



## Prediction of humpback whale group densities along the Brazilian coast using spatial autoregressive models

**HELOISE J. PAVANATO**,<sup>1,2</sup>  Laboratório de Estatística Ambiental, Universidade Federal do Rio Grande, Campus Carreiros, Rio Grande, Rio Grande do Sul 96203-900, Brazil; **FERNANDO P. MAYER**, Laboratório de Estatística e Geoinformação, Universidade Federal do Paraná, Centro Politécnico, Curitiba, Paraná 81531-980, Brazil; **LEONARDO L. WEDEKIN** and **MÁRCIA H. ENGEL**, Projeto Baleia Jubarte/Instituto Baleia Jubarte, Caravelas, Bahia 45900-000, Brazil; **PAUL G. KINAS**, Laboratório de Estatística Ambiental, Universidade Federal do Rio Grande, Campus Carreiros, Rio Grande, Rio Grande do Sul 96203-900, Brazil.

### ABSTRACT

At the breeding grounds of most baleen whales the patchiness and gaps in spatial distribution results from interactions between behavior patterns and environmental conditions. We evaluated the influence of environmental factors (bathymetry and distance from shore with quadratic terms, and wind speed), effort, and spatial autocorrelation effects to predict humpback whale group density in the Southwest Atlantic Ocean. Count data of groups by grid cells were fitted with conditional autoregressive models (CAR). Bayesian inference was performed *via* integrated nested Laplace approximation. The best-fit model contained distance from shore and its quadratic term, bathymetry, and the autoregressive component. Occupancy probability was high for the Abrolhos Bank, some cells from the northeast continental shelf and southeast margin, but gaps in occurrence were identified. High densities were estimated in the east continental margin, with the highest density in the Abrolhos Bank, in some cells of the northeast continental margin and in the southernmost area. We report that intermediate distances from the coast, and shallow waters were preferred for breeding and calving activities. We suggest that CAR models may incorporate aggregation mechanisms into habitat modeling and may provide advances in marine mammal analyses by accounting for residual autocorrelation.

**Key words:** *Megaptera novaeangliae*, distribution, habitat modeling, Atlantic Ocean, continental margin, breeding ground, spatial analysis, autocorrelation, INLA, Bayesian inference.

Information about the spatial structure of a population is an important component to guide conservation and management procedures (Stamps 2009). Spatially, some conditions may be recognized as optimal since the performance of individuals might be higher in some locations compared with others (Begon *et al.* 2006). The

<sup>1</sup>Corresponding author: (e-mail: hpavanato@maths.otago.ac.nz).

<sup>2</sup>Current address: Department of Mathematics & Statistics, University of Otago, P.O. Box 56, Dunedin 9016, Otago, New Zealand.

measure of habitat importance assumes that selection of areas is linked to breeding and survival parameters that characterize fitness (Garshelis 2000, Bled *et al.* 2011). Therefore, habitats of high quality refer to sites where individuals may expect a higher probability of fitness (Bled *et al.* 2011).

Most cetacean species respond to variability in marine ecosystems through changes in distribution patterns (Forney 2000) rather than changes in survival and reproductive patterns (Redfern *et al.* 2006). This response is conditioned by their high mobility (Acevedo-Gutiérrez 2009) and dependence on a large amount of prey, which likely responds to bottom-up effects (Estes *et al.* 2006). Because of this, the main factors guiding cetacean distribution are food availability, predation risk, and calf thermoregulation (Corkeron and Connor 1999, Acevedo-Gutiérrez 2009). There are extensive studies that deal with prey availability directly or indirectly, that is, biotic or abiotic features influencing this availability (*e.g.*, Doniol-Valcroze *et al.* 2007, Cañadas and Hammond 2008, Gregr and Trites 2008, Pendleton *et al.* 2009, Bailey and Thompson 2010, Becker *et al.* 2012, Dalla Rosa *et al.* 2012). For example, narwhals (*Monodon monoceros*) of Baffin Bay, West Greenland, select optimal foraging areas in dense pack ice to feed on Greenland halibut, *Reinhardtius hippoglossoides*, regardless of open water availability (Laidre *et al.* 2011).

For baleen whales, which partition feeding and breeding activities through seasonal migration, the spatial distribution in high and low latitudes is driven by different factors (Clapham 2001). While in high latitude feeding grounds, the effects of fronts, sea surface temperature, primary productivity, and bathymetry may predict aggregation of individuals because of high density of prey (Redfern *et al.* 2006), the distribution in low latitude breeding grounds may be driven by energy conservation and avoidance of potential predators in high risk areas (Corkeron and Connor 1999, Ford and Reeves 2008, Clapham 2001, Steiger *et al.* 2008). For example, the highest densities of southern right whales, *Eubalaena australis*, occur in areas protected against waves and wind (Elwen and Best 2004, Seyboth *et al.* 2015), where mother-calf pairs occupy shallow waters near to shore more often than do other groups (Elwen and Best 2004). For North Atlantic right whales a thermal limit in the population distribution seems to exist (Keller *et al.* 2006).

In addition to environmental adequacy, distribution may also be dependent on a series of biological processes playing out on different levels. Population life history, including reproduction, territoriality, and dispersion can cause aggregation and gaps in the spatial distribution (Latimer *et al.* 2006). Baleen whales have a social structure of small, unstable groups wherein the dynamics vary among species and associated breeding strategies (Clapham 2000). Hence, habitat preference results from complex interactions among behavior patterns, biological requirements, and environmental conditions (Ersts and Rosenbaum 2003).

Like most baleen whales, the humpback whale, *Megaptera novaeangliae*, is a migratory species moving seasonally between summer feeding grounds in high latitudes and winter breeding and calving grounds in low latitudes (Dawbin 1956, 1966; Mackintosh 1965, Clapham 2000). The International Whaling Commission (IWC) recognizes seven humpback whale breeding grounds in the Southern Hemisphere (IWC 1998). The population of interest in this study is the Southwest Atlantic humpback whale population (SWA, breeding stock A), which feeds east of the Scotia Sea, around South Georgia and South Sandwich Islands, and breeds along the coast of Brazil (Stevick *et al.* 2006, Zerbini *et al.* 2006, Engel *et al.* 2008, Engel and Martin 2009).

Humpback whale breeding areas are usually associated with reef complexes, islands, and coastal areas (Dawbin 1956, Whitehead and Moore 1982). The pattern for SWA humpback whale population is similar, with high population occurrence on the Abrolhos Bank (Andriolo *et al.* 2006, 2010). It has been hypothesized that the suitability of Abrolhos Bank is related to its low-current issuing from its reefs, and the protection against wind from its islands (Martins *et al.* 2001, Morete *et al.* 2007).

Here, we consider the influence of environmental variables on habitat use together with spatial correlation patterns to predict the density of humpback whales in the Southwest Atlantic Ocean. Specifically, we predicted group density by considering taking into account the effects of bathymetry, wind speed, distance from shore, and autocorrelation between spatial units, assuming that these features might reflect a specific humpback breeding and calving strategy. We expected to find a positive effect on group density of distance from shore and spatial autocorrelation, and a negative influence of bathymetry and wind speed.

## METHODS

### *Study Area*

The study area comprised the southwestern Atlantic continental margin (continental shelf and slope) from the state of Sergipe ( $10^{\circ}08'S$ ) to the state of Rio de Janeiro ( $23^{\circ}12'S$ ), Brazil, covering the 500 m isobath and reaching the 1,000 m isobath in some areas (Fig. 1).

The northeast and east continental shelves are characterized by a complex substrate derived from volcanic and tectonic activities. Between  $10^{\circ}08'S$  and  $15^{\circ}45'S$  the shelf gradually narrows with a mean width of 50 km (Zembruski *et al.* 1979, Lessa and Cirano 2006). A shelf expansion, originated from volcanos and biogenic growth, occurs at the Royal Charlotte and Abrolhos banks, reaching 110 and 200 km of width respectively (Zembruski *et al.* 1979, Fainstein and Summerhayes 1982). The Abrolhos Bank is considered the largest and richest coral reef area of the South Atlantic Ocean (Leão 1999, Castro and Pires 2001).

The continental slope related to the northeast and east shelves has a mean width of 30 km and a mean depth of 2,000 m. From  $20^{\circ}19'S$  on the shelf narrows again until  $22^{\circ}00'S$ , which is considered the southern limit of the east shelf (Zembruski *et al.* 1979). In this area the slope is wider and less abrupt than the northern counterpart. From this limit the shoreline orientation changes direction from northeast-southwest to east-west, and the shelf is known as the southeast continental shelf. It has a mean width of 80 km and smooth topography with shelf break depth varying between 120 and 180 m (Castro and Miranda 1998).

### *Survey*

The sighting survey took place off the Brazilian coast between 26 August and 13 September 2011 during the peak of abundance of the humpback whale breeding season (Martins *et al.* 2001, Morete *et al.* 2003). The sampling design was planned according to distance sampling protocol (Buckland *et al.* 2001) to estimate population size of the SWA humpback whale population (Pavanato *et al.* 2017). Distance data were not used here, however (see *Group Density Prediction* below for details).

A high-wing aircraft (Aerocommander) equipped with bubble windows was used to survey the northeast, east, and southeast continental shelves and the corresponding continental slopes. Whenever possible, parallel transects were designed. In order

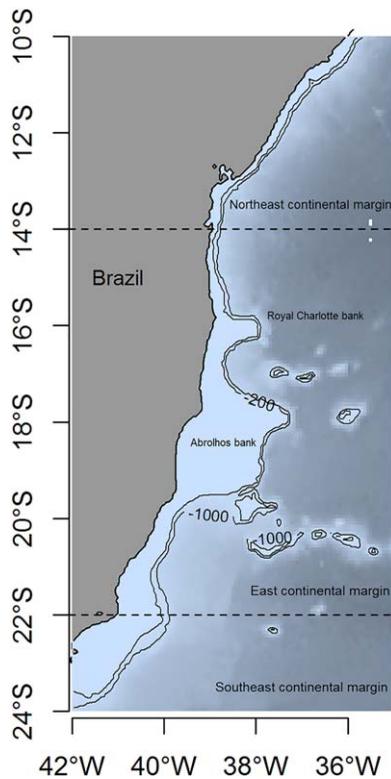


Figure 1. Map showing the study area. The isobaths of 200 m and 1,000 m are shown (black lines) and the limits of the continental margins are indicated (thin black lines).

to maximize sampling effort in the narrow northeast margin, transects were designed in a zigzag shape. The aircraft flew during favorable weather conditions at a height of 500 ft (152.40 m) and airspeed of 110 knots (204 km/h).

The survey was conducted with four researchers on board. Two observers were located at the right and left bubble windows of the aircraft with one data recorder and one resting person both seated in the rear row. To avoid fatigue effect, positions were rotated according to a fixed protocol before starting a new transect. For each sighting, the declination angle from the aircraft to the whale or group of whales was measured by a hand-held clinometer when passed abeam. Geographic position and group size were also registered. Group was defined as two or more individuals swimming side by side within up to two body lengths of their nearest neighbor, with coordinated speed and direction of movements (Clapham 2000).

#### *Environmental Data Acquisition and Processing*

We explored grids built at two spatial scales, a coarse resolution of  $0.5^\circ$ , and a fine resolution of  $0.1^\circ$  of latitude and longitude, which correspond to 56 km and 11 km at the equator respectively. After implementing all analytical procedures described below, we did not recognize substantial scale dependence on the predictor variable; therefore, we opted to carry out inference and prediction using the grid

cell of  $0.1^\circ$  of latitude and longitude because this resolution afforded a more detailed inference (see details in Appendix S1, Fig. S5–S13).

Effort was calculated as the total transect lengths (km) within a given grid cell. Bathymetric data were obtained from the global database ETOPO-1 with 1 arc-minute resolution (Amante and Eakins 2009). To match the data with the scale of the grid cells, we averaged bathymetry to obtain a single value by spatial resolution. Distances from shore (km) were calculated with the Haversine approach as the great-circle distance between the midpoint of grid cells for the different scales and shoreline coordinates (Hijmans 2017). Wind speed data were obtained from daily surface observations derived from an advanced scatterometer (ASCAT) aboard the Meteorological Operational Polar (MetOp-A) satellite at a spatial resolution of  $0.25^\circ$  of latitude and longitude. We used the nearest neighbor interpolation to regrid wind speed for the spatial resolution of  $0.1^\circ$ , and the values were averaged for the spatial resolution of  $0.5^\circ$ . Group sizes were averaged for the two spatial resolutions. Covariate maps are provided in Figures S14–S16.

### *Group Density Prediction*

Group density prediction for the SWA humpback whale population was based on conditional autoregressive (CAR) models that allow the modeling of autocorrelated spatial effects under the response variable (Besag 1974). Regarding humpback whales, the choice of CAR models seems reasonable for at least two reasons: it accounts for the influences of environmental variables and inserts a component describing the spatial correlation between whale groups. Given by a generalization of the standard conditional autoregressive models, the intrinsic CAR models (ICAR) support certain types of nonstationarity in which the variance-covariance matrix is not positive definite, but is semidefinite (Besag 1974, Held and Rue 2010). For these reasons, the ICAR models were chosen here.

Let  $s_j$  be the geographic location of a sighted group  $y$ . Summing up the number  $y(s_j)$  of sighted groups within the grid cell  $i$ , we defined

$$Y_i = \sum_{j \in i} y(s_j), \text{ for } i = 1, \dots, G, \quad (1)$$

where  $G$  is the number of grid cells.

Assuming these counts depend on the environment—mean bathymetry ( $b$ ), distance from shore ( $d$ ), and wind speed ( $w$ ) (all numeric variables), then a Poisson distribution results in

$$Y_i \sim \text{Pois}(\lambda_i) \quad (2)$$

$$\log(\lambda_i) = \beta_0 + \sum_{z=1}^Z \beta_{z,z_i} + \rho_i + \log(f_i), \text{ for } z = 1, \dots, Z, \quad (3)$$

where  $\lambda_i$  is the expectation,  $Z$  is the total number of covariates in a model, and  $f$  is the effort in km. The term  $\log(f_i)$  is an offset. Quadratic terms for  $b$  and  $d$  were also included. Since the variance of counting data was higher than the mean, a negative binomial error distribution was also tested.

The spatial dependence between groups can be achieved by adding a random structure defined as

$$\rho_i | \rho_j \sim \text{Normal} \left( \frac{\sum_{j \in \delta_i} \alpha_{ij}}{\alpha_{i+}}, \frac{\sigma_\rho^2}{\alpha_{i+}} \right), \quad (4)$$

where  $\alpha_{i+}$  denotes the total number of neighboring cells of  $i$  and  $\alpha_{ij} = 1$  if cells  $i$  and  $j$  share the same boundary, otherwise  $\alpha_{ij} = 0$ . The conditional variance  $\alpha_\rho^2$  is a hyperparameter. The component  $\rho_i$  is conditional upon the effects of the neighborhood through an autoregressive distribution. The neighborhood effect is defined as the increasing density near high-density cells and decreasing density near low-density cells or absences (Dormann 2009).

We used eight adjacent cells as neighbors (also known as Queen neighborhood). For cells that share borders with cells containing the continental (west) shoreline the number of neighbors was less than eight; for cells that share their boundary with nonsampled cells we built the grid to also achieve eight neighbors, to avoid border effects (Lim *et al.* 2007).

Therefore, in a first level only the cells referring to the observations were modeled. In a second level, the spatial correlation was estimated even for those cells with null effort. This structure of spatial dependence allows investigating the contribution of random effects even in cells that were not sampled, settling gaps and irregular sampling intensities (Gelfand *et al.* 2006).

For mapping purposes, a density index ( $\hat{D}_i$ ) was calculated, representing the fraction of the total number of groups of the SWA humpback whale population per cell. By defining the abundance of groups  $\hat{N}_g = \hat{N}/g$ , where  $\hat{N}$  is the population size (posterior mean) estimated elsewhere (Pavanato *et al.* 2017), and  $g$  is the mean group size,  $\hat{D}_i$  was defined as

$$\hat{D}_i = \frac{\lambda_i}{G} \hat{N}_g, \quad (5)$$

$$a_i \sum_{i=1}^G \lambda_i$$

where  $a_i$  is the area in  $\text{km}^2$ , and  $\lambda_i$  is the expected number of groups from the selected ICAR model for each cell  $i$ .

Since we assumed a Poisson or a negative binomial distribution for  $Y_i$ , we calculated the occupancy probability (*i.e.*, the occurrence probability of at least one group by cell) as

$$\Pr [Y_i > 0] = 1 - e^{-\hat{D}_i a_i}. \quad (6)$$

It is noteworthy to mention that for grid cells in which  $f_i = 0$  the density and occurrence probability were predicted, and for this reason, those cells might have higher associated variances.

#### *Parameter Estimation and Model Selection*

Unknown parameters of ICAR models were estimated through Bayesian inference *via* integrated nested Laplace approximation (INLA, Rue *et al.* 2009). Default vague priors were established for the ICAR model through the *besag* function in R-INLA package (Rue *et al.* 2009, Lindgren and Rue 2015) available in R software (R Core Team 2015).

The steps to select covariates with important effects over group density were the following: (1) avoid collinearity effects by calculating the Spearman correlation coefficients ( $r$ ) among all pairs of covariates and do not use jointly within a single model whenever  $r > 0.6$ ; (2) given this constraint, adjust models from the null to the global; (3) use the Watanabe-Akaike information criterion (WAIC, Watanabe 2010) for model selection, where the smallest WAIC indicates a better model whenever the difference to the nearest alternative model exceeds one; (4) whenever the WAIC from one model to the subsequent one decreased or  $\Delta\text{WAIC} \leq 1$ , do not maintain the new covariate; (5) further evaluate quality of model fit with the conditional predictive ordinates (CPO) and the probability integral transforms (PIT, Martino and Rue 2010).

The WAIC is given by:

$$\text{WAIC} = -2 \sum_{i=1}^G \log p_{\text{post}} y_i + 2p_{\text{WAIC}} \quad (7)$$

where  $p_{\text{post}} y_i = E_{\text{post}} p(y_i | \theta)$  is the posterior predictive density for each data point  $y_i$  over the entire posterior parameter space  $\theta$ , and  $p_{\text{WAIC}}$  is the effective number of parameters defined as the summation over all data points of the posterior variance for the log predictive density for  $y_i$ . That is,

$$p_{\text{WAIC}} = \sum_{i=1}^G \log p(y_i | \theta) \quad (8)$$

Differently from Akaike information criteria (AIC) and the deviance information criteria (DIC), which are conditioned on some point estimate (maximum likelihood estimate for AIC and posterior mean for DIC), WAIC averages over the posterior distribution.

The use of CPO is suited to measuring predictive power and for detecting outliers in a data set. Formally,  $\text{CPO} = p(y_i | y_{[-i]})$ , for  $I = 1, \dots, G$ , is the predictive cross-validated (or “leave-one-out”) predictive density at  $y_i$ , where the posterior distribution was based on  $y_{[-i]}$ , defined as the vector of all observations except  $i$ . If  $y_i$  is an outlier,  $\text{CPO}_i$  will be small meaning that the measure of probability of distance between data and model is large. Visual examination of individual  $\text{CPO}_i$  helps to identify potential problems in the model specification. To use the CPOs jointly as a means to evaluate predictive power of a model, the averaged  $-\log \text{CPO}(\overline{\text{LCPO}})$  is indicated (Held and Rue 2010, Roos and Held 2011).

$$\overline{\text{LCPO}} = -(n^{-1}) \sum_{i=1}^G \log(\text{CPO}_i) \quad (9)$$

Likewise for WAIC, a smaller value of  $\overline{\text{LCPO}}$  indicates a better model.

The PIT histogram is useful to compare and validate candidate models, wherein in well-calibrated models PIT values should have a uniform distribution (Martino and Rue 2010, Blangiardo and Cameletti 2015).

$$\text{PIT}_i = \Pr [y_i^{\text{new}} < y_i | y_{[-1]}] \quad (10)$$

In the case of count data, the predictive distribution is discrete and the PIT histogram is not uniform under the hypothesis of an ideal forecast (Czado *et al.* 2009). Therefore, an adjusted nonrandomized version of PIT can be computed (Czado *et al.* 2009):

$$\text{PIT}_i^{\text{adj}} = \text{PIT}_i + 0.5 \Pr [y_i^{\text{new}} < y_i | y_{[-1]}] \quad (11)$$

The adjusted PIT values can be interpreted in the same way as in applications with continuous outcome data.

The importance of the spatial autocorrelation component was evaluated by comparing the best-fit ICAR model with a similar generalized linear model (GLM) through WAIC. To compare the covariate (fixed) and spatial (random) effects, we mapped the mean  $\rho_i$  and the mean linear predictors from the best-fit model. Finally, we calculated the mean and standard deviation of the fixed and random effects by continental margin to infer the main force driving aggregation of humpback whale groups in each one.

## RESULTS

We obtained 488 cells within the grid of  $0.1^\circ$  of latitude and longitude for the study area. We augmented the number of cells to achieve eight neighbors and, therefore, we worked over 1,148 cells. The number of groups and effort are represented in Figure 2.

The global model included all covariates since  $r < 0.6$  for all covariate pairs. Based on the WAIC and  $\overline{\text{LCPO}}$ , Poisson models provided a better fit to the data than negative binomial models (Table 1).

According to the WAIC and  $\overline{\text{LCPO}}$ , model E was the best ( $d + d^2 + b$ ) (Table 1). Analyzing the PIT histograms from the Poisson (Fig. S1), all models seem to be reasonably well calibrated as no extreme values were computed.

According to model E, distance from shore ( $d$ ) had a quadratic effect on group density, indicating expected maximum density at a distance from shore of 171 km (Fig. 3). Bathymetry ( $b$ ) had a negative effect on group density (Fig. 3). It is important to note that the maximum density in logarithmic scale for distance from shore has a large uncertainty. Therefore, distances from shore from 140 to 236 km are within the possible range of maximum density.

The WAIC of the ordinary GLM that is analogous to model E but without incorporating the spatial structure was 1,237.84, showing a considerable predictive decay

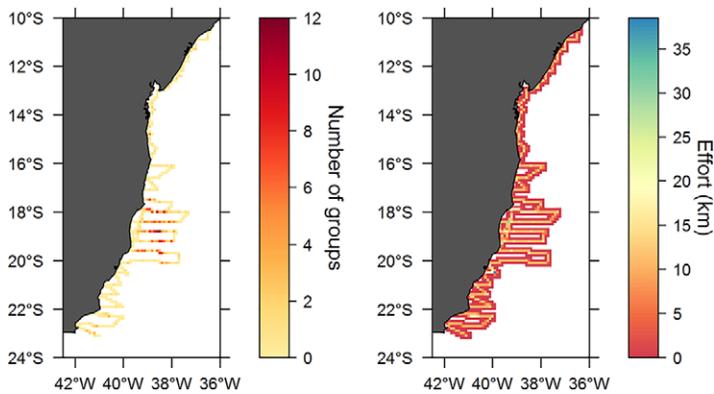


Figure 2. Number of groups and effort (km) by grid cell.

Table 1. Summary of candidate models including number of parameters ( $p_{\text{WAIC}}$ ), Watanabe-Akaike information criterion (WAIC) and logarithmic conditional predictive ordinates (LCPO).

Model <sup>a</sup>	$p_{\text{WAIC}}$	WAIC	LCPO
Poisson			
A: null	82.70	883.63	458.13
B: $d$	82.38	880.77	457.20
C: $d + d^2$	76.45	877.29	451.79
D: $d + d^2 + w$	75.83	877.08	450.79
E: $d + d^2 + b$	73.63	868.16	445.93
F: $d + d^2 + b + b^2$	73.24	868.20	445.77
Negative binomial			
G: null	62.41	919.47	478.99
H: $d$	62.53	916.59	483.66
I: $d + d^2$	53.12	910.26	463.45
J: $d + d^2 + w$	51.99	909.80	460.87
K: $d + d^2 + b$	51.14	899.69	457.44
L: $d + d^2 + b + b^2$	51.11	899.26	587.70

<sup>a</sup> $d$  = distance from shore,  $w$  = wind speed,  $b$  = bathymetry.

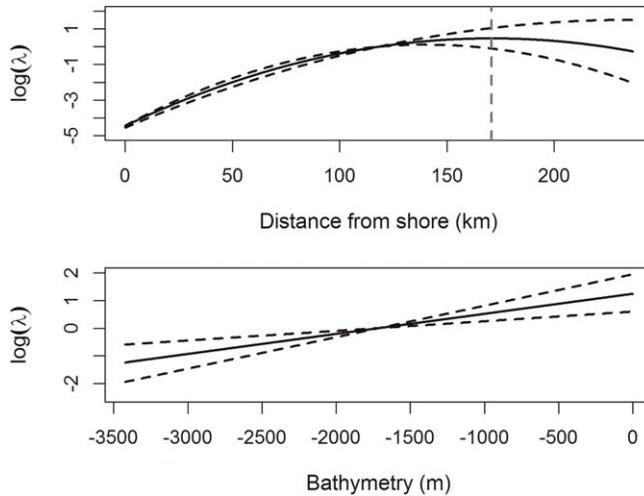


Figure 3. Covariate effects on group density in logarithmic scale. Continuous black lines indicate the posterior means, black dashed lines indicate the 95% probability intervals; gray dashed line indicates the optimum (*i.e.*, the highest density) for the covariate with quadratic term (upper panel).

compared to the ICAR model ( $\Delta\text{WAIC} = 355.53$ ). The variance explained by the spatial structure in the ICAR model was of 22%.

The density predicted (posterior mean) by model E is shown in Figure 4 along with the count data. There is good agreement between the predicted and observed number of groups in those grid cells where comparison was possible. Large variations between grid cells indicate aggregations in some areas, most marked between grid cells 400 and 900 (Fig. S2). These correspond to the east continental margin (Fig. 1).

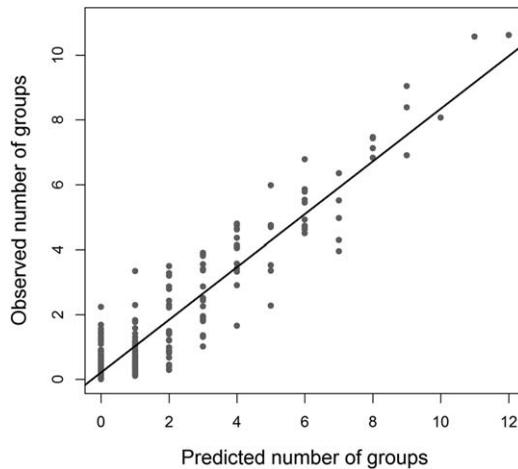


Figure 4. Observed and predicted (posterior mean) number of groups by grid cell.

The map of predicted group density per km<sup>2</sup> displays the highest aggregations at the Abrolhos Bank, in some cells of the east and northeast continental margin, and in the southernmost covered area in the state of Rio de Janeiro (Fig. 5). In general, the smallest densities were estimated for cells located too close or too far from the shoreline. Minimum and maximum predicted densities were 0 and 0.5 per km<sup>2</sup>. To provide the plausibility of the candidate models, density prediction analogous to Figure 5 are available for models A (null) and F (global), which would be the two extreme outcomes (Fig. S3, S4).

The occupancy probability of groups per grid cell was heterogeneous, ranging from 0 to 1, with a mean of 0.7 (Fig. 6). A remarkable gap in the occurrence (*i.e.*, cells in which the probability of occurrence were low) was identified for the area located between the south of Abrolhos Bank and northern limit of the southeast continental margin, between 20°S and 22°S. Regarding longitude, cells located extremely far from the shoreline (>200 km) and/or at deep waters (>1,500 m) also had the smallest probability of occurrences.

The estimates of the covariate and spatial autocorrelation effects are shown in Figure 7, while the relationship between them is shown in Figure 8. The covariate effect was high for the majority of cells, with the highest prediction in the Abrolhos Bank and in the southernmost area far from the shoreline; a small covariate effect was predicted for cells far from the coast in the latitude of the Abrolhos Bank and for some cells between the south of Abrolhos Bank and northern limit of the southeast continental margin. The spatial effect was high in the northeast continental margin, except for the northernmost cells, and in the Abrolhos Bank. The smallest spatial autocorrelation effect was predicted at the south of the Abrolhos Bank and in the Royal Charlotte Bank. A positive relation between both effects seems to occur (Fig. 8).

A comparison between the covariate and spatial autocorrelation effects by continental margin (Table 2) suggests that the autocorrelation was mostly important in the northeast, and east margins than in the southeast. By comparison, the covariate effect was smaller in the northeast and southeast margins than in the east

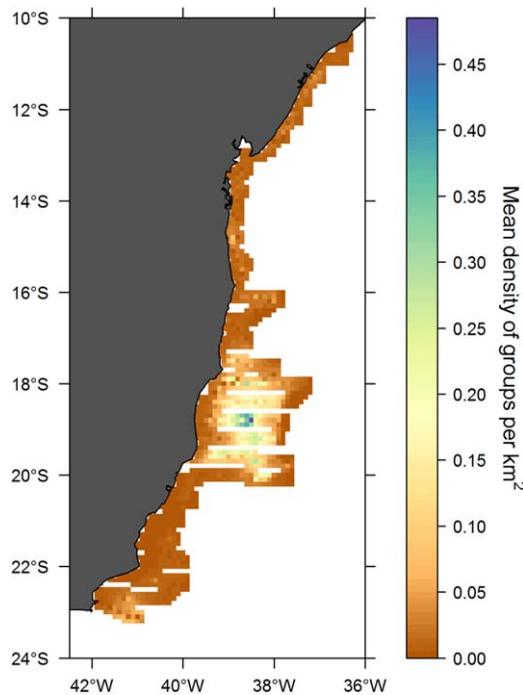


Figure 5. Density of groups (posterior mean) by  $\text{km}^2$  obtained from model E ( $d + d^2 + b$ ).

continental margin. Marginal posterior standard deviations, however, were relatively large in all cases, indicating large uncertainties in the reported differences.

## DISCUSSION

### *Modeling Approach and Inference*

The predictive distribution of humpback whale group density and occupancy probability off the Brazilian coast was based on models which combined components with potential to reflect diverse types of data variability, including environmental effects (distance from shore, wind speed, and bathymetry), spatial autocorrelation, and heterogeneous sampling intensity. From an investigative point of view, the inclusion of one or more of these components may be satisfactory. Nevertheless, predictive inference may be misleading if group density or occupancy processes need more explanation than that achieved by environmental covariates (Guisan and Zimmermann 2000, Guisan *et al.* 2006, Latimer *et al.* 2006).

Substantial residual autocorrelation can remain after a model is fitted to data due to incomplete explanatory power of the linear predictors that are available. To handle the residual autocorrelation, a suitable solution is obtained by the inclusion of a set of spatially correlated random effects in the form of a Bayesian hierarchical model (Lee 2013). The spatial correlation pattern implies that close cells have similar behavior compared to those located farther away. In other words, a high-density

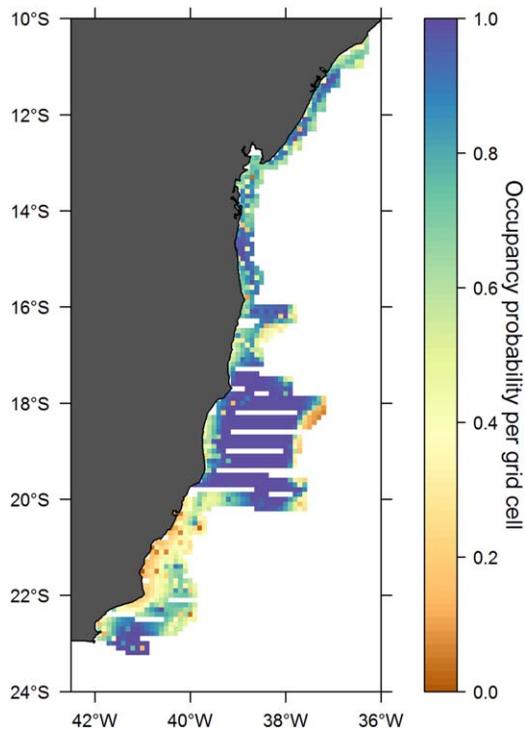


Figure 6. Probability of occurrence (posterior mean) by grid cell obtained from model E ( $d + d^2 + b$ ).

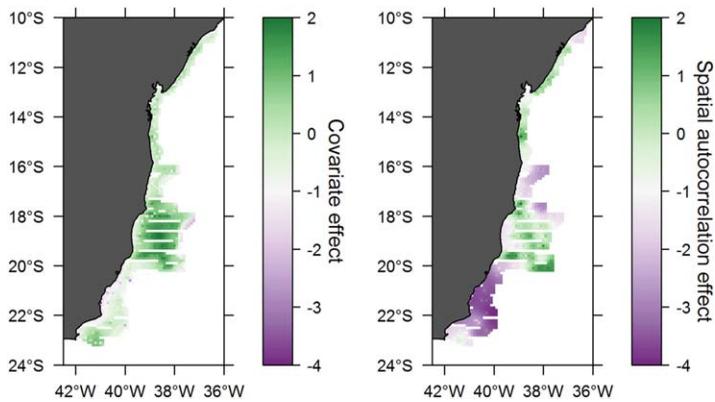


Figure 7. Posterior mean of the covariate effect (fixed) and spatial autocorrelation effect (random) in logarithmic scale by grid cell obtained from model E ( $d + d^2 + b$ ).

cell may yield an increment in near cells, while a low-density cell may induce a decrease in its neighborhood (Dormann 2009).

Conditional autoregressive models have already been used to make predictions in epidemiology, presence/absence data of vegetation, terrestrial animals, and birds

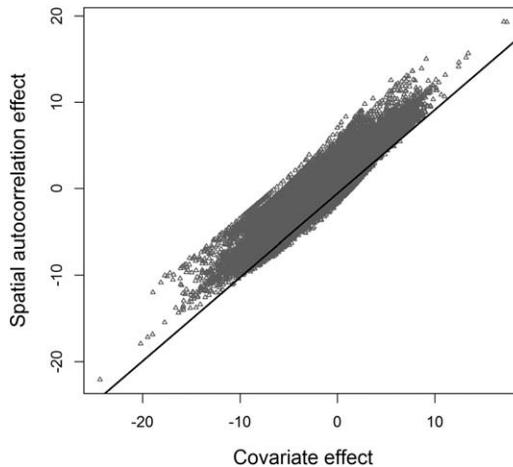


Figure 8. Relationship between the posterior distributions of the spatial autocorrelation and covariate effect obtained from model E ( $d + d^2 + b$ ). The black line indicates a linear model fitted between both effects.

Table 2. Summary of the spatial and covariate effects in the northeast, east, and southeast continental margins.

Continental margin	Mean(SE)	SD(SE)	Mean(CE)	SD(CE)
Northeast	1.81	1.69	0.20	1.99
East	1.10	3.10	0.38	5.09
Southeast	0.41	2.06	0.21	2.94

Note: SE corresponds to spatial effect and CE to covariate effect.

(Lichstein *et al.* 2002, Latimer *et al.* 2006, Carroll *et al.* 2010). Redfern *et al.* (2006) have suggested that CAR models constitute a potential technique for cetacean modeling. Here, cetacean counting data were modeled using a Gaussian autoregressive conditional structure subjected to the influence of eight (first order) neighbors to account for the spatial effects in each grid cell. This structure may be further adjusted to account for the influence of second- and higher-order neighbors through a decay function, assigning smooth weights as the distances increase.

The INLA procedure has been largely used in spatial data analysis for providing an interesting computational approach to obtain marginal posterior distributions, being a much faster alternative to Markov chain Monte Carlo methods (MCMC, Martino and Rue 2010, Beguin *et al.* 2012, Blangiardo *et al.* 2013, Bivand *et al.* 2015). In contrast to INLA, the MCMC algorithms, when applied to complex hierarchical spatial models, usually have difficulties related to convergence and computational time, which may be a disadvantage to those ecologists who aspire to use Bayesian inference in a spatial context (Beguin *et al.* 2012). Nevertheless, Bayesian hierarchical models still lack robust methods for model comparison since AIC and the Bayesian information criterion (BIC) are not recommended for complex hierarchical structures because the true number of independent parameters is hard to estimate in these models (Burnham and Anderson 2002). The DIC (Spiegelhalter

*et al.* 2002) was developed to better suit this challenge, but lack of robustness seems to remain (Royle and Dorazio 2008, Spiegelhalter *et al.* 2014). The WAIC and can be viewed as improvements on the DIC (Vehtari *et al.* 2017), and for this reason were chosen here. For a review of these criteria see Gelman *et al.* (2014).

### *Habitat Selection and Density Pattern*

The difference between the effects of spatial autocorrelation and covariates (distance from shore, and bathymetry) were related to the type of continental margins: for the northeast margin the spatial autocorrelation effect was predominant; for the southeast margin the covariate effect was predominant; and for the east continental margin both effects were important (Table 2). It is worthy to point out that high mean covariate and spatial autocorrelation effects were predicted for the east margin, an area in which the highest density was estimated (Fig. 5). In contrast, in areas where one effect was predominant (southeast and northeast margin) density tended to be smaller, except for some high-density cells.

The best-fit model included distance from shore and its quadratic term, and bathymetry. Likewise, associations between shallow waters at a given distance from the shoreline and the occurrence of groups, mainly mother-calf pairs, were found for other breeding grounds (Ersts and Rosenbaum 2003, Félix and Haase 2005, Oña *et al.* 2016). Specifically for the Abrolhos Bank, a high proportion of mother-calf pairs has been found in shallow waters near the Abrolhos Archipelago (Martins *et al.* 2001; Morete *et al.* 2003, 2007). From a broad scale, our results point towards high group densities in shallow waters at a distance from the coast of 171 km (95% PI: 140–236 km), which encompass the Abrolhos Bank.

Nonindependence of data points near each other yield spatial autocorrelation, which may be connected through ecological and nonecological processes (Dormann 2009). Nonmodeled covariates and a misleading functional relationship (*e.g.*, consider a linear effect when the “true” relationship between the response and explanatory variable is nonlinear) constitute features driving spatial autocorrelation (Dormann *et al.* 2007). For instance, water transparency might be a missing effect that leads to an inflation of the spatial autocorrelation. In addition, because we did not distinguish between density and detection processes, the effect of covariates may be responding to both. It is unlikely that distance from shore and bathymetry had an effect on detection probability; on the other hand, wind speed had the potential to affect the detection process. From our best-fit model, wind speed was not selected. Otherwise, caution should be taken to interpret wind speed effect.

Another possibility to explain the importance of the spatial autocorrelation is that a synergistic effect has been at play because biological processes may generate spatial patterns, leading to a spatial structuring of the population (Latimer *et al.* 2006, Dormann *et al.* 2007). Since territorial defense of feeding sources is absent and predation is probably negligible, the population spatial structure may be viewed as an outcome of the adopted breeding strategy which, for humpback whales, is controlled by a polygynous system in which males search for a relatively dominant position over females (Winn and Reichley 1985). Singer males may remain alone, probably to attract females for copulation and to maintain distance from other males, or may actively look out for and, eventually, escort a mature female, aggregating in pairs or triplets when a female is accompanied by a calf (Winn and Reichley 1985). Furthermore, males may constitute coalitions to compete for the main escort position (Winn and Reichley 1985), when they might

show aggressive behavior (Tyack and Whitehead 1983). Males that may not compete successfully might adopt a dispersal strategy (Rosenbaum *et al.* 2009) to low-density areas with fewer partners, but also with less competition (Clapham 2001). Thereby, an aggregated structure at the breeding ground would favor the selection for both sexes and may be one of the mechanisms driving the spatial autocorrelation for groups off the Brazilian coast.

Moreover, the spatial autocorrelation effect may evidence another important aspect driving humpback whale habitat selection. Based on the temporal structure of the population, lactating females are the first to migrate to the breeding grounds, followed by mature and immature individuals and, finally, by pregnant females (Dawbin 1966). Mother and calf pairs are thought to gather in the best habitat conditions for calving and nursing. Since females could be in estrus at this time (Chittleborough 1965), they may attract males, and this suggests individuals are spatially distributed according to their role within the population. The aggregation of groups often happens to be around the habitat that is selected by females to maximize the success of calf growth.

For the SWA humpback whale population, the highest concentration of groups has been identified on the Abrolhos Bank (Andriolo *et al.* 2006, 2010). Here, the predicted density corroborates the previous studies; it also suggests that areas relatively far from the bank (*e.g.*, cells from northeast and southeast margins) comprised moderate densities ( $\sim 0.1$  groups per cell) and occurrence was almost certain (Figs. 5 and 6). Therefore, we recommend that future aerial or vessel surveys expand their sampling protocol to include the north and south limits of the present survey to prospect for new areas of occurrence.

There is a consensus that the SWA humpback whale population has been consistently increasing and expanding after whaling was prohibited (Siciliano 1995, Zerbini *et al.* 2004, Andriolo *et al.* 2010, Wedekin *et al.* 2010, Ward *et al.* 2011, Zerbini *et al.* 2011, Bortolotto *et al.* 2016, Pavanato *et al.* 2017). Nevertheless, the scenario of population growth and associated habitat expansion may increase negative conflicts with anthropogenic activities given the low representativeness of the marine protected areas along the Brazilian breeding ground (Wedekin *et al.* 2010, Martins *et al.* 2013, Castro *et al.* 2014). The anthropogenic threats for the SWA humpback whale population include collisions with vessels, whale watching disturbance, entanglement in fishing gear, cetacean stranding associated with seismic surveys, and chemical pollution (Morete *et al.* 2007, Engel *et al.* 2004, Martins *et al.* 2013, Bezamat *et al.* 2015).

From a precautionary standpoint and in the face of these possible negative interactions, we recommend that future spatial distribution models also include proxies of the presence and intensity of human activities as explanatory covariates to quantify their potential effects on the occurrence or density of humpback whales off the Brazilian coast. In addition, extending the analyses to include multiple years using spatio-temporal models would provide important information on how the population has been expanding. Besides being important for preventing conflicts with human activities, spatial models may guide efficient management and conservation actions, and, eventually, define representative marine protected areas.

#### ACKNOWLEDGMENTS

We thank the observers I. L. Serrano and M. C. C. Marcondes for their help during the aerial survey. We are grateful to A. N. Zerbini, E. R. Secchi, L. Dalla Rosa, M. Kéry, and

J. Redfern, the editors D. Boness and P. Hammond, Charlotte Boyd, and two anonymous reviewers for their valuable reviews of this manuscript. We are also grateful to A. Detoni for her help with data acquisition and E. Krainski for his help with INLA. We appreciated the detailed English revision provided by A. Lamont. Petrobras is the official sponsor of Projeto Baleia Jubarte/Instituto Baleia Jubarte. The aerial surveys were funded by Veracel Celulose S.A. The Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) provided a scholarship to the first author during her M.Sc. degree while working under the guidance of the last author.

#### LITERATURE CITED

- Acevedo-Gutiérrez, A. 2009. Habitat Use. Pages 524–529 in W. F. Perrin, B. Würsig and J. G. M. Thewissen, eds. *Encyclopedia of marine mammals*. Academic Press, San Diego, CA.
- Amante, C., and B. W. Eakins. 2009. ETOPO-1 1 arc-minute global relief model: Procedures, data sources and analysis. U.S. Department of Commerce, NOAA Technical Memorandum NESDIS NGDC-24. 19 pp.
- Andriolo, A., C. C. A. Martins, M. H. Engel, *et al.* 2006. The first aerial survey to estimate abundance of humpback whales (*Megaptera novaeangliae*) in the breeding ground off Brazil (Breeding Stock A). *Journal of Cetacean Research and Management* 8:307–311.
- Andriolo, A., P. G. Kinas, M. H. Engel, C. C. A. Martins and A. M. Rufino. 2010. Humpback whales within the Brazilian breeding ground: Distribution and population size estimate. *Endangered Species Research* 11:233–243.
- Bailey, H., and P. Thompson. 2010. Effect of oceanographic features on fine-scale foraging movements of bottlenose dolphins. *Marine Ecology Progress Series* 418:223–233.
- Becker, E. A., D. G. Foley, K. A. Forney, J. Barlow, J. V. Redfern and C. L. Gentemann. 2012. Forecasting cetacean abundance patterns to enhance management decisions. *Endangered Species Research* 16:97–112.
- Begon, M., C. R. Townsend and J. L. Harper. 2006. *Ecology from individuals to ecosystems*. Blackwell Publishing, Oxford, U.K.
- Beguín, J., S. Martino, H. Rue and S. G. Cumming. 2012. Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution* 3:921–929.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 36:192–236.
- Bezamat, C., L. L. Wedekin and P. C. Simões-Lopes. 2015. Potential ship strikes and density of humpback whales in the Abrolhos Bank breeding ground, Brazil. *Aquatic Conservation: Marine and Freshwater Ecosystems* 25:712–725.
- Bivand R., V. Gómez-Rubio and H. Rue. 2015. Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software* 63:1–31.
- Blangiardo, M., and M. Cameletti. 2015. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, Chichester, U.K.
- Blangiardo, M., M. Cameletti, G. Baio and H. Rue. 2013. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology* 7:39–55.
- Bled, F., J. A. Royle and E. Cam. 2011. Hierarchical modeling of an invasive spread: The Eurasian collared-dove *Streptopelia decaocto* in the United States. *Ecological Applications* 21:290–302.
- Bortolotto, G. A., D. Danilewicz, A. Andriolo, E. R. Secchi and A. N. Zerbini. 2016. Whale, whale, everywhere: Increasing abundance of western South Atlantic humpback whales (*Megaptera novaeangliae*) in their wintering grounds. *PLOS ONE* 11(10):e0164596.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas. 2001. *Introduction to distance sampling*. Oxford University Press, Oxford, U.K.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd edition. Springer, New York, NY.

- Cañadas, A., and P. S. Hammond. 2008. Abundance and habitat preferences of the short-beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: Implications for conservation. *Endangered Species Research* 4:309–331.
- Carroll, C., D. S. Johnson, J. R. Dunk and W. J. Zielinski. 2010. Hierarchical Bayesian spatial models for multispecies conservation planning and monitoring. *Conservation Biology* 24:1538–1548.
- Castro, B. M., and L. B. Miranda. 1998. Physical oceanography of the western Atlantic continental shelf located between 4°N and 34°S. Pages 209–251 in A. R. Robinson and K. H. Brink, eds. *The sea*. John Wiley and Sons, New York, NY.
- Castro, C. B., and D. O. Pires. 2001. Brazilian coral reefs: What we already know and what is still missing. *Bulletin of Marine Science* 69:357–371.
- Castro, F. R., N. Mamede, D. Danilewicz, Y. Geyer, J. L. A. Pizzorno, A. N. Zerbini and A. Andriolo. 2014. Are marine protected areas and priority areas for conservation representative of humpback whale breeding habitats in the western South Atlantic? *Biological Conservation* 179:106–114.
- Chittleborough, R. G. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). *Australian Journal of Marine and Freshwater Research* 16:33–128.
- Clapham, P. J. 2000. The humpback whale: Seasonal feeding and breeding in a baleen whale. Pages 173–196 in J. Mann, R. C. Connor, P. L. Tyack and H. Whitehead, eds. *Cetacean societies: Field studies of dolphins and whales*. University of Chicago Press, Chicago, IL.
- Clapham, P. J. 2001. Why do baleen whales migrate? A response to Corkeron and Connor. *Marine Mammal Science* 17:432–436.
- Corkeron, P. J., and R. C. Connor. 1999. Why do baleen whales migrate? *Marine Mammal Science* 15:1228–1245.
- Czado, C., T. Gneiting and L. Held. 2009. Predictive model assessment for count data. *Biometrics* 65:1254–1261.
- Dalla Rosa, L., J. K. B. Ford and A. W. Trites. 2012. Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Continental Shelf Research* 36:89–104.
- Dawbin, W. H. 1956. The migrations of humpback whales which pass the New Zealand coast. *Transactions and Proceedings of the Royal Society of New Zealand* 84:147–196.
- Dawbin, W. H. 1966. The seasonal migratory cycle of humpback whales. Pages 145–169 in K. S. Norris, ed. *Whales, dolphins and porpoises*. University of California Press, Berkeley, CA.
- Doniol-Valcroze, T., D. Berteaux, P. Larouche and R. Sears. 2007. Influence of thermal fronts on habitat selection by four rorqual whale species in the Gulf of St. Lawrence. *Marine Ecology Progress Series* 335:207–216.
- Dormann, C. F. 2009. Response to comment on “Methods to account for spatial autocorrelation in the analysis of the species distributional data: A review.” *Ecography* 32:379–381.
- Dormann, C. F., M. J. McPherson, M. B. Araújo, *et al.* 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* 30:609–628.
- Elwen, S. H., and P. B. Best. 2004. Environmental factors influencing the distribution of southern right whales (*Eubalaena australis*) on the south coast of South Africa I: Broad scale patterns. *Marine Mammal Science* 20:567–582.
- Engel, M. H., and A. R. Martin. 2009. Feeding grounds of the western South Atlantic humpback whale population. *Marine Mammal Science* 25:964–969.
- Engel, M. H., M. C. C. Marcondes, C. C. A. Martins, F. O. Luna, R. P. Lima and A. Campos. 2004. Are seismic surveys responsible for cetacean strandings? An unusual mortality of adult humpback whales in Abrolhos Bank, northeastern coast of Brazil. Unpublished report to the Scientific Committee of the International Whaling Commission SC/56 E.

- Engel, M. H., N. J. R. Fagundes, H. C. Rosenbaum, *et al.* 2008. Mitochondrial DNA diversity of the southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antarctic feeding areas. *Conservation Genetics* 9:1253–1262.
- Ersts, P. J., and H. C. Rosenbaum. 2003. Habitat preferences reflect social organization of humpback whales (*Megaptera novaeangliae*) on a wintering ground. *Journal of Zoology* 260:337–345.
- Estes, J. A., D. P. DeMaster, D. F. Doak, T. M. Williams and R. L. Brownell, Jr. 2006. Whales, whaling, and ocean ecosystems. University of California Press, Berkeley, CA.
- Fainstein, R., and C. P. Summerhayes. 1982. Structure and origin of marginal banks off Eastern Brazil. *Marine Geology* 46:199–215.
- Félix, F., and B. Haase. 2005. Distribution of humpback whales along the coast of Ecuador and management implications. *Journal of Cetacean Research and Management* 7:21–31.
- Ford, J. K. B., and R. R. Reeves. 2008. Fight or flight: Antipredator strategies of baleen whales. *Mammal Review* 38:50–86.
- Forney, K. A. 2000. Environmental models of cetacean abundance: Reducing uncertainty in population trends. *Conservation Biology* 14:1271–1286.
- Garshelis, D. L. 2000. Delusions in habitat evaluation: Measuring use, selection, and importance. Pages 111–164 in L. Boitani and T. K. Fuller, eds. *Research techniques in animal ecology*. Columbia University Press, New York, NY.
- Gelfand, A. E., J. A. Silander, S. Wu, A. Latimer, P. O. Lewis, A. G. Rebelo and M. Holder. 2006. Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis* 1:41–92.
- Gelman, A., J. Hwang and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24:997–1016.
- Gregg, E. J., and A. W. Trites. 2008. A novel presence-only validation technique for improved Steller sea lion *Eumetopias jubatus* critical habitat descriptions. *Marine Ecology Progress Series* 365:247–261.
- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.
- Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. M. C. C. Overton, R. Aspinall and T. Hastie. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* 43:386–392.
- Held, L., and H. Rue. 2010. Conditional and intrinsic autoregressions. Pages 207–208 in A. E. Gelfand, P. Diggle, P. Guttorp and M. Fuentes, eds. *Handbook of spatial statistics*. Chapman & Hall/CRC Press, Boca Raton, FL.
- Hijmans, R. J. 2017. geosphere: Spherical trigonometry. R package version 1.5–7. Available at <https://CRAN.R-project.org/package=geosphere>.
- IWC (International Whaling Commission). 1998. Report of the Scientific Committee. Report of the Meeting of the International Whaling Commission 48:53–118.
- Keller, C. A., W. B. Brooks, C. K. Slay, C. R. Taylor and B. J. Zoodsma. 2006. North Atlantic right whale distribution in relation to sea-surface temperature in the Southeastern United States calving grounds. *Marine Mammal Science* 22:426–445.
- Laidre, K. L., and M. P. Heide-Jørgensen. 2011. Life in the lead: Extreme densities of narwhals *Monodon monoceros* in the offshore pack ice. *Marine Ecology Progress Series* 423: 269–278.
- Latimer, A. M., W. Shanshan, A. E. Gelfand and J. A. Silander, Jr. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16:33–50.
- Leão, Z. M. A. N. 1999. Abrolhos—O complexo recifal mais extenso do oceano Atlântico Sul [Abrolhos—The largest reef complex in the South Atlantic Ocean]. *Sítios Geológicos e Palaontológicos do Brasil* 90:1–26.
- Lee, D. 2013. CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software* 55:1–24.

- Lessa, G. C., and M. Cirano. 2006. On the circulation of a coastal channel within the Abrolhos coral-reef system Southern Bahia (17°40'S), Brazil. *Journal of Coastal Research* 39:450–453.
- Lichstein, J. W., T. R. Simons, S. A. Shiner and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Lim, J., X. Wang and M. Sherman. 2007. An adjustment for edge effects using an augmented neighborhood model in the spatial auto-logistic model. *Computational Statistics & Data Analysis* 51:3679–3688.
- Lindgren, F., and H. Rue. 2015. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* 63:1–25.
- Mackintosh, N. A. 1965. *The stocks of whales*. Fishing News Books, London, U.K.
- Martino, S., and H. Rue. 2010. Case studies in Bayesian computation using INLA. Pages 99–114 in P. Mantovan and P. Secchi, eds. *Complex data modeling and computationally intensive statistical methods*. Springer, Milan, Italy.
- Martins, C. C. A., M. E. Morete, M. H. Engel, A. C. Freitas, E. R. Secchi and P. G. Kinas. 2001. Aspects of habitat use patterns of humpback whales in the Abrolhos Bank, Brazil, breeding ground. *Memoirs of the Queensland Museum* 47:83–90.
- Martins, C. C. A., A. Andriolo, M. H. Engel, P. G. Kinas and C. H. Saito. 2013. Identifying priority areas for humpback whale conservation at Eastern Brazilian Coast. *Ocean & Coastal Management* 75:63–71.
- Morete, M. E., R. M. Pace, C. C. A. Martins, A. C. Freitas and M. H. Engel. 2003. Indexing seasonal abundance of humpback whales around Abrolhos Archipelago, Bahia, Brazil. *Latin American Journal of Aquatic Mammals* 2:21–28.
- Morete, M. E., T. L. Bisi and S. Rosso. 2007. Temporal pattern of humpback whale (*Megaptera novaeangliae*) group structure around Abrolhos Archipelago breeding region, Bahia, Brazil. *Journal of the Marine Biological Association of the United Kingdom* 87: 87–92.
- Oña, J., E. C. Garland and J. Denking. 2016. Southeastern Pacific humpback whales (*Megaptera novaeangliae*) and their breeding grounds: Distribution and habitat preferences of singers and social groups off the coast of Ecuador. *Marine Mammal Science* 33:219–235.
- Pavanato, H. J., L. L. Wedekin, F. R. Guilherme-Silveira, M. H. Engel and P. G. Kinas. 2017. Estimating humpback whale abundance using hierarchical distance sampling. *Ecological Modelling* 358:10–18.
- Pendleton, D. E., A. Pershing, M. W. Brown, C. A. Mayo, R. D. Kenney, N. R. Record and T. V. N. Cole. 2009. Regional-scale mean copepod concentration indicates relative abundance of North Atlantic right whales. *Marine Ecology Progress Series* 378: 211–225.
- R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Redfern, J. V., M. C. Ferguson, E. A. Becker, *et al.* 2006. Techniques for cetacean-habitat modeling. *Marine Ecology Progress Series* 310:271–295.
- Roos, M., and L. Held. 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* 6:259–278.
- Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical modeling and inference in ecology*. Academic Press, San Diego, CA.
- Rosenbaum, H. C., C. Pomilla, M. Mendez, *et al.* 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLOS ONE* 4(10):e7318.
- Rue, H., S. Martino and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71:319–392.
- Seyboth, E., K. R. Groch, E. R. Secchi and L. Dalla Rosa. 2015. Habitat use by southern right whales, *Eubalaena australis* (Desmoulins, 1822), in their main northernmost calving area in the western South Atlantic. *Marine Mammal Science* 31:1521–1537.

- Siciliano, S. 1995. Preliminary report on the occurrence and photo-identification of humpback whales in Brazil. Report of the International Whaling Commission 45:138–140.
- Spiegelhalter, D. J., N. J. Best, B. P. Carlin and A. Van Der Linde. 2002. Bayesian measure of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583–639.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. Van Der Linde. 2014. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76:485–493.
- Stamps, J. 2009. Habitat selection. Pages 38–44 in S. A. Levin, S. R. Carpenter, H. C. Godfray, *et al.*, eds. *The Princeton guide to ecology*. Princeton University Press, Princeton, NJ.
- Steiger, G. H., J. Calambokidis, J. M. Straley, *et al.* 2008. Geographic variation in killer whale attacks on humpback whales in the North Pacific: Implications for predation pressure. *Endangered Species Research* 4:247–256.
- Stevick, P. T., J. Allen, P. J. Clapham, *et al.* 2006. Population spatial structuring on the feeding grounds in North Atlantic humpback whales (*Megaptera novaeangliae*). *Journal of Zoology* 270:244–255.
- Tyack, P., and H. Whitehead. 1983. Male competition in large groups of wintering humpback whales. *Behaviour* 83:132–154.
- Vehtari A., A. Gelman and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27:1413–1432.
- Ward, E., A. N. Zerbini, P. G. Kinas, M. H. Engel and A. Andriolo. 2011. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). *Journal of Cetacean Research and Management (Special Issue 3)*:145–149.
- Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11:3571–3594.
- Wedekin, L. L., M. C. Neves, C. Baracho, M. R. Rossi-Santos and M. H. Engel. 2010. Site fidelity and movements of humpback whales (*Megaptera novaeangliae*) on the Brazilian breeding ground, southwestern Atlantic. *Marine Mammal Science* 26:787–802.
- Whitehead, H., and M. J. Moore. 1982. Distribution and movements of West Indian humpback whales in winter. *Canadian Journal of Zoology* 60:2203–2211.
- Winn, H. E., and N. E. Reichley. 1985. Humpback whale *Megaptera novaeangliae* (Borowski, 1781). Pages 241–273 in S. H. Ridgway and S. R. Harrison, eds. *Handbook of marine mammals*. Volume 3. The sirenians and baleen whales. Academic Press, San Diego, CA.
- Zembruscki, S. G. 1979. Geomorfologia da margem continental sul brasileira e das bacias oceânicas adjacentes [Geomorphology of the Brazilian southern continental margin and adjacent ocean basins]. Pages 169–191 in H. A. F. Chaves, ed. *Reconhecimento global da margem continental brasileira: Projeto REMAC [Global recognition of the Brazilian continental margin: REMAC Project]*. Petrobrás, Rio de Janeiro, RJ.
- Zerbini, A. N., A. Andriolo, J. M. Da Rocha, *et al.* 2004. Winter distribution and abundance of humpback whales (*Megaptera novaeangliae*) off northeastern Brazil. *Journal of Cetacean Research and Management* 6:101–107.
- Zerbini, A. N., A. Andriolo, M. P. Heide-Jørgensen, *et al.* 2006. Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. *Marine Ecology Progress Series* 313:295–304.
- Zerbini, A. N., E. Ward, M. H. Engel, A. Andriolo and P. G. Kinas. 2011. A Bayesian assessment of the conservation status of humpback whales (*Megaptera novaeangliae*) in the western South Atlantic Ocean (breeding Stock A). *Journal of Cetacean Research and Management (Special Issue 3)*:129–144.

Received: 13 May 2016  
Accepted: 2 December 2017

## SUPPORTING INFORMATION

The following supporting information is available for this article online at <http://onlinelibrary.wiley.com/doi/10.1111/mms.12492/suppinfo>.

*Appendix S1.* Comparison between both spatial resolutions.

*Figure S1.* PIT histograms for candidate models A (null) to H (global).

*Figure S2.* Observed (red line) and predicted (posterior mean: thin black line; 95% probability interval: thin blue line) number of groups by grid cell.

*Figure S3.* Density of groups (posterior mean) by km<sup>2</sup> obtained from model A (null).

*Figure S4.* Density of groups (posterior mean) by km<sup>2</sup> obtained from model F (global).

*Figure S5.* Number of groups and effort (km) by grid cell for the 0.5° spatial resolution.

*Figure S6.* PIT histograms for candidate models A (null) to H (global) for the 0.5° spatial resolution.

*Figure S7.* Covariate effects on group density in logarithmic scale for the 0.5° spatial resolution. Continuous black lines indicate the posterior means; thin black lines indicate the 95% probability intervals.

*Figure S8.* Observed and predicted (posterior mean) number of groups by grid cell for the 0.5° spatial resolution.

*Figure S9.* Observed (red line) and predicted (posterior mean: thin black line; 95% probability interval: thin blue line) number of groups by grid cell for the 0.5° spatial resolution.

*Figure S10.* Density of groups (posterior mean) by km<sup>2</sup> obtained from model E ( $d + b$ ) for the 0.5° spatial resolution.

*Figure S11.* Probability of occurrence (posterior mean) by grid cell obtained from model E ( $d + b$ ) for the 0.5° spatial resolution.

*Figure S12.* Posterior mean of the covariate effect (fixed) and spatial autocorrelation effect (random) in logarithmic scale by grid cell obtained from model E ( $d + b$ ) for the 0.5° spatial resolution.

*Figure S13.* Relationship between the posterior distributions of the spatial autocorrelation and covariate effect obtained from model E ( $d + b$ ) for the 0.5° spatial resolution. The black line indicates a linear model fitted between both effects.

*Figure S14.* Distance from shore (km) by grid cell.

*Figure S15.* Wind speed (m/s) by grid cell.

*Figure S16.* Bathymetry (m) by grid cell.

*Table S1.* Summary of candidate models including: number of parameters ( $p_{\text{WAIC}}$ ), Watanabe-Akaike Information Criterion (WAIC) and logarithmic Conditional Predictive Ordinates ( $\overline{\text{LCPO}}$ ).

*Table S2.* Summary of the spatial and covariate effects in the northeast, east, and southeast continental margins.