Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia

Luis Bedriñana-Romano1,2 | Rodrigo Hucke-Gaete1,2 |
Francisco Alejandro Viddi1,2 | Juan Morales3 | Rob Williams4 | Erin Ashe4 |
José Garcés-Vargas1,5 | Juan Pablo Torres-Florez2,6 | Jorge Ruiz1,2

1Facultad de Ciencias, Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile
2NGO Centro Ballena Azul, Valdivia, Chile
3Laboratorio Ecotono, INIBIOMA-CONICET, Universidad Nacional del Comahue, Bariloche, Argentina
4Oceans Initiative, Seattle, WA, USA
5Departamento de Genetica e Evolución, Universidade Federal de São Carlos, São Carlos, SP, Brazil

Correspondence
Luis Bedriñana-Romano, Facultad de Ciencias, Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile.
Email: luis.bedrinana.romano@gmail.com

Funding information
Fondo Nacional de Desarrollo Regional, Gobierno Regional de Los Lagos, Grant/Award Number: BIP N°30040215-0 ID1857-17LP07; Whitley Fund for Nature; Kilverstone Wildlife Charitable Trust; World Wildlife Fund (WWF) Germany

Abstract
Aim: Species distribution models are useful tools for depicting important habitat, assessing abundance and orienting conservation efforts. For small populations in poorly studied ecosystems, available data are often scarce and patchy. To overcome this limitation, we aim to evaluate the use of different data types within a hierarchical Bayesian framework with the goal of modelling the abundance and distribution of a small and highly migratory population of blue whale (BW, Balaenoptera musculus) summering in Chilean Northern Patagonian (CNP).

Location: CNP, Eastern South Pacific (ESP).

Methods: We constructed a Bayesian hierarchical species distribution Model (HSDM), combining a binomial N-mixture model used to model BW groups counts in line-transect data (2009, 2012 and 2014) with a logistic regression for modelling presence-availability data (2009–2016), allowing both models to share covariate parameters for borrowing strength in estimations.

Results: Distance to areas of high chlorophyll-a concentration during spring before summering season (AHCC-s) was the most important and consistent explanatory variable for assessing BW abundance and distribution in CNP. Incorporating accessorial presence-only data reduced uncertainty in parameters estimation when comparing with a model using only line-transect data, although other covariates of secondary importance failed to be retained in this model.

Main conclusions: Our results remark the capability of HSDM for integrating different data types providing a potential powerful tool when data are limited and heterogeneous. Results indicate that AHCC-s, and possibly thermal fronts, could modulate BW abundance and distribution patterns in CNP. Preliminary model-based delimitations of possible priority conservation areas for BW in CNP overlap with highly used vessel navigation routes and areas destined to aquaculture.

KEYWORDS
abundance assessment, blue whale, Chilean Patagonia, habitat selection, hierarchical Bayesian models, marine conservation, species distribution models
INTRODUCTION

Located in the Eastern South Pacific (ESP), Chilean Northern Patagonia (CNP) represents a region of both elevated primary and subsequent secondary productivity, which are mainly mediated by the interplay of a large input of low salinity micronutrient-loaded superficial fresh water, high salinity macronutrient-loaded oceanic deeper water and drastic seasonal changes in light regimes (González et al., 2011; Iriarte, González, & Nahuelhual, 2010; Torres et al., 2011). These complex oceanographic processes sustain a high but poorly studied marine biodiversity, including the most important summering and nursery area for the endangered blue whale (BW, *Balaenoptera musculus*) in the ESP (Hucke-Gaete, Osman, Moreno, Findlay, & Ljungblad, 2004; Hucke-Gaete, Ruiz, & Alvarex, 2010; Reilly et al., 2008). It also hosts one of the largest aquaculture industries in the world (Buschmann et al., 2006; Niklitschek, Soto, Lafon, Molinet, & Toledo, 2013).

*Balaenoptera musculus* are a local priority for conservation and marine spatial planning initiatives (Hucke-Gaete et al., 2010; Outeiro et al., 2015) due to potential threats, including ship strikes, fishery and aquaculture net entanglement, poorly regulated whale watching activities and noise (Colpaert, Briones, Chiang, & Sayigh, 2016; Hoyt & Ilíghéz, 2008; Hucke-Gaete et al., 2010, 2013; Van Waerebeek et al., 2007; Viddi, Harcourt, & Hucke-Gaete, 2015). Thereby, reliable spatially explicit density maps for spatial planning and risk assessment are needed (Pennino et al., 2017; Redfern et al., 2013; Williams et al., 2014). Additionally, a new US trade rule requires countries, including Chile, to demonstrate that their fishery and aquaculture activities are comparable in effectiveness to the US Marine Mammal Protection Act or risk losing the ability to export seafood products to the lucrative US market (Williams, Burgess, Ashe, Gaines, & Reeves, 2016). These national and international efforts create a pressing need for information on BW abundance and distribution in the region.

During 2009, private and public agencies joined efforts to generate the first abundance estimate for BW in CNP covering most of its coastal waters, as part of a broader initiative setting a baseline for the implementation of a multiple-use marine protected area (Hucke-Gaete et al., 2010). After 1,278 km of dedicated marine survey, only 33 BW groups were recorded on-effort yielding an abundance estimate of 222 (115–430) for an area of 34,899 km², using classic design-based distance sampling methods, a collection of independent BW sightings database and oceanographic data obtained from satellite, over a period of 8 years (2009–2016).

To test the hypothesis that areas of high chl-a concentration and/or areas of recurrent TF trigger higher BW abundance and therefore modulates its distribution in CNP, we constructed a hierarchical species distribution model (HSDM) integrating classical distance sampling methods, a collection of independent BW sightings database (presence-only), and oceanographic data obtained from satellite, over a period of 8 years (2009–2016).

METHODS

2.1 Line-transect surveys

Line-transect surveys were carried out off CNP (41°00’S–45°30’S) during austral summer of 2009 and early autumns of 2012 and 2014, covering distinct but overlapping areas and never exceeding ca. 25 km from the nearest coast (Figure 1). Field protocols
followed standard line-transect survey methods (Buckland et al., 2001) with some specific modifications for small-boat surveys (Dawson, Wade, Slooten, & Barlow, 2008; Williams et al., 2017). A 17-m motor vessel was used for most surveys, except for the exposed western coast of Chiloe Island during 2009 where a 17-m sailboat was used. The observer team comprised three persons, plus a fourth person operating the computer to enter data. An angle board mounted on the deck was used to measure radial angle to the group of animals, and visual distance estimate to the animals was estimated at first sighting. To calibrate visual distance estimation, radial distance estimates from observer-specific estimations were corrected (Williams, Leaper, Zerbini, & Hammond, 2007) using known distances from landmarks (i.e., islands, lighthouses, salmon farms and other vessels) derived from the vessels’ radar.

2.2 | Presence-only data

Focal-group marine surveys were undertaken in the Corcovado Gulf and Reloncavi Sound during summer/autumn of 2004–2016 (Figure 1). These surveys were undertaken mostly on board a 7.6-m semi-rigid hull inflatable vessel. Additional surveys from exploratory flights or cruise expeditions that were not specifically aimed for marine mammal research were also available, covering other areas of CNP. During all these surveys, observers recorded the groups’ geographical position using a handheld GPS, time of sighting and group size. Only data from 2009 to 2016 were used in subsequent modeling because before 2009, data were restricted to CGMC. This yielded a total of 180 BW groups GPS locations; however, as these data are likely to be highly spatially and temporally auto-correlated, we used a raster with grid cells of 8 km per side from the study area as sampling units and assigned a value of 1 on each grid cell that presented at least one BW sighting on it for each year. Grid cells with BW sightings in more than 1 year were not considered duplicates as their associated covariates were different. This resulted in a database of 81 blue whale presences for analysis.

2.3 | Satellite oceanographic and topographic covariates

Chl-$\alpha$ concentration data were extracted from satellite level 3 images from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the Aqua satellite (Data set ID: erdMH1chlamday), corresponding to monthly averages in a grid size of 4.64 km. Based on these images, four different covariates were constructed, (1) spring average chl-$\alpha$ concentration: generated by a composite of satellite images from September, October and November (austral spring) from the year before (2008–2015) each selected field season (2009–2016); (2) summer average chl-$\alpha$ concentration: the same as the later but using images from January, February and March of each selected field season (2009–2016); (3) distance to areas of high chlorophyll-$\alpha$ concentration during spring (AHCC-s): consisted in distance to polygons enclosing areas with an average chl-$\alpha$ concentration equal or higher than 5 mg/m$^3$ (Montero et al., 2011) during spring months; and (4) AHCC-su: the same as the later but using summer months.

Daily averages level 4 SST satellite images were obtained from Multi-Scale Ultra-High Resolution (MUR) SST Analysis database (Data set ID: jplMURSST41). MUR-SST maps merge data from different satellites, combined with in situ measurements, using the Multi-Resolution Variational Analysis statistical interpolation (Chin, Milliff, & Large, 1998), in a grid size of 0.01° (ca. 1 km$^2$). From MUR-SST maps, thermal fronts (TF) were identified using a single-image-edge-detection (SIED) algorithm with a threshold detection of 0.5°C (Cayula & Cornillon 1992; Cayula & Cornillon 1995). SIED tools were available in ArcMap 10.1 through the Marine Geospatial Ecology Tools (Roberts, Best, Dunn, Treml, & Halpin, 2010). Areas of thermal front recurrence (ATFR) were constructed using SIED on daily images ranging from 1 January to 30 April for each year skipping every third day for time-saving purposes during data analysis. A composite of this set was used to account for the times each grid cell within the study area was catalogued as holding a TF during the 4 months span.
of every year. These new rasters presented grid cell values in a range of 0–16 TF detections within the 4 months’ time span. Percentiles for all rasters were very similar across years showing that 90% of the grid cells for the entire study area presented three or less TF detections. Therefore, we selected a threshold value of 4 for constructing polygons yielding ATFR, which only represent areas where the strongest and most conspicuous TF occur. Distance to these ATFR polygons was used as a covariate in the subsequent models.

Summer SST was extracted from level 3 monthly composites from the Aqua MODIS satellite database in a grid size of 4 km (Data set ID: erdMH1sstdmday). Data from January to March of every year (2009–2016) were averaged and this composite was used for data extraction. Raw data for depth were obtained from the Chilean Navy (Servicio Hidrográfico y Oceanográfico de la Armada), from which a triangular irregular network (TIN) model was created using 3D Analyst in ArcGIS and the resultant raster was used to extract depth values. Distance to the coast (DTC) was extracted in QGIS (QGIS Development Team, 2009) and was also used as a covariate.

### 2.4 Modelling approach

We divided on-effort tracks from line-transect surveys into contiguous equal-sized sampling segments (Hedley & Buckland, 2004; Williams, Hedley, & Hammond, 2006) of 8 km per side (64 km²). For each one of these sampling segments (hereafter tracks), a response variable, BW group counts and environmental covariates were extracted assuming the centroid of each track as the spatial point from which the covariates were extracted. Before analyses, all variables were standardized and correlations were assessed through Pearson correlation analysis.

Based on the binomial N-mixture model from Chelgren et al. (2011), true BW group abundance $N_i$ for each track $i$ was modelled by a Poisson distribution, which we modified to be modelled through a zero-inflated version

$$Pr\{N_i > 0\} = \sum_{j=0}^{\infty} \frac{\lambda_j^n e^{-\lambda_j}}{n!},$$

where $\psi$ is the probability of a non-zero true abundance and $\lambda_j$ is the usual Poisson parameter, which depends on the exponential of a linear function of covariates

$$\lambda_j = \text{Habitat}_i \times e^{(\beta_0 + \beta_1X_i)}.$$

Habitat$_i$ is an offset term accounting for effective area sampled at each transect (subtracting land cover when required). $\beta_0, \beta_1$ are intercepts which are calculated for each year $y$, $\beta$ is a vector of parameters coefficients and $X_i$ is the corresponding design matrix. Intercepts $(\beta_0, \beta_1)$ were assumed to come from a normal distribution, for which we estimated its respective mean and variance hyperparameters. If required, model selection was performed through a “model identity” variable with each category representing a unique set of covariates, allowing to draw a posterior probability for each one from within the HSDM (Kruschke, 2014; Royle, Chandler, Sollmann, & Gardner, 2013).

Instead of assuming the observed number of BW groups $n_i$ in each track as the true local group abundance ($N_i$), a second part relates $n_i$ to $N_i$ as a binomial outcome with probability of success determined by detection probability $p_i$, thus making $N_i$ a latent variable.

$$Pr\{n_i|N_i, p_i\} = \left(\frac{N_i}{n_i}\right) p_i^n (1-p_i)^{N_i-n_i}$$

Assuming a truncation distance of 4 km, the probability of detection $p_i$ was derived from un-binned perpendicular distances $y_d$ from each $d$ detection, using a half-normal distribution with a single parameter $\Sigma$. Even when we did not use covariates for modelling $\Sigma$, subscript still applies to $p_i$ to account for differences in Habitat, (Chelgren et al., 2011). Error in distance estimation was assessed by regressing estimated training distances to a series of landmarks by observers against true distances provided by the vessel’s radar. The slope of this linear regression (Figure S1.1) was used to divide angular distances to BW groups previously to multiplying by the sine of the angle in perpendicular distance estimation (Hammond et al., 2002).

Maps of whale densities predictions and associated uncertainties from 2009 to 2016 were calculated using rasters of the entire study area with a grid-cell size of 8 km per side. Model’s estimated parameters were used to predict the number of whales in each grid cell, based on their standardized covariates values, estimated mean group size and estimated intercept parameter for the corresponding year if line-transect data were available or drawing an intercept value from the normal distribution of intercepts if not. As the bulk of the data was gathered in 2009, we provide an overall abundance estimate only for this year, which was equal to the sum of all grid-cell values in that year.

As a measure of goodness-of-fit, we conducted a posterior predictive check (PPCheck, Gelman, Meng, & Stern, 1996) based on chi-squared tests, which allowed us to calculate the ratio between the sum of discrepancy measures in observed and simulated data, the $c$-hat parameter, and a Bayesian $p$-value, which is the probability to obtain a test statistic that is at least as extreme as the observed test statistic computed from the actual data (should be around 0.5 for a good fit, Kery & Royle, 2015). All previously described steps were undertaken from within the HSDM as a one-stage approach (Miller, Burt, Rexstad, & Thomas, 2013).

Presence-availability set-ups are often modelled through a logistic regression that uses values of “one” for the recorded presence of the species and “zeros” for a sample of randomly selected points within the study area, termed pseudo-absences or availability data. Aarts, Fieberg, and Matthisiopoulos (2012) have shown that count, presence–absence and logistic regression models are all approximations of the inhomogeneous Poisson point process, for which the linear predictor function is proportional to the expected density of observations. Based on this, an alternative model (model 2) incorporated a logistic regression model for the presence-only data sharing covariate parameters with those of the main count model to borrow strength in estimations. To assess how the subjective process of
availability data selection (Beyer et al., 2010) influenced the new covariate parameters, we constructed five alternative model variants using different number of availability points and changing the geographical areas where these points were extracted from (Figure 2). As sample size was small, we also repeatedly ran the model leaving one data value out every time to check how the parameters posterior distribution fluctuates.

All models were fit in R (R Development Core Team 2015) and JAGS (Plummer, 2003) for Markov Chain Monte Carlo estimation methods. Vague priors were used for all parameters. Three chains were run in parallel through 100,000 iterations each. The first 20,000 samples were discarded as burn-in, and one of every two remaining samples was retained, for a total of 120,000 samples to form the posterior distribution of model parameter estimates. See Appendix S1 for more details about methods and results.

The 2009 abundance estimate and associated credible interval were used to estimate a precautionary minimum abundance estimate ($N_{min}$) to estimate a sustainable annual allowable harm limit from all anthropogenic sources of mortality, namely the "potential biological removal" ("PBR," Wade, 1998). Under US legislation, PBR is defined as the product of a minimum estimate of abundance ($N_{min}$) times one half of the maximum net productivity of a stock ($0.5 R_{max}$), times a recovery factor ($F_r$) between 0.1 and 1.0 (Wade, 1998). Guidelines for assessing marine mammal stocks are well established in the United States, and we follow convention sing the 80th percentile of the distribution as our value for $N_{min}$ and a 4% default value for $R_{max}$ (Wade, 1998). The recovery factor is a precautionary adjustment term governing the desired rate of recovery. We follow recommendations for recovery factors for endangered marine mammals and use a value of 0.1 for $F_r$ (Taylor, Michael, Heyning, & Barlow, 2003).

3 | RESULTS

On-effort tracks comprised 106 sampling units for 2009 (848 km), 35 for 2012 (272 km) and 47 for 2012 (368 km) (Figure 1). A total of 44 BW sightings were observed while on effort during all three line-transect surveys (2009 = 34, 2012 = 2, and 2014 = 8). Three sightings from 2009 were excluded after truncation of perpendicular distance data (4 km) yielding 41 sightings for analysis, including fitting the half-normal detection function (Figure S1.2). Group size ranged between 1 and 3 individuals, with a mean of 1.5.

Sea surface temperature was correlated with spring chl-a concentration and all chl-a-related covariates were correlated among them ($r ≥ .5, p < .01$). Only AHCC-s was retained by model 1 when used independently of other chl-a-related covariates in trial runs of this model. As AHCC-su presented a similar value for the covariate parameter but including zero in its posterior distribution, we wanted to address whether the inclusion of more data in model 2 modified this. Using the model identity variable showed that AHCC-s presented a probability of 0.82 against 0.18 of AHCC-su as the most likely to explain the data in trial runs of model 2. Based on these results, we used AHCC-s in all further models yielding five variables to be tested in both models, AHCC-s, ATFR, SST, DTC and depth.

Model 1 retained AHCC-s, ATFR, SST and DTC. An interaction parameter between the two most important covariates (AHCC-s and ATFR) was evaluated but not retained by the model. Only AHCC-s was retained by model 2 regardless of modifications on availability data selection (Table S1.1). AHCC-s was the only covariate retained in all models, experiencing a reduction in the CI and SD of parameters involved in calculating $\lambda$, when incorporating accessory presence-only data (Table 1). For 2009, the year with more data available, model 2 predicted larger total abundance (442, CI: 236–744) when comparing to model 1 (373, CI: 191–652). PPCheck results indicated that c-hat and Bayesian p-value presented values very near 1 and .5, respectively, indicating a good fit of the model to the data (Table 1 and Figure S1.3). Removing one sampling unit from analysis at a time did not produced large differences in parameters posterior distribution (Figure S1.4).

Plots of predicted density using both models showed a large variation in BW distribution among years (Figures 3 and 4). Although some areas such as Ancud Gulf and the Western Coast of Chiloe showed some consistency in concentrating higher BW densities, overall predictions uncertainty for those years where line-transect data were not available was large (Figure 3). To get a more straightforward comparison between models, we ran both models using only AHCC-s as covariate and focused only on those years where line-transect data were available to show how predictions uncertainty was reduced when incorporating presence-only data (Figures 2 and 4).

The point estimate of abundance in 2009 was 373 (model 1). Using 274 as the 20th percentile of the posterior distribution of $N$, we estimate a potential biological removal of 0.548, or one human-caused death or serious injury every 1.8 years.

4 | DISCUSSION

4.1 | Modelling approach

Literature provides several options to model abundance and distribution from non-systematic line-transect effort and sightings data (Hedley, Buckland, & Borchers, 1999; Redfern et al., 2006; Miller et al., 2013; and original references cited therein). Pooling all available information into one single count model would have added strength in covariate parameter estimation; however, this would have resulted in averaging all possible intercepts of the function modelling $\lambda$ and hence precluding depicting differences in abundance between years. Presence-availability designs would have provided an easy way of incorporating presence-only data as well, but at the expense of discarding information on the observation process and increased risk of bias in intercept estimation (Aarts et al., 2012; Lele & Keim, 2006).

Within a hierarchical Bayesian framework, we accommodated uncertainty on distance estimation, using different types of data to borrow strength for covariates parameters estimations (without
losing possible differences in absolute densities between years), allowed for alternative model comparison, estimated overall abundance and assessed goodness-of-fit, all with full uncertainty propagation capacity in one single modelling step. Further research should be conducted to understand the limitations of this approach as model comparison has not yet been broadly developed for hierarchical models (Kery & Royle, 2015).

The main difference between the more orthodox model 1 and the alternative model 2 was the exclusion of all other explanatory covariates besides AHCC-s, which could be the result of the spatial and temporal bias of presence-only data (see next section). Incorporating presence-only additional data reduced uncertainty associated with the most important covariate parameter, and less pronouncedly, to all intercepts (Table 1). The increase in the mean, SD and credible interval size for total abundance estimation during 2009 when using model 2 is interpreted as the loss of explanatory power that was provided mainly by ATFR in model 1 (Table 1). When using only AHCC-s as covariate in both models, we see that all parameters show smaller SD and credible interval size in model 2 (Figure 2). As any other method, our approach should not be regarded as an invitation to avoid properly designed systematic surveys, which should be always pursued (Buckland et al., 2001; Redfern et al., 2006). However, the capability of HSDM to integrate several sources of information might prove a valuable tool for assessing species abundance and distribution patterns when systematic and homogeneous data are limited. Eventually, new presence-only data can be added to improve covariate parameters. Additional line-transect data can be added to improve intercepts and their hyperparameters, thereby improving predictability for absolute abundance. Other nonlinear types of functions for modelling $\lambda$ and $p$, more sophisticated functions than

![Total Abundance 2009](image)

**FIGURE 2** Parameters posterior distributions for model 1 (blue line) and 2 (red line) using only distance to areas of high chlorophyll concentration during spring (AHCC-s) as covariate. For comparison purposes, the posterior distribution of $\beta_1$ AHCC-s is presented under different approaches, using a logistic regression for all combined data including line-transect data (black), using model 2 with 300 and 500 availability points samples from the entire study area (yellow, difference is indistinguishable), using model 2 with 300 and 500 availability points samples extracted only from Chiloe Inner Sea (red, the one concentrating more probability mass around the mean correspond to the variation using 500 samples) and using model 2 with all availability points (green). The mean accompanied with standard deviation and credible intervals (in parentheses) for each model is presented at the bottom of each plot.
suggesting that at the observed scale, krill availability is not necessary correlated with chol-a concentration. The location of high krill densities is mediated by advection processes, krill vertical migration, food availability and predator avoidance; therefore, at finer scales, spatio-temporal mismatch between chol-a concentration and krill abundance could occur if net export relocates krill biomass to neighbouring areas or if krill grazing impact phytoplankton abundance (Dorman, Powell, Sydeman, & Bograd, 2011; Mackas, Denman, & Abbott, 1985; Santora, Sydeman, Schroeder, Wells, & Field, 2011). The former could be the case here as results provided support for AHCC-s as the most important and consistent covariate (Table 1, Figure 2). This would imply that higher krill aggregations occur at or nearby areas of high productivty, which has been reported at least for Ancud Gulf and CGMC (Buchan & Quiñones, 2016; González et al., 2010, 2011), and that time-lags between two and 4 months are linking spring high chol-a concentration and BW occurrence in summer/autumn season, which is consistent with previous work on BWs elsewhere (Croll et al., 2005; Visser et al., 2011). Thermal fronts might be the result of tide or wind-induced upwelling and therefore co-occur within AHCC while TF concentrating effect might enhance primary production, making these features possibly correlated (Acha et al., 2004; Letelier, Pizarro, & Nuñez, 2009). As the interaction parameter between AHCC and ATFR parameters in this study was not retained in the model and ATFR was observed to occur in areas where AHCC was absent, we can argue that the influence of
FIGURE 3  Predicted blue whale abundance (Nt) and uncertainty (sd) at each grid cell (8 km per side) using model 1, from 2009 (top left) to 2016 (bottom right). Dots indicate blue whale groups’ locations recorded through line-transect surveys (black) or through focal-group marine surveys (white).
FIGURE 4 Comparison of predicted blue whale abundance (Nt) and uncertainty (sd) at each grid cell (8 km per side) when using model 1 (line-transect only) vs. model 2 (incorporating presence-only data through a logistic regression). Focusing only in those years where line-transect data were available, we show that uncertainty (posterior SD) was reduced in model 2. Dots indicate blue whale groups’ locations recorded through line-transect surveys (black) or through focal-group marine surveys (white).
TF on BW distribution is not necessarily a by-product of AHCC, but an independent phenomenon that might elicit a herding effect on prey. Studies performed on sympatric large whale species have found that the distribution of BW is more highly correlated with distance to TF than other species, presumably because unlike the main prey item for this species (krill), fish and other mobile prey species might escape the concentrating effect of TF (Doniol-Valcroze et al., 2007). The real effect of TF on modulating BW distribution here might have been underestimated as our gradient threshold of 0.5°C/km was extremely conservative, taking into consideration that BWs appear to be sensitive to TFs produced by SST gradients between 0.03 and 0.3°C/km (Etnoyer et al., 2006). This lack of resolution, the fact that these strong TFs tended to be more abundant in the Western coast of Chiloe Island and that presence-only data were restricted to Chiloe Inner Sea, could explain why ATFR was removed from model 2, but further research is needed to confirm this supposition.

4.3 | Interannual differences in BW abundance and distribution

Galletti-Vernazzani, Jackson, Cabrera, Carlson, and Brownell (2017) found no trends in population size for BW summering in CNP and hypothesized that apparent trends are better explained by temporal differences in habitat use. In the absence of systematic effort across years, our model predictions provide some insights into the interannual variations in BW distribution in the area.

During 2010 and 2012, both models predicted a decrease in BW abundance in CGMC (Figures 3 and 4). For both years, a drop in chl-α concentration during previous spring, a decrease in SST (for 2010), a decrease in the number of TF detected and salp outbreaks were observed (Figure S1.5, Buchan & Quiñones, 2016; Giesecke et al., 2014; Lara, Saldías, Tapia, Iriarte, & Broitman, 2016). These changes in oceanographic characteristics within CNP were hypothesized to be produced by an atypical pattern of oceanic sea surface currents mediated by negative anomalies in southern annular mode (SAM, Giesecke et al., 2014). SAM modulates intraseasonal to interannual changes in atmospheric conditions in the Southern Hemisphere, producing shifts in the westerly winds patterns and a significant decrease in SST in subtropical zones during negative anomalies (Hall & Visbeck, 2002; Marshall 2003; Lovenduski & Gruber, 2005) which did occur during 2010 and 2012 (Figure S1.6).

Although preliminary, these observations suggest that a reduction in spring chl-α concentration and weakening of TF produced by more homogeneous colder waters might result in a decrease in BW habitat suitability through diminished krill recruitment and/or availability. This reduction in BW abundance might be seized by other large whales that not rely exclusively on krill, as data on relative abundance for large whale species in CGMC from 2004 to 2012 showed that for most years, BW was by far the most frequent large whale species, a pattern that was disrupted during 2010 and 2012 when the relative abundance of humpback whales (Megaptera novaeangliae) and sei whales (Balaenoptera borealis) was higher (Table S2.2).

4.4 | Implications for blue whale conservation and future research

To date, all BW abundance estimates in ESP place the population in the mid-hundreds which suggests that the population has not yet reached pre-whaling levels (Galletti-Vernazzani et al., 2017; Hucke-Gaete et al., 2010; Torres-Florez, Hucke-Gaete, Rosenbaum, & Figueroa, 2014; Williams et al., 2011; this study). Although none of these estimations resulted from sampling the population’s entire distributional range, our estimates for 2009 are within the lowest range of those from Galletti-Vernazzani et al. (2017) for the same area. We consider this is an agreement for independent studies, with differences most likely explained by the utilized methods, as these authors used mark–recapture models that can account for
animals not typically using the study area (Pradel, Hines, Lebreton, & Nichols, 1997). A population of a few hundred whales, still recovering from commercial exploitation, has limited resilience to human-caused mortality and serious injury. Our analyses suggest that the population cannot withstand sustained mortality of even one whale annually from human activities.

Vessel collisions with large whales and salmon farm net entanglements have been reported in CNP (Hucke-Gaete et al., 2013; Van Waerebeek et al., 2007). Hence, additional information is needed on ship strikes and entanglement in fishing and aquaculture gear, to assess whether human-caused mortality and injury could be exceeding sustainable limits. So far, areas exhibiting most of expected BW abundance overlap with shipping routes and aquaculture activities (Figure 5). As such, more properly designed surveys are desired to validate current estimations of abundance and distribution patterns, while simultaneously assessing ship strike and entanglement risk to inform marine spatial planning initiatives (Becker et al., 2012; Redfern et al., 2006; Williams et al., 2017).

ACKNOWLEDGEMENTS

This research would not have been possible without the funding provided by the Fondo Nacional de Desarrollo Regional, Gobierno Regional de Los Lagos, BIP N°30040215-0 ID1857-17LP07, Whitley Fund for Nature, Kliverstone Wildlife Charitable Trust and WWF Germany to RHG. S. Buchan provided extra BW presence-only data. A. Zerbini provided comments on an earlier version of this manuscript. We are grateful to L/M Noctiluca crew, M. Novy, J. Barros, R. Contreras, N. Subercaseaux and R. Westcott whose commitment made this research possible. We used Logger 2000, developed by IFAW. LBR held a doctoral CONICYT-Chile fellowship.

DATA ACCESSIBILITY

Jags code for fitting the model is available in Appendix S1. Linetransact data, presence-availability data and accompanying covariate data are available in Data S1.

ORCID

Luís Bedriñana-Romano http://orcid.org/0000-0002-6105-0920

REFERENCES


Biologists often use distribution models to estimate abundance and evaluate the conservation status of marine mammals. Such methods are crucial for understanding and protecting biodiversity in marine environments.

### Additional Information

**How to cite this article:** Bedriñana-Romano L, Hucke-Gaete R, Viddi FA, et al. Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia. *Divers Distrib.* 2018;00:1–14. [https://doi.org/10.1111/ddi.12739](https://doi.org/10.1111/ddi.12739)

---

**BIOSKETCH**

Luis Bedriñana-Romano is a Ph.D. student at Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile. His current research interests are habitat selection drivers. For this, he is currently combining species distribution models, movement models and theoretical simulations to gain a deeper insight into this ecological process. The research team’s general focus is developing quantitative models for estimating abundance and distribution of marine mammals and using these models for conservation and management.

Author contributions: LB, RH and FAV conceived the idea. RW and EA designed systematic marine surveys. JPT and JR coordinated field campaigns. LB, RH, FAV, JPT, JR, RW and EA collected the data. LB, JM and JG analysed the data. All authors participated in manuscript writing.