦ Ventura F, Matthiopoulos J, Jeglinski JWE (2019) Minimal overlap between areas of high conservation priority for endangered Galapagos pinnipeds and the conservation zone of the Galapagos marine reserve. *Aquat Conserv* 29:115–126
- ✦ Walker CG, Mackenzie ML, Donovan CR, O'Sullivan MJ (2011) SALSA: a spatially adaptive local smoothing algorithm. *J Stat Comput Simul* 81:179–191
- ✦ Wedekin LL, Engel MH, Andriolo A, Prado PI and others (2017) Running fast in the slow lane: rapid population growth of humpback whales after exploitation. *Mar Ecol Prog Ser* 575:195–206
- ✦ Whitehead H, Jonsen ID (2013) Inferring animal densities from tracking data using Markov chains. *PLOS ONE* 8: e60901
- Wood SN (2017) Generalized additive models: an introduction with R, 2nd edn. Chapman Hall/CRC, Boca Raton, FL
- ✦ Zerbini AN, Andriolo A, Heide-Jørgensen MP, Pizzorno JL and others (2006) Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. *Mar Ecol Prog Ser* 313:295–304
- ✦ Zerbini AN, Adams G, Best J, Clapham PJ, Jackson JA, Punt AE (2019) Assessing the recovery of an Antarctic predator from historical exploitation. *R Soc Open Sci* 6:190368

*Editorial responsibility: Lisa T. Ballance,  
Newport, Oregon, USA*

*Reviewed by: S. Derville and 1 anonymous referee*

*Submitted: March 18, 2023*

*Accepted: August 8, 2023*

*Proofs received from author(s): September 29, 2023*