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

Guilherme A. Bortolotto[1,2,3]*, Alexandre N. Zerbini[3,4,5,6], Len Thomas[2], Artur Andriolo[3,6], Philip S. Hammond[1,2]

[1]Sea Mammal Research Unit & Scottish Oceans Institute, University of St Andrews, KY16 8LB, Fife, UK

[2]Centre for Research into Ecological and Environment Modelling, University of St Andrews, KY16 9LZ, Fife, UK

[3]Instituto Aqualie, Juiz de Fora, 36036-330, MG, Brazil

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

[5]Marine Ecology and Telemetry Research, Seabeck, WA 98380, USA

[6]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

*Corresponding author: gabdo@st-andrews.ac.uk

Short title: humpback whale distribution from tracking data

Abstract

Statistical modelling of animal distribution has been widely applied to explain how mobile species use their habitats. The distribution and habitat use of humpback whales *Megaptera novaeangliae* off the eastern coast of Brazil have previously been investigated by modelling visual survey data. Here we modelled their distribution in their breeding range using individual tracking data to compare ecological inferences with those from previous models from line transect data. A Generalized 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

26 temperature (SST), current and wind speeds near the surface, distances to shelf-break and the coast,

27 sea bottom depth and slope, and a factor variable representing "shelter". Two modelling exercises

28 were developed: a Habitat Use Model (HUM) and a Distribution Model (DIM). Covariates retained

29 in the selected HUM were SST, distance to coast and shelf-break, current and wind speeds, and

30 shelter. Covariates retained in the selected DIM were latitude/longitude, current speed, and distances

31 to shelf-break and coast. The modelled relationships between whale occurrence and environmental

32 covariates using tracking data were similar to those using line transect data. Distribution maps were

33 also similar, supporting higher densities around the Abrolhos Archipelago and to its south. We

34 showed that habitat use and distribution of this population in the area could be similarly inferred by

35 modelling either line transect or tracking data. Using these two approaches in conjunction can

36 strengthen the understanding of important ecological aspects of animal populations.

1. INTRODUCTION

Animal distribution 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 environment 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 the distribution and habitat use of animals. For marine mammals, developments to improve equipment and the increasing availability of analysis tools have facilitated their application (Aarts et al. 2008, Russell et al. 2016, Jonsen 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. That means that the

3

63 portion of space that is sampled is only that which is visited by those animals, meaning that there is

64 no information in the data about places that are not visited. Another important difference between the

65 two data types is that line transect data are a snapshot of the location of animals at the time of the

66 survey (i.e. it is cross-sectional data), while tracking data consists of a set of repeated measurements

67 of the location of a sample of animals that have been telemetered (i.e. it is longitudinal data, on

68 multiple individuals). Despite the above differences, the analysis of contrasting data types can provide

69 different views of a species underlying distribution (Matthiopoulos et al. 2022)

70 For telemetry tracking data to be useful for habitat use and distribution modelling, the design of

71 analysis and interpretation of results must consider many potential biases (Elith et al. 2020). For

72 example, the distribution of locations can be greatly influenced by the distribution of tag deployments,

73 resulting in tagging location bias (Block et al. 2011). Also important is the serial autocorrelation in

74 locations along tracks because observations are naturally space-time series (Matthiopoulos & Aarts

75 2010). It is crucial to consider whether tracked individuals represent a small portion of the population

76 for which inferences are intended, meaning that although the sample size of locations may appear

77 large (i.e. many locations), sample sizes for individuals can be small (i.e. few animals). That also

78 raises concerns regarding whether tracked animals realistically represent their population with respect

79 to distribution and habitat use (Sequeira et al. 2019).

80 Animal distribution and habitat use can be inferred using a presence-absence approach. However,

81 that requires information on both where animals where present (i.e. places surveyed where animals

82 were detected) and where animals were absent (i.e. places surveyed with no animals detected).

83 Because of the lack of information on real absences in tracking data (e.g. Pirotta et al. 2011), "pseudo-

84 absence" locations, which represent the available habitat not-visited, are needed for presence-absence

85 distribution modelling of tracking data (Elith et al. 2020). When using pseudo-absences, the number

86 of absences to include in the models is controlled by the user, therefore the ratio of pseudo-absence

87 per presence to be used must be decided. That decision should be guided by model "coefficient

88 stability analysis" (when alternative models with an increasing number of pseudo-absences per

89 presence are ran to visually investigate the stability of the output coefficients; see Supporting

90 Information S4 in Ventura et al. 2019), while ensuring that pseudo-absences have the corresponding

91 fraction of the weight of a presence in the models (i.e. the weighted sum of presences equals the

92 weighted sum of pseudo-absences; Barbet-Massin et al. 2012).

93 A decision about how to define the area of inference is also required. Some studies have used a

94 percentage of kernel density estimates to represent habitats frequently used by the animals (e.g.

95 Pendoley et al. 2014, Mei et al. 2017, Thorne et al. 2017); that percentage is, again, user-controlled.

96 Adopting a pre-defined area, where the distribution and habitat use of a population have already been

97 studied using different data, can contribute to a better understanding of how those animals relate to

98 their environment.

99 Most humpback whale *Megaptera novaeangliae* populations are recovering after the end of global

100 commercial whaling (Zerbini et al. 2019) and the population breeding in coastal waters of Brazil

101 during winter and spring is no exception (Bortolotto et al. 2016, Wedekin et al. 2017). However, the

102 coincident increase in human activities in Brazilian coastal waters, especially those related to oil and

103 gas production (Bortolotto et al. 2017), means that there is a need to understand how these activities

104 may affect distribution and habitat use to inform whether management actions may be necessary to

105 avoid impact on the population. Distribution of this population in the area was investigated in the past

106 using descriptive analysis (Zerbini et al. 2006, Martins et al. 2001, 2013, Castro et al. 2014, Gonçalves

107 et al. 2018), from a difference in densities within line transect survey blocks (Andriolo et al. 2010) or

108 for very restricted areas (Martins et al. 2001, Gonçalves et al. 2018). Analysis of line transect data to

109 investigate the distribution of humpback whales off the coast of Brazil has shown that density was

110 strongly related to sea temperature and bathymetric features (Bortolotto et al. 2017, Pavanato et al.

111 2018). Telemetry data for the species in the area has been used to investigate the relative usage of

112 protected areas (Castro et al. 2014) and movements (Zerbini et al. 2006), but not distribution.

113   Here we present distribution and habitat use models of telemetry tracking data of humpback whales

114   in Brazil and compare model outputs with those from a study using line-transect survey data (i.e.

115   Bortolotto et al. 2017). Comparing the results on distribution and habitat from different approaches

116   can strengthen the understanding of important ecological aspects of animal populations, especially

117   when results support each other. Therefore, management actions, such as spatial planning of natural

118   resources exploration (e.g. oil exploration, fishing), and the need for conservation strategies for

119   populations may be better informed.

120   **2. MATERIALS & METHODS**

121   The area of inference for modelling humpback whale distribution and habitat use in this study was

122   determined by adapting the survey area from Bortolotto et al. (2017), originally comprising the

123   continental shelf from 5°S to 23°S, to allow a direct comparison. The original area was restricted

124   considering both the extent of spatial covariates available and the distribution of tagging locations.

125   Since in a presence/pseudo-absence approach for spatial modelling the area of inference is assumed

126   to be the habitat available for the animals (Aarts et al. 2008), and because the tagging locations were

127   not distributed evenly within the original study area (Figure 1), the northernmost portion of the

128   original area, originally extending north to Natal (Fig. 4 in Bortolotto Ret al. 2017), was excluded to

129   minimize the chances of failing to meet that assumption.

130   *2.1. Data acquisition: whale tracking*

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

132   coast of Brazil. Tagging operations occurred every year between August and December except for

133   2004, when no tagging happened (Supplement, Table S1). Animals were tagged only in good weather

134   conditions, calm seas, and light to moderate winds (i.e. sea state less than 4 in the Beaufort scale).

135   Implantable (n = 108) and LIMPET (Low Impact Minimum Percutaneous Electronic Transmitter;

136   n = 5) tags from Wildlife Computers (Redmond, WA, USA) were used. Implantable tags were

137 attached to the animals using a carbon fiber pole or with an Air Rocket Transmitter System (ARTS;

138 Heide-Jørgensen et al. 2001); LIMPET tags were attached using a crossbow (Andrews et al. 2008).

139 The targeted tag location on the animal was the base of the dorsal fin because this area is most

140 frequently exposed when animals surface to breathe, maximizing the chances of transmitter-satellite

141 signal linking. The tagging boat approached target animals to distances between 3 and 10 m. When

142 the pole was used, a maximum distance of 5 m was required. Photographs of the tagged animal's tail

143 fluke, dorsal fin and attached tag were taken for individual identification and quality control of

144 tagging methods. Tracking data from tagged whales were obtained via the Argos satellite system

145 (www.argos-system.org, *Collecte Localisation Satellites SA*), and tags were programmed to transmit

146 in various duty cycles (e.g. transmitting every second/other day, every four days; Supplement, Table

147 S1), to maximize the longevity of animal tracking. Sexes were genetically identified from skin

148 biopsies collected with crossbow and a modified dart (Dalla Rosa et al. 2008).

149 ### *2.2. Tracking data processing*

150 Because the present whale tracks were obtained via the Argos satellite system, which uses "Doppler-

151 based positioning" (Lopez et al. 2015), each location was estimated with an associated uncertainty.

152 To minimize the number of unrealistic locations, a speed filter was applied to the whale tracking data,

153 implemented with R package *trip* and assuming a maximum swimming speed of 12 km h^{-1} for

154 humpback whales (Garrigue et al. 2010). To reduce the occurrence of long gaps between locations,

155 for which no information was available, tracks were split into two or more if time gaps were longer

156 than 10 days. To account for irregular time intervals and uncertainty in estimated locations, humpback

157 whale tracks were re-estimated using a hierarchical form of a "first-Difference-Correlated-Random

158 Walk model" (DCRW; Jonsen 2016). That model was fitted with R package *bsam* (Version 1.1.2;

159 Jonsen 2016) and used to predict two locations per day (i.e. time step of 12 hours) for each track.

160 Default *bsam* package model diagnostics and plots of predicted locations *versus* original locations

161 were inspected to check for both model convergence and whether derived locations were aligned with

162 observed locations. Data from individuals for which the model did not converge were discarded from

163 the analysis. A summary of the data used in the analysis is presented in Table 1 (see Supplement,

164 Table S1, for details). The uncertainty associated with the estimated locations was not considered in

165 the distribution and habitat use models.

166 Because the objectives here were related to investigating distribution in the breeding area and to

167 compare results to a previous study, derived locations beyond the limits of the survey area, in offshore

168 waters or during migration, for example, were censored from the dataset (Figure 1) before the

169 distribution models were built.

170 *2.3. Covariates*

171 Environmental covariate values used in presence/pseudo-absence spatial models were extracted from

172 published datasets for the post-processed track locations. Candidate covariates considered in the

173 models were chosen to permit comparison to a previous modelling study using line transect data from

174 humpback whales in the same region (Bortolotto et al. 2017). Current speed close to the surface

175 (Curr.sp) values were extracted from the *OSCAR Third Degree Sea Surface Velocity* dataset (ESR,

176 2009), with resolution $0.33 \times 0.33°$ (latitude × longitude) in 5-day intervals. Daily values for sea

177 surface temperature (SST), with resolution $0.01 \times 0.01°$, were extracted from *JPL MUR SST project*

178 dataset (JPL MUR MEaSUREs Project, 2010). Daily wind speed at the surface of the sea (Wind.sp)

179 values were extracted from the *Era-Interim* dataset (Dee et al. 2011), with horizontal resolution

180 $0.125 \times 0.125°$. Depth values were extracted form ETOPO1 (Amante & Eakins 2009), for which the

181 resolution is $0.1 \times 0.1°$. Slope was derived from ETOPO1. Distances from coast (Dist.coast) and from

182 the shelf break (Dist.shelf), represented here by the 500 m depth contour, were measured with the

183 *gDistance* function, *rgeos* R package (version 0.3-26; Bivand & Rundel 2017). To represent regions

184 within the area where animals could be sheltered from rougher weather and colder waters, a factor

185 covariate for "shelter" was created by combining values of SST and wind speed in six classes, defined

186 by quantiles of these covariates, *sensu* Bortolotto et al. (2017). The logarithm of depth was used in

187 the models.

188 Values for the dynamic covariates (i.e. Curr.sp, SST and Wind.sp) were obtained by matching the

189 timestamps of each location with the nearest (in time and horizontal space) covariate value.

2.4. Data analysis and modelling

191 Processed track locations were considered as presences in spatial models. To provide contrasting data,

192 pseudo-absences were generated randomly within the survey area to represent where animals could

193 have been, assuming that the survey area was accessible to all tagged animals. To select the most

194 appropriate ratio of presence to pseudo-absences, different scenarios (i.e. 1:1, 1:3,1:5 and 1:10) were

195 explored to verify model "coefficients stability" (see Supporting Information S4 in Ventura et al.

196 2019). In every model scenario, weighting was applied so the weight of pseudo-absences followed

197 the corresponding fraction, to ensure that the number of presences was equal to the sum of the weights

198 of the pseudo-absences. The model fitted to the data with five pseudo-absences per presence produced

199 the same fitted coefficients as with ten, but different from smaller ratios (1:1 and 1:3), indicating that

200 five pseudo-absences per presence was sufficient to represent the underlying background of the

201 inference area. To match a set of five pseudo-absences to one specific presence, all locations within

202 a set of five pseudo-absences were given the same timestamp as the corresponding presence.

203 Covariate pairs that were strongly correlated (> 0.7), or that had high (> 10; Hair et al. 2014) variance

204 inflation factor scores (*vif* function, *car* R package) when together in a model, were not included

205 simultaneously in the same model.

206 Two objectives guided the inclusion of candidate covariates in the spatial models. For the Distribution

207 Model (DIM), the objective was to create the best distribution map as possible, so all available

208 covariates were considered. For the Habitat Use Model (HUM), the objective was to investigate the

9

209 relationship of whale occurrence and important ecological characteristics, and all covariates but

210 geographic position (latitude and longitude) were considered.

211 To compare the present models with those in Bortolotto et al. (2017), latitude and longitude were not

212 considered in the HUM. This procedure was adopted because these covariates have no logical

213 biological interpretation for habitat use, and because they were both strongly correlated with SST.

214 The DIM included those covariates as potential explanatory terms because its objective was to

215 identify those features that describe most variability in whale occurrence and to generate the best

216 predictive map for potentially informing management. The present DIM map was created to be

217 compared to the AEM map in Bortolotto et al. (2017).

218 Presence vs. pseudo-absence was modelled as a binomial random variable, with probability of

219 presence a smooth function of the continuous covariates, first within a Generalized Additive Model

220 (GAM; Wood 2017) framework and then fitting the GAM using a Generalized Estimating Equation

221 (GEE; Hardin & Hilbe 2002) approach to account for possible autocorrelation, as detailed below.

222 Continuous covariates were included in the models using b-splines, and a logit link function was used.

223 Weighted regression was used, where presences were assigned a weight of 1 and pseudo-absences a

224 weight of 0.2. Model fitting was done using software R (R Core Team, 2017), with functions specified

225 below.

226 For each of the DIM and HUM, covariate selection proceeded in three stages. First, the full model

227 (i.e. model with all candidate covariates) was fitted, with the number and location of each b-spline

228 knot selected using the Spatially Adaptive Local Smoothing Algorithm (SALSA; Walker et al. 2011),

229 implemented with the *MRSea* R package (version 1.0.beta; Scott-Hayward et al. 2017). The

230 maximum number of knots were restricted to eight for one-dimensional smoothers and to 20 for two-

231 dimensional smoothers (i.e. for geographic position), to prevent overfitting of smooth terms

232 (Wood 2017). Continuous covariates for which no knot was indicated as significant, at α = 0.05, were

233 not considered in subsequent steps.

234 In the second stage, a backwards covariate selection process was followed. The model from the first

235 step was refitted using a GEE to accommodate residual autocorrelation. For this, the data were

236 organized in correlation panels, with a panel for each set of presences within a track (one panel per

237 track) and a different panel for each pseudo-absence (one panel per pseudo-absence). Using this panel

238 structure was congruent with the assumption that model residuals within a track were correlated, but

239 that residuals in different tracks were not, and that residuals for pseudo-absences were mutually

240 independent. Model fitting used the *geeglm* function within the *geepack* R package (version 1.2-1;

241 Højsgaardet al. 2006). The fitted model's QICu score (Pan, 2001) was calculated. Then, a series of

242 models were fitted, leaving one covariate out at a time, and QICu scores were calculated for each of

243 these models. The model with the lowest QICu score was retained. If this model had fewer covariates

244 than the full model, then the process was repeated to drop another covariate. This was repeated until

245 dropping covariates did not result in any further decrease in QICu.

246 In the third stage, the statistical significance of remaining covariates was assessed using marginal p-

247 values, via the function *getPvalues* from *MRSea* R package (Scott-Hayward et al., 2017). Any

248 covariates that were non-significant using an α-level of 0.05 were dropped and the GEE re-fitted.

249 For the selected DIM and HUM, model performance was verified with Receiver Operating

250 Characteristic (ROC) curves and confusion matrices, as per Pirotta et al. (2011), using R package

251 *ROCR* (version 1.0-7; Sing et al. 2009). The ROC and confusion matrix can be used to calculate

252 percentages of false positives and false negatives expected for the model, by comparing the predicted

253 values to the observed.

254 For comparison to the distribution maps presented in Bortolotto et al. (2017), prediction grids

255 (8 by 8 km cells) containing covariate values from 2008 and 2012 (the years for which distribution

256   models from sighting data were developed in that study) were used for predictions using the selected

257   DIM. The contribution of each covariate in the final HUM was visualized with partial plots, with

258   confidence intervals based on the GEE estimated uncertainty, adapting custom R code from Pirotta

259   et al. (2011).

260   **3.   RESULTS**

261   Data from 113 tags, deployed from 2003 to 2012, were available for the analysis. Model outputs from

262   package *bsam* (e.g. posterior densities and model convergence plots) indicated poor model fit and/or

263   convergence for whale tracks with less than 12 locations or less than five days of tracking. Those

264   tracks were therefore excluded from the analysis. After filtering, interpolated locations from 87 tracks

265   were available for investigating distribution and habitat use (Figure 1), comprising 62 tracks from

266   females (mothers with calves), 19 from males and six from animals of unidentified sex (Table 1).

267   Because five pseudo-absences were created per each presence (i.e. interpolated locations within the

268   survey area; n = 3,071), 15,355 pseudo-absences were used, totaling 18,426 locations to be modelled.

269   The final HUM model included smooth terms for current speed, SST, wind speed, coast distance and

270   distance to the shelf-break, and the factor variable shelter (Table 2). As a result of using GEEs for

271   dealing with autocorrelation in the data, confidence intervals for fitted relationships between the

272   response variable and the covariates were very wide (i.e. high uncertainty), except for SST (Figure

273   2). Depth had no statistically significant knots when the full model was fitted with SALSA and was

274   not considered further. The fitted relationship for the SST covariate showed a clear peak around 24-

275   25ºC (Figure 2), similar to Bortolotto et al. (2017).

276   The final DIM included latitude and longitude, current speed, and distances to coast and shelf break

277   (Supplement, Table S4). Shelter and wind had no significant knots at the first step of covariate

278   selection, and depth had a non-significant marginal p-value in the last step of covariate selection for

279   the DIM (Table 2). Overall, higher probabilities of encountering whales were predicted for the region

12

around the Abrolhos archipelago and to the south, supporting findings from Bortolotto et al. (2017) (Figure 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 the area under the ROC curve (AUC), 0.691 and 0.732, respectively (Table 3).

4. DISCUSSION

We investigated the distribution and habitat use of humpback whales off Brazil with spatial models applied to locations obtained through animal tracking and to spatial covariates. To generalize 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. Also, because here we considered whales on a breeding area, the unbalance in the reproductive stage (or animal sex) in our tracked animals (see Results; Supplement Table S1) very likely prevents them from accurately representing the population in general. That is because, animals in different reproductive stages are known to use their 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 in describing mainly the differences in the

13

305 implementation and interpretation of modelling of either data from individual tracking or from line

306 transect sampling.

307 *Model considerations and constraints*

308 Spatiotemporal correlation in track locations and other common issues in applying presence/absence

309 modelling for distribution and habitat use (Aarts et al. 2008) were carefully considered in the analysis.

310 Despite differences in the sampling unit of the tracking and survey data (individual vs. space) and

311 statistical tools used, some of the outputs from modelling track locations in a presence/pseudo-

312 absence modelling approach support findings from distribution models fitted to line transect data

313 (Bortolotto et al. 2017). Although the uncertainty in most of the covariate-response (telemetry data)

314 relations fitted here was much higher (Figure 2), SST showed a clear peak around the same range of

315 temperatures in both the present study and in the line transect modelling (Supplement, Figure S3),

316 which probably reflects how strongly habitat use for these animals is related to temperature in their

317 breeding grounds (see Bortolotto et al. 2017 for a discussion on the role of SST for habitat use of

318 humpback whales). The apparent agreement between the two methods is not only reassuring for what

319 is known about habitat use for this population, but also shows that similar, possibly complementary,

320 conclusions can be drawn on this regardless of which of the two methods is used, given enough

321 sample size and careful consideration of potential bias sources. However, differences such as the

322 patch of present lower probability of whale occurrence that overlapped with the area presenting the

323 highest densities from Bortolotto et al. (2017), need to be considered at smaller scales.

324 It is important to note that, by creating pseudo-absences from random locations in the entire area, it

325 was assumed that the entire survey area was available to the animals at any time (Soberón and

326 Peterson 2005). However, this is an arbitrary decision and the available area from the animals'

327 perspective could be different (Aarts et al. 2008, Hazen et al. 2021). For example, an animal tagged

328 in the vicinity of the Abrolhos archipelago (Figure 1) would take at least two days to reach the

14

329   southernmost portion of the survey area, swimming in a straight line at 12 km h^{-1} (i.e. the maximum

330   travel speed [Garrigue et al. 2010] used as a criteria in the data pre-processing/filtering).

331   The restriction of the area of inference (survey area) in comparison to that of Bortolotto et al. (2017;

332   compare north extents of the upper and the lower panels of Figure 3) was adopted to reduce the effects

333   of violating the area availability assumption (Hazen et al. 2021). Further restrictions could be

334   investigated at the expense of inferring over a smaller region and considering fewer data. Another

335   possibility to reduce issues from non-uniformly distributed tagging locations could be to truncate the

336   first days from each track. This was not adopted here because many tracks presented relatively short

337   periods of locations within the survey area and such truncation would eliminate a substantial amount

338   of information from the dataset. The extent of the area of inference also has a major influence on

339   where pseudo-absences are randomly placed to represent the background environment where animals

340   were not observed. The restriction of the area considered in this study potentially also minimized the

341   effects related to that issue, because areas further north from the limits of the survey area (Figure 1)

342   were visited by only two animals (out of 113). Also, because of imbalance in the tagging locations

343   and number of animals tagged across years (Table 1) data from all years were pooled to model

344   distribution. Therefore, it was impossible to evaluate temporal variation in distribution and habitat

345   use patterns. Data were assumed to be representative for the population as a whole and for the period

346   between 2003 to 2013. Year was not considered as a covariate in the analysis and inferences presented

347   here must be interpreted as the overall distribution pattern for that population. Investigating temporal

348   variation in distribution could help understanding the potential expansion of population range in the

349   breeding area (Pavanato et al. 2018), but spatiotemporally balanced data within the area of inference

350   is needed for that. One possible option is to restrict the area of inference in ways that allow meeting

351   the above criteria, at the expense, again, of inferring over a much-reduced area. There is evidence that

352   this population was increasing at near the upper limit of the rate expected during the period considered

353   here (Wedekin et al. 2017), and we suggest that the potentially resulting distributional shifts would

354   be better assessed with line transect data modelling (Supplement, Table S5). To investigate the

355   potential expansion of the breeding range, however, tracking data could indicate areas used by

356   animals that were not designed to be surveyed.

357

358   The first step of covariate selection (i.e. observing "robust standard errors" from model fitting with

359   SALSA) led to discarding of covariates that were not estimated precisely enough to be considered

360   significant, even before the residual autocorrelation was accounted for. Because serial autocorrelation

361   may cause non-important covariates to appear more significant (Aarts et al. 2008), GEEs permit

362   realistic estimation of uncertainty for covariate-response relations. However, such fitted relations can

363   be difficult to interpret (Pirotta et al. 2011), especially for those covariates with wide confidence

364   intervals in their fitted relationships (Figure 2). The second step of covariate selection (observing the

365   *QICu* scores) had been used in previous studies analyzing similar data to the present and is a

366   reasonable way to account for the residual auto-correlation issue (Pirotta et al. 2011, Jones et al.

367   2017). The last step of covariate selection (marginal *p-values*) was the final check for contribution of

368   covariates to the models. The combination of the three criteria adopted improves the chances of only

369   retaining important covariates related to animal distribution. Despite the above criteria, all covariates

370   but SST presented large confidence intervals in their fitted relationships (Figure 2), precluding

371   detailed interpretation of those results.

372   ***Habitat use from humpback whale tracking data***

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

374   not considered in the HUM. The present final HUM adds wind speed to the already identified

375   important covariates related to habitat use of humpback whales in the area (Martins et al. 2001,

376   Bortolotto et al. 2017, Pavanato et al. 2018). However, the combination of SST and wind speed,

377   represented by shelter, had been already identified as important in the line transect models. Using

16

378   autoregressive models applied to line transect data, Pavanato et al. (2018) identified bathymetry and

379   distance to the shore to be related to the occurrence of humpback whale groups in the area. Wind

380   speed was also considered as a candidate covariate in that study, but it was not retained in that final

381   model. In the present study, it is unclear what is the relation between wind and whale occurrence from

382   the partial plot for that covariate (Figure 2) or from its estimated coefficients (Supplement, Table S3).

383   Looking at the fitted coefficients for shelter (Figure 2), wind seems to be important when the water

384   is relatively colder, which agrees well with the results found from modelling line transect data.

385   However, present results for shelter suggest a contradicting interpretation of its importance than from

386   the line transect data models from Bortolotto et al. (2017) which is likely due to the high uncertainty

387   in the shelter coefficients (Figure 2). Other covariates in the final HUM include current speed and

388   distance from the shelf and to the shore, which may be related to calf survival probabilities, protection

389   against predators or in habitat selection for specific reproduction-related groups (Corkeron & Connor

390   1999, Félix & Botero-Acosta 2011). Here, tracking data were used to investigate the population

391   distribution overall, therefore sex and other individual characteristics were not incorporated in the

392   models. The reason for not using this information was that there is no clear way to allocate sex and

393   other individual characteristics to pseudo-absences within the GEEs framework. One option would

394   be to fit models separately for males and females because this information is available (Supplement,

395   Table S1). However, sex was unbalanced in the present data with most animals being females.

396   Distribution patterns found in Bortolotto et al. (2017) and now supported here, are consistent with

397   what could be expected for population distribution driven by female habitat selection.

398   Alternative approaches to investigate habitat use from tracking data include inferring animal

399   behavior, which can be related to specific biological and environmental features (McClintock et al.

400   2015, Jonsen 2016, Roncon et al. 2018, Jonsen et al. 2019). There are also methods to study spatial

401   distribution from tagging data which do not require creating pseudo-absences, such as point process

402   models (Johnson et al. 2013) or by considering animal tracks in a grid over the study area as a result

403   of a Markov process (Whitehead & Jonsen 2013). Movement models can be used to quantify the

404   relationship between covariates and the probability of animals being in specified behavioral states

405   (Roncon et al. 2018, Jonsen et al. 2019), therefore, to investigate habitat use, not necessarily from a

406   spatial perspective. In that sense, direct comparisons to line transect models (Table S5) as presented

407   here would be more challenging. For instance, that type of models would be very useful to investigate

408   habitat use in more dynamic systems, such humpback whale feeding grounds where whales are

409   constantly tracking suitable foraging conditions (Bamford et al. 2022) in relation to highly dynamic

410   ocean features (i.e. upwellings, eddies, fronts). Management implications

411   It is certainly important to investigate the potential effect of human activities on the occurrence of

412   whales in the area (Pavanato et al. 2018). To evaluate that, data from before the presence of such

413   activities are needed, or drastic enough changes in the distribution or intensity of them must happen

414   before being able to identify a redistribution, coinciding with the change in activities. However, that

415   the two methods considered here agreed in terms of distribution and habitat use inferences, means

416   that either could be used to informing conservation or management actions. Alternatively, a simpler

417   analysis of overlap between whale distribution and the distribution of potential harmful human

418   activities (Martins et al. 2013) could also contribute to such investigation.

419   Modelling animal distribution using line transect data may have some important advantages in data

420   collection, survey design and modelling techniques, such as being able to control the places visited

421   in the survey design (Table S5**Error! Reference source not found.**). Line transect surveys also allow

422   deriving abundance estimates when the assumptions of distance sampling are dealt with (Buckland

423   et al. 2015). In contrast, satellite tagging of whales allows data to be collected remotely, meaning that

424   the field work necessary for the method is restricted to that of tagging the animals. Also, the data may

425   allow different approaches for investigating habitat use and distribution, because they include

426   information about movement of individuals over time (Matthiopoulos & Aarts 2010). Because close

427   proximity to the animal is required for tagging, other useful detailed information may also be

428   simultaneously collected. For example, it is common to collect skin and blubber biopsies during the

429   tagging procedure. Important information such as sex, hormone levels and contaminants can be

430   determined from those samples (Heide-Jørgensen et al. 2006, Reisinger et al. 2014). Tagging whales

431   is a more invasive field procedure (Alves et al. 2010, Andrews et al. 2019), but provides data that are

432   impossible to obtain from line transect surveys only.

433   The two methods compared here are complementary and may be used in conjunction to expand the

434   spatiotemporal coverage of studies on distribution or habitat use, therefore providing better

435   information to evaluate the need of, and for implementing conservation and management actions

436   when needed. Survey efforts to investigate distribution of a population can be split between tagging

437   some animals in one portion of the area of interest and surveying another portion, enhancing data

438   collection while reducing logistical costs. Therefore, distribution and habitat use investigations

439   aiming at informing conservation of large whales can be more easily and realistically implemented.

440

5. LITERATURE CITED

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 novaeangliae* (Cetacea: Balaenopteridae), to satellite transmitter deployment procedures. Zoologia 27:1–6

Amante C, Eakins BW (2009) ETOPO1 1 Arc-Minute Global Relief Model: procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M (accessed 1 Mar 2016)

Andrews RD, Baird RW, Calambokidis J, Goertz CEC, Gulland FMD, Heide-Jørgensen MP, Hooker SK, Johnson M, Mate B, Mitani Y, Nowacek DP, Owen K, Quakenbush LT, Raverty S, Robbins J, Schorr GS, Shpak OV, Townsend Jr FI, Uhart M, Wells RW, Zerbini AN (2019). Best practice guidelines for cetacean tagging. Journal of Cetacean Research and Management 20:27–66

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 Biology 31:1461–1468

Andriolo A, Kinas PG, Engel MH, Martins CCA, Rufino AM (2010) Humpback whales within the Brazilian breeding ground: distribution and population size estimate. Endangered Species Research 11:233–243

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many?. Methods in Ecology and Evolution 3:327-338

Bivand RS, Pebesma E, Gomez-Rubio V (2013) Applied spatial data analysis with R (2nd edition). Springer, New York

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, Hazen EL, Foley DG, Breed GA, Harrison A, Ganong JE, Swithenbank A, Castleton M, Dewar H, Mate BR, Shillinger GL, Schaefer KM, Benson SR, Weise MJ et al. (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. Marine Ecology Progress Series 585:213–227

Bamford CCG, Jackson JA, Kennedy AK, Trathan PN, Staniland IJ, Andriolo A, Bedriñana-Romano L, Carroll EL, Martin S, Zerbini AN (2022) Humpback whale (*Megaptera novaeangliae*) distribution and movements in the vicinity of South Georgia and the South Sandwich Islands Marine Protected Area. Deep Sea Research Part II: Topical Studies in Oceanography 198: 105074

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, Grecian WJ, Hastie GD, Jessopp M, Matthiopoulos J, McConnell BJ, Miller DL, Morris CD (2022) Sympatric seals, satellite tracking and protected areas: habitat-based distribution estimates for conservation and management. Frontiers in Marine Science 9:875869

506 Cartwright R, Gillespie B, Labonte K, Mangold T, Venema A, Eden K, Sullivan M (2012) Between
507   a rock and a hard place: habitat selection in female-calf humpback whale (*Megaptera*
508   *novaeangliae*) pairs on the Hawaiian breeding grounds. PLoS One 7:e38004

509 Castro FR, Mamede N, Danilewicz D, Geyer Y, Pizzorno JLA, Zerbini AN, Andriolo A (2014) Are
510   marine protected areas and priority areas for conservation representative of humpback whale
511   breeding habitats in the western South Atlantic? Biological Conservation 179:106–114

512 Corkeron PJ, Connor RC (1999) Why do baleen whales migrate? Marine Mammal Science 15:1228–
513   1245

514 Dalla Rosa L, Secchi ER, Maia YG, Zerbini AN, Heide-Jørgensen MP (2008) Movements of satellite-
515   monitored humpback whales on their feeding ground along the Antarctic Peninsula. Polar
516   Biology 31:771-781Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S,
517   Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, Van de Berg L,
518   Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ et al. (2011) The ERA-Interim
519   reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of
520   the Royal Meteorological Society 137:553–597

521 Derville S, Torres LG, Garrigue C (2018) Social segregation of humpback whales in contrasted
522   coastal and oceanic breeding habitats. Journal of Mammalogy 99:41-54

523 Elith J, Graham C, Valavi R, Abegg M, Bruce C, Ferrier S, Ford A, Guisan A, Hijmans RJ, Huettmann
524   F, Lohmann L, Loiselle B, Moritz C, Overton J, Peterson AT, Phillips S, Richardson K, Williams
525   S, Wiser SK, Wohlgemuth T, Zimmermann NE (2020) Presence-only and presence-absence data
526   for comparing species distribution modeling methods. Biodiversity Informatics 15:69–80

527 ESR (2009) OSCAR third degree resolution ocean surface currents, PO.DAAC., Ver. 1.
528   podaac.jpl.nasa.gov/dataset/OSCAR_L4_OC_third-deg (accessed 4 Apr 2016)

529   Félix F, Botero-Acosta N (2011) Distribution and behaviour of humpback whale mother-calf pairs

530   during the breeding season off Ecuador. Marine Ecology Progress Series 426:277–287

531   Garrigue C, Zerbini AN, Geyer Y, Heide-Jørgensen MP, Hanaoka W, Clapham P (2010) Movements

532   of satellite-monitored humpback whales from New Caledonia. Journal of Mammalogy 91:109–

533   115

534   Gonçalves MIC, Sousa-Lima RS, Teixeira NN, Morete ME, Carvalho GH, Ferreira HM, Baumgarten

535   JE (2018) Low latitude habitat use patterns of a recovering population of humpback whales.

536   Journal of the Marine Biological Association of the United Kingdom 98 (special issue 5):1–10

537   Hair Jr JF, Black WC, Babin BJ, Anderson RE (2014) Multivariate Data Analysis. 7th edition. Essex:

538   Pearson.

539   Hardin JW, Hilbe JM (2012) Generalized Estimating Equations (2nd edition). Chapman Hall/CRC,

540   Boca Raton

541   Hazen EL, Abrahms B, Brodie S, Carroll G, Welch H, Bograd SJ (2021) Where did they not go?

542   Considerations for generating pseudo-absences for telemetry-based habitat models. Movement

543   Ecology 9:1-13

544   Heide-Jørgensen MP, Laidre KL, Jensen MV, Dueck L, Postma LD (2006). Dissolving stock

545   discreteness with satellite tracking: bowhead whales in Baffin Bay. Marine Mammal Science

546   22:34–45

547   Heide-Jørgensen MP, Nordøy ES, Øien N, Folkow LP, Kleivane L, Blix AS, Jensen MV, Laidre KL

548   (2001) Satellite tracking of minke whales (*Balaenoptera acutorostrata*) off the coast of northern

549   Norway. Journal of Cetacean Research and Management 3:175–178

550   Højsgaard S, Halekoh U, Yan J (2016) Package 'geepack', Version 1.2-1

551 Johnson DS, Hooten MB, Kuhn CE (2013) Estimating animal resource selection from telemetry data
552   using point process models. Journal of Animal Ecology 82:1155–1164

553 Jones EL, Rendell L, Pirotta E, Long JA (2016) Novel application of a quantitative spatial comparison
554   tool to species distribution data. Ecological Indicators 70:67–76

555 Jones EL, Sparling CE, McConnell BJ, Morris CD, Smout S (2017) Fine-scale harbour seal usage for
556   informed marine spatial planning. Scientific Reports 7:1–11

557 Jonsen I (2016) Joint estimation over multiple individuals improves behavioural state inference from
558   animal movement data. Scientific Reports 6:1–9

559 Jonsen ID, McMahon CR, Patterson TA, Auger-Méthé M, Harcourt R, Hindell MA, Bestley S (2019)
560   Movement responses to environment: fast inference of variation among southern elephant seals
561   with a mixed effects model. Ecological Society of America 100:e02566

562 JPL MUR MEaSUREs Project (2010) GHRSST level 4 MUR global foundation sea surface
563   temperature analysis, PO.DAAC, Ver. 2. podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-
564   MUR (accessed 14 Jun 2016)

565 Lopez R, Malardé JP, Danès P (2015) Improving Argos Doppler location using multiple-model
566   smoothing. Animal Biotelemetry 3:32

567 Martins CCA, Andriolo A, Engel MH, Kinas PG, Saito CH (2013) Identifying priority areas for
568   humpback whale conservation at Eastern Brazilian Coast. Ocean and Coastal Management
569   75:63–71

570 Martins CCA, Morete ME, Engel, MH, Freitas AC, Secchi ER, Kinas PG (2001) Aspects of habitat
571   use patterns of humpback whales in the Abrolhos Bank, Brazil, breeding ground. Memoirs of the
572   Queensland Museum 47:563–570

573 Matthiopoulos J, Aarts G (2010). The spatial analysis of marine mammal abundance. In Boyd IL,

574   Bowen D, Iverson. (eds.) Marine mammal ecology and conservation: a handbook of techniques.

575   Oxford University Press, Oxford

576 Matthiopoulos J, Wakefield E, Jeglinski JW, Furness RW, Trinder M, Tyler G, McCluskie A, Allen

577   S, Braithwaite J, Evans T (2022) Integrated modelling of seabird-habitat associations from multi-

578   platform data: a review. Journal of Applied Ecology 59:909–920

579 McClintock BT, London JM, Cameron MF, Boveng PL (2015) Modelling animal movement using

580   the Argos satellite telemetry location error ellipse. Methods in Ecology and Evolution 6:266–277

581 Mei Z, Chen M, Li Y, Huang S, Haung J, Han Y, Zhu B, Li C, Wang K, Wang D (2017) Habitat

582   preference of the Yangtze finless porpoise in a minimally disturbed environment. Ecological

583   Modelling 353:47–53

584 Miller DL, Burt ML, Rexstad EA, Thomas L (2013) Spatial models for distance sampling data: recent

585   developments and future directions. Methods in Ecology and Evolution 4:1001–1010

586 Pan W (2001) Akaike's information criterion in generalized estimating equations. Biometrics

587   57:120–125

588 Pavanato HJ, Mayer FP, Wedekin LL, Engel MH, Kinas PG (2018) Prediction of humpback whale

589   group densities along the Brazilian coast using spatial autoregressive models. Marine Mammal

590   Science 34:1–21

591 Pendoley KL, Schofield G, Whittock PA, Ierodiaconou D, Hays GC (2014) Protected species use of

592 a coastal marine migratory corridor connecting marine protected areas. Marine Biology 161:1455–

593 1466

594 Pirotta E, Matthiopoulos J, MacKenzie M, Scott-Hayward L, Rendell L (2011) Modelling sperm

595   whale habitat preference: a novel approach combining transect and follow data. Marine Ecology

596   Progress Series 436:257–272

597 R Core Team (2017) R: a language and environment for statistical computing. R Foundation for

598   Statistical Computing, Vienna, Austria

599 Redfern JV, Ferguson MC, Becker EA, Hyrenbach KD, Good C, Barlow J, Kaschner K, Baumgartner

600   MF, Forney KA, Ballance LT, Fauchald P, Halpin P, Hamazaki T, Pershing AJ, Qian SS, Read

601   A, Reilly SB, Torres L, Werner F (2006) Techniques for cetacean-habitat modeling. Marine

602   Ecology Progress Series 310:271–295

603 Reisinger RR, Oosthuizen WC, Péron G, Toussaint DC, Andrews RD, De Bruyn PJN (2014) Satellite

604   tagging and biopsy sampling of killer whales at subantarctic Marion Island: effectiveness,

605   immediate reactions and long-term responses. PLoS ONE 9:1–14

606 Roberts JJ, Best BD, Mannocci L, Fujioka EI, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole

607   TV, Khan CB, McLellan WA (2016) Habitat-based cetacean density models for the US Atlantic

608   and Gulf of Mexico. Scientific reports 6:22615

609 Roncon G, Bestley S, McMahon CR, Wienecke B, Hindell MA (2018) View from below: inferring

610   behavior and physiology of Southern Ocean marine predators from dive telemetry. Frontiers in

611   Marine Science 5:1–23

612 Runge CA, Watson JEM, Butchart SHM, Hanson JO, Possingham HP, Fuller RA (2015) Protected

613   areas and global conservation of migratory birds. Science 350:1255–1258

614 Russell DJF, Hastie GD, Thompson D, Janik V, Hammond PS, Scott-Haywood L, Matthiopoulos J,

615   Jones E, McConnell BJ (2016) Avoidance of windfarms by harbour seals is limited to pile driving

616   activities. Journal of Applied Ecology 53:1642–1652

617 Scott-Hayward L, Oedekoven C, Mackenzie M, Walker C, Rexstad E (2017) MRSea Package

618   (version 1.0-beta): statistical modelling of bird and cetacean distributions in offshore renewables

619   development areas. University of St. Andrews: Contract with Marine Scotland: SB9

620   (CR/2012/05)

621 Sequeira AMM, Heupel MR, Lea MA, Eguíluz VM, Duarte M, Meekan MG, Thums M, Calich HJ,

622   Carmichael RH, Costa DP, Ferreira LC, Fernandéz-Gracia J, Harcourt R, Harrison AL, Jonsen I,

623   McMahon CR, Sims DW, Wilson R P, Hays GC (2019). The importance of sample size in marine

624   megafauna tagging studies. Ecological Applications 29:1344–1360

625 Sing T, Sander O, Beerenwinkel N, Lengauer T (2009) ROCR: visualizing classifier performance in

626   R, version 1.0-4

627 Soberón J, Peterson TA (2005) Interpretation of models of fundamental ecological niches and

628   species' distributional areas. Biodiversity Informatics 2:1-10

629 Thorne LH, Foley HJ, Baird RW, Webster DL, Swaim ZT, Read AJ (2017) Movement and foraging

630   behavior of short-finned pilot whales in the Mid-Atlantic Bight: importance of bathymetric

631   features and implications for management. Marine Ecology Progress Series 584:245–257

632 Trudelle L, Cerchio S, Zerbini AN, Geyer Y, Mayer FX, Jung J, Hervé MR, Pous S, Adam O,

633   Charrassin JB. (2016) Influence of environmental parameters on movements and habitat

634   utilization of humpback whales (*Megaptera novaeangliae*) in the Madagascar breeding ground.

635   Royal Society Open Science 3:160616

636 Ventura F, Matthiopoulos J, Jeglinski JWE (2019) Minimal overlap between areas of high

637   conservation priority for endangered Galapagos pinnipeds and the conservation zone of the

638   Galapagos marine reserve. Aquatic Conservation: Marine and Freshwater Ecosystems 29:115–

639   126

640   Walker CG, Mackenzie ML, Donovan CR, O'Sullivan MJ (2011) SALSA: a spatially adaptive local

641   smoothing algorithm. Journal of Statistical Computation and Simulation 81:179–191

642   Wedekin LL, Engel MH, Andriolo A, Prado PI, Zerbini AN, Marcondes MMC, Kinas PG (2017)

643   Running fast in the slow lane: rapid population growth of humpback whales after exploitation.

644   Marine Ecology Progress Series 575:195–206

645   Whitehead H, Jonsen ID (2013) Inferring animal densities from tracking data using Markov chains.

646   PLoS ONE 8:1–5

647   Wood SN (2017) Generalized additive models: an introduction with R (2nd edition). Chapman

648   Hall/CRC, Boca Raton

649   Zerbini AN, Adams G, Best J, Clapham PJ, Jackson JA, Punt AE (2019) Assessing the recovery of

650   an Antarctic predator from historical exploitation. Royal Society Open Science 6:190368

651   Zerbini AN, Andriolo A, Heide-Jørgensen MP, Pizzorno JL, Maia YG, VanBlaricom GR, DeMaster

652   DP, Simões-Lopes PC, Moreira S, Bethlem C (2006) Satellite-monitored movements of

653   humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. Marine Ecology

654   Progress Series 313:295–304

655   **Table 1.** Summary of the data set (presences only) used for modelling distribution. (Max

656   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*	1	16	20-Aug	27-Aug	16
Total/Overall	87	3071	—	—	—

657   *locations from a whale tagged in 2012, which was tracked for more than 300 days.

658　**Table 2.** Covariates retained in models for both present tracking data and for line transect data.

659　(HUM = Habitat Use Model, DIM = Distribution Model, AEM = Abundance Estimation Model from

660　Bortolotto et al. 2017).

Covariate	Tracking data (present)		Line transect data	
	HUM	DIM	HUM	AEM
Current speed	*	*	*	*
Depth				
Distance to shelf break	*	*	*	*
Distance to coast	*	*	*	*
Shelter	*		*	*
Slope				
SST	*		*	
Wind speed	*			*
Latitude and longitude	—	*	—	*

661

662 **Table 3.** Performance of models fitted to telemetry tracking data of humpback whales. (HUM =

663 Habitat Use Model, DIM = Distribution Model, AUC = area under the ROC curve).

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

664

665
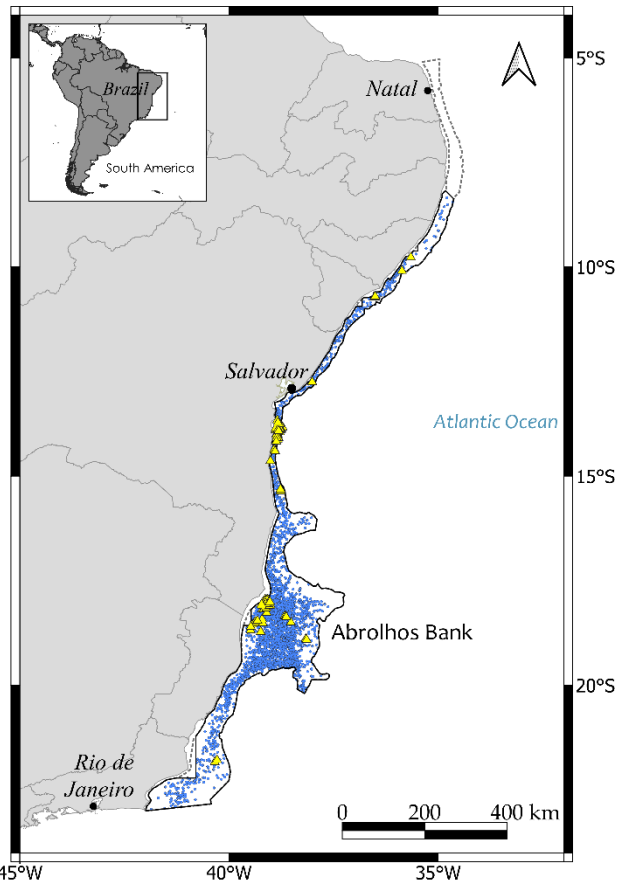
Figure 1. Humpback whale locations used in the distribution and habitat use analysis (blue dots) and tagging locations (yellow triangles). The area of inference (black solid line) was adapted from Bortolotto et al. (2017) (grey dotted line), considering the uneven tagging locations and environmental covariates extent.
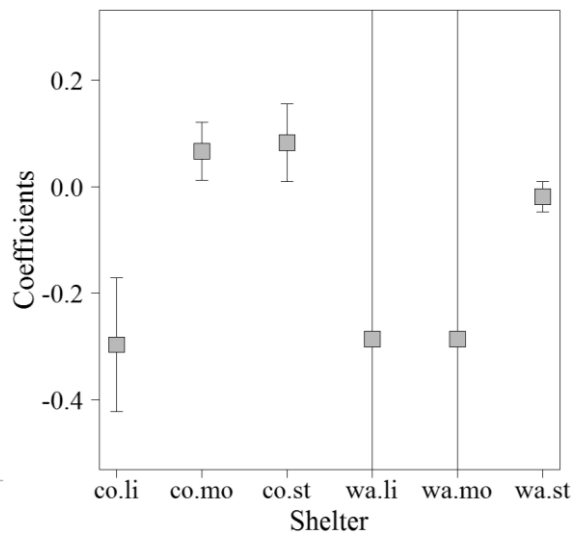
670

671

672
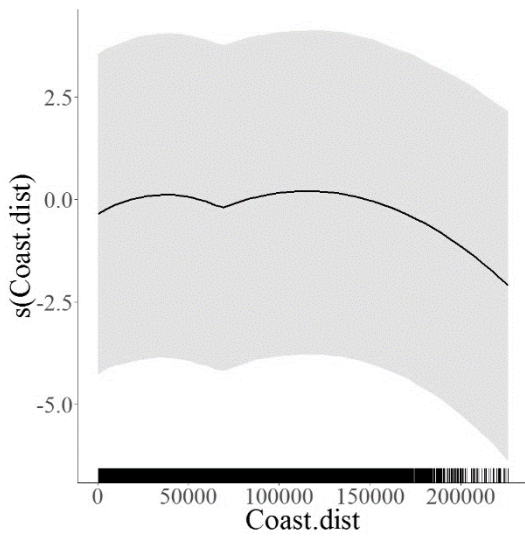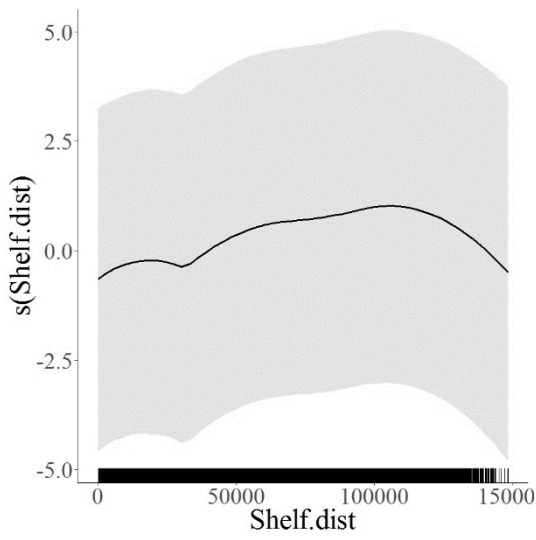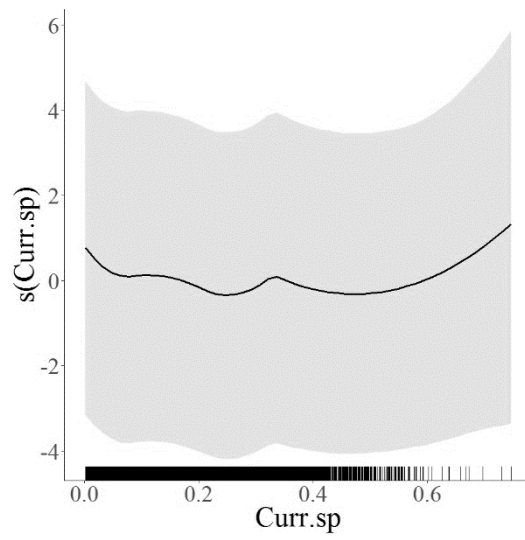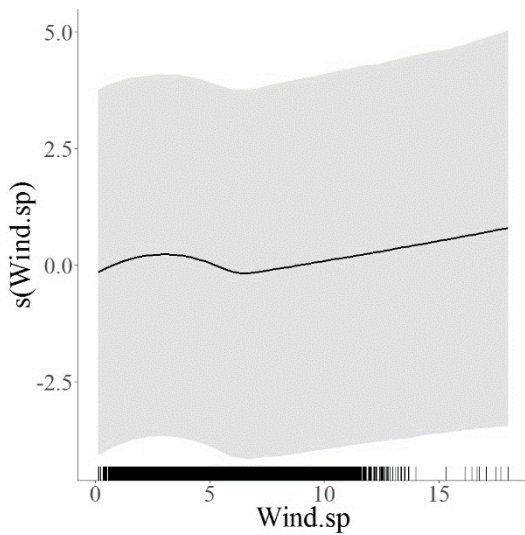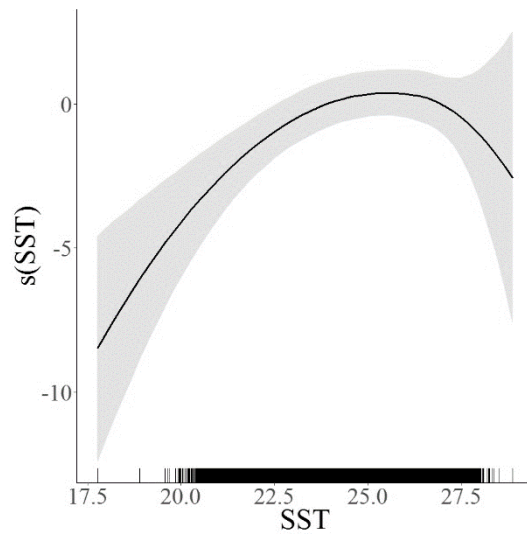
Figure 2. Fitted relationships for smooth functions of covariates in the final habitat use model (HUM).

Error bars in the "Shelter" plot represent 95% normal confidence intervals. (co = cold, wa = warm,

675   li = light, mo = moderate, st = strong, Curr.sp = current speed, SST = sea surface temperature,

676   Wind.sp = Wind speed, Coast.dist = distance to the coast, Shelf.dist = distance to the shelf-break).

677

678

36
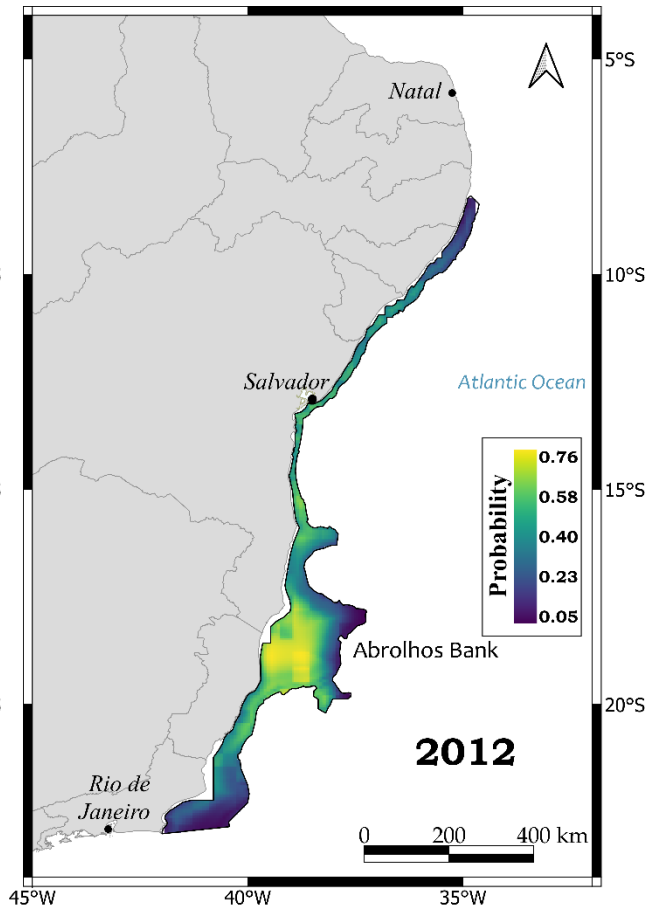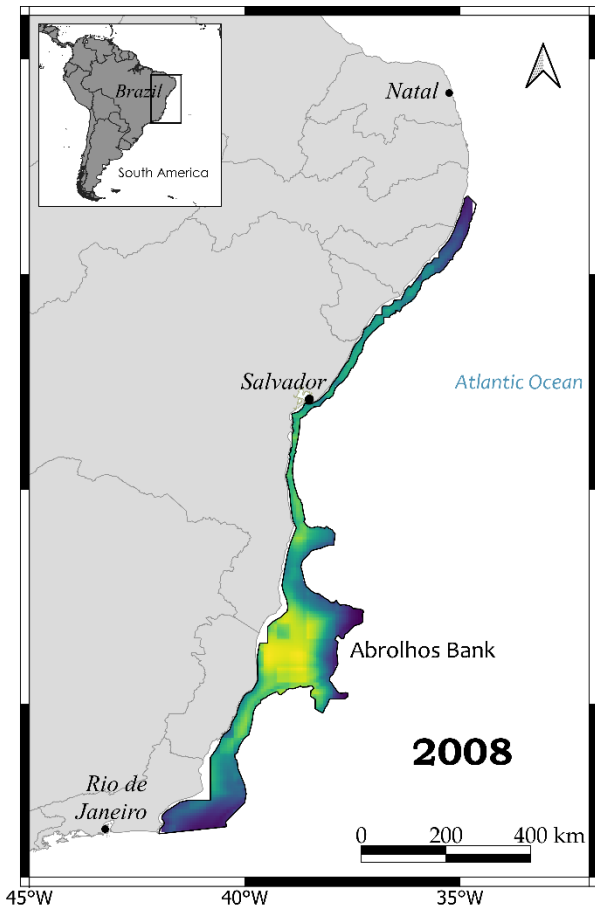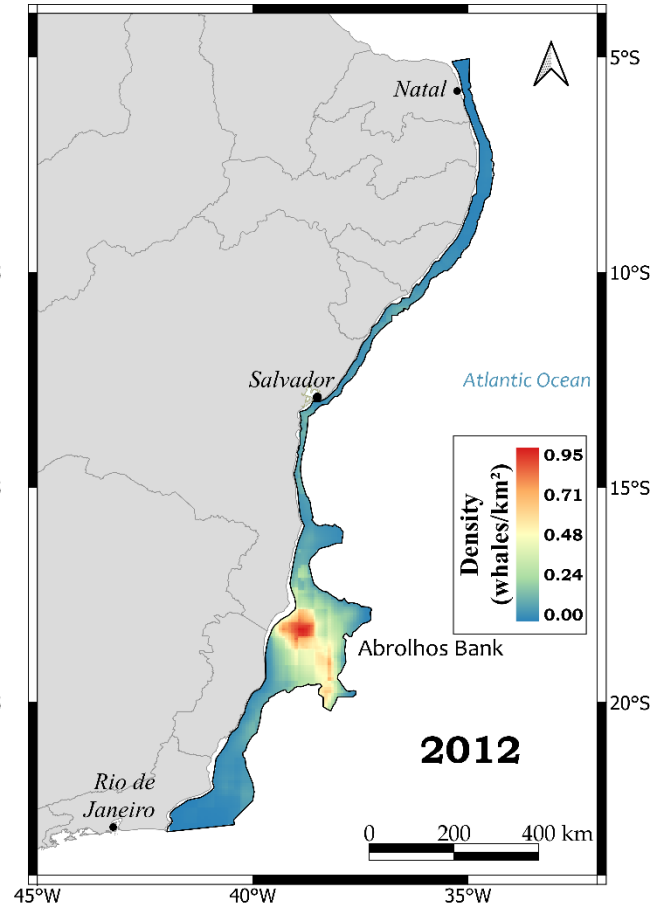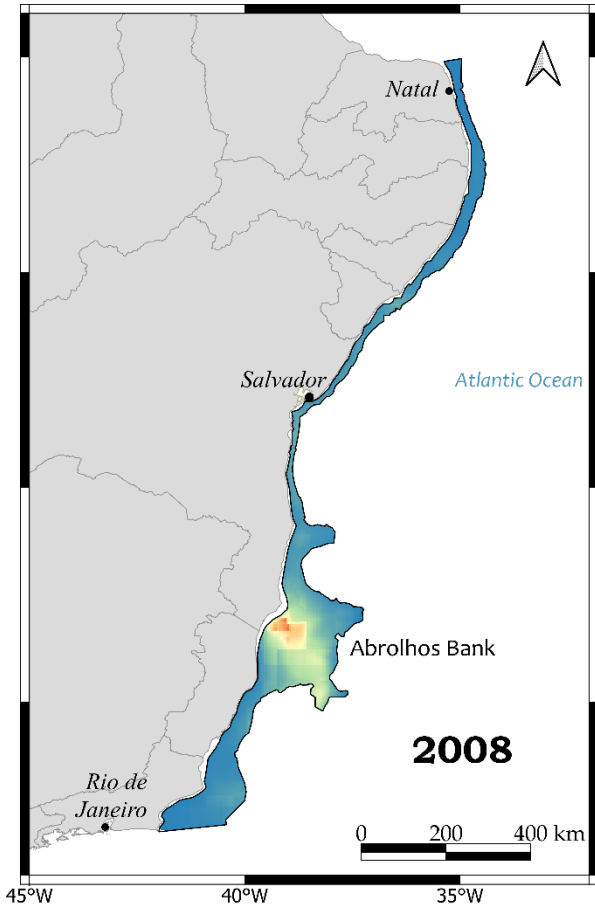
679 **Figure 3.** Upper panel: Occurrence probability surfaces, predicted using the present distribution

680 model (DIM) of tracking data; Lower panel: "AEM" (Abundance Estimation Model) maps adapted

681 from Bortolotto et al. (2017), showing humpback whale density surfaces, for comparison. See

682 Supplement, Figure S2, for uncertainty maps of present DIMs.

683