Predicting distribution and relative abundance of mid-trophic level organisms using oceanographic parameters and acoustic backscatter

Pablo C. Escobar-Flores[1,*], Richard L. O'Driscoll[2], John C. Montgomery[1]

[1]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[2]National Institute of Water and Atmospheric Research, Private Bag 14-901, Kilbirnie, Wellington 6241, New Zealand

ABSTRACT: Data on the distribution and abundance of mid-trophic level organisms (MTLOs) in the pelagic open-ocean ecosystem are normally sparse or absent. Consequently, ecosystem models are limited in their ability to support decision-making for issues ranging from fisheries management to ecosystem resilience to climate change. We used acoustic data collected at 38 kHz frequency across the Southern Ocean (SO) between 2008 and 2014 to develop explanatory and predictive models for acoustic backscatter, a proxy for MTLO abundance in the epi- and mesopelagic zones. Boosted regression trees and generalised additive mixed models were used to develop simple predictive models for backscatter in the epi- and mesopelagic zones, using sea surface temperature, time of day (day/night) and depth. The resulting models predicted backscatter reasonably well in the Pacific sector of the SO, and in an independent dataset in the Indian sector of the SO. Our predictive models may provide a tool for inferring abundance and distribution of MTLOs in other parts of the SO.

KEY WORDS: Mid-trophic level · Southern Ocean · Mesopelagic fish · Myctophids · Boosted regression trees · Generalised additive mixed models

INTRODUCTION

The Southern Ocean (SO) has profound effects in the global ocean circulation, climate patterns and biogeochemical processes of carbon and nutrients (Rintoul et al. 2012). Because of the rapid climate-induced changes that the SO is experiencing, the Southern Ocean Observing System programme was established, and is currently in its implementation phase (Rintoul et al. 2012, Constable et al. 2014, 2016, Turner et al. 2014, Newman et al. 2015). This programme will increase the sampling effort of the biological components of the SO ecosystem, although the remoteness of the SO restricts potential observation. Alternatively, modelling approaches can make use of existing limited observations to interpolate in unsampled locations, assess the state of an ecosystem

by combining different datasets and investigate trends or changes over time (Rintoul et al. 2012). In this study, we developed explanatory and predictive statistical models for acoustic backscatter (s_a) in the epi- and mesopelagic zones in the SO, that can be used as an interpolation tool for unsampled areas.

Mid-trophic level organisms (MTLOs) or micronekton (referred to here as the mid-trophic level of open-ocean ecosystems), play a key role in the pelagic marine ecosystem by linking primary producers and tertiary consumers (Kloser et al. 2009, Catul et al. 2011). Their ecological significance is related to their large biomass (net and acoustic-derived biomass estimates of mesopelagic fish alone range from ~1–10 billion tons; Gjøsaeter & Kawaguchi 1980, Irigoien et al. 2014) and their role as both predators and prey (Mann 1984, Beamish et al. 1999).

In addition to their ecological functional role, MTLOs participate actively in the biochemical cycles of the open-ocean ecosystems. They facilitate energy, nutrient and carbon fluxes across the water column, and affect oxygen consumption at different depth strata through their diurnal vertical migration (DVM) (Bianchi et al. 2013, Davison et al. 2013, Irigoien et al. 2014). Micronekton is comprised of organisms (1−20 cm) such as small fish, squid, crustaceans and gelatinous animals (Lehodey et al. 2010). Mesopelagic fish are the largest unexploited biomass in marine ecosystems worldwide (Gjøsaeter & Kawaguchi 1980), and recent estimates suggest that their biomass is about 1 order of magnitude higher than previously thought (Irigoien et al. 2014). Mesopelagic fish abundance is usually underestimated by net sampling due to escape and avoidance (Collins et al. 2012, Kaartvedt et al. 2012); therefore, other sampling techniques such as fisheries acoustics have been used (e.g. Irigoien et al. 2014).

Fisheries acoustics is commonly used for studying fish abundance and distribution (Simmonds & MacLennan 2005). Mesopelagic fish often dominate acoustic scattering layers throughout the water column (e.g. Barham 1966), and acoustic data have the potential to provide information about the mid-trophic levels (Lehodey et al. 2010, 2015, Handegard et al. 2013). Most lanternfish (family Myctophidae) have gas-filled swimbladders, making them good sound-reflecting targets (Yasuma et al. 2006); they are therefore considered one of the most important components of the oceanic deep scattering layers (Catul et al. 2011). Collecting acoustic data onboard dedicated research vessels for ecosystem monitoring endeavours and research studies of marine organisms can be costly for research institutes, so vessels of opportunity have become common research platforms in acoustics (Karp 2007). Vessels of opportunity expand the acoustic sampling effort spatially and temporally, and the data collected can be used to understand the relative distribution and abundance and dynamics of the MTLOs (e.g. Escobar-Flores et al. 2013).

Fisheries acoustics is based on the linearity principle that backscatter is proportional to an organism's density. Although this principle holds true for targets with similar acoustic properties, in large-scale surveys where mid-trophic community composition is unknown and likely to change spatially (e.g. Koubbi et al. 2011a), it may not apply (e.g. Davison et al. 2015). This relates to the variability of the acoustic contribution of different organisms (target strength), which is species-specific and even changes between individuals of the same species (ontogenetic changes), and to resonance scattering effects from different organisms at varying sizes (e.g. Peña et al. 2014, Davison et al. 2015, Kloser et al. 2016). Lanternfishes display a variety of swimbladder shapes, as well as changes in age-related swimbladder content and atrophy (Marshall 1960, 1971, Davison 2011), complicating the interpretation of acoustic data as a proxy for biological density. Despite this uncertainty, large-scale patterns of acoustic-derived measurements can provide relative indices of distribution and abundance of organisms consistently over time (Griffiths et al. 2005), and be used in large-scale studies to characterise the general status of marine ecosystems. Quantifiable indicators can be overlapped with groups of species or trophic levels, providing information about the performance of management strategies, or associate their performance to the state of other trophic levels (Trenkel et al. 2011). Understanding patterns of prey is important to determine the spatio-temporal dynamics of large apex predators in the open ocean (Bertrand et al. 2003, Lehodey et al. 2010).

Ecosystem models including MTLOs, e.g. ATLAN-TIS (Fulton et al. 2004), APECOSM (Maury et al. 2007) and SEAPODYM (Lehodey et al. 2008, 2010), which are constrained by the lack of observations of mid-trophic levels, will benefit from increased acoustic data collection (Handegard et al. 2013). Although programmes for developing cost-effective sampling platforms have been proposed (e.g. Mid-trophic Automatic Acoustic Sampler [MASS], Handegard et al. 2009) — and there are collaborative projects to process and make acoustic data accessible to users through online repositories (e.g. the Integrated Marine Observing System [IMOS] collaborative program in Australia, www.imos.org.au, and the European Mesopelagic Southern Ocean Prey and Predator [MESOPP] project, www.mesopp.eu) — the vast extent of the pelagic open-ocean ecosystem cannot be entirely covered by opportunistic data. Consequently, modelling approaches have been advocated to overcome the scarcity of available observations (Rintoul et al. 2012).

The aim of this study was to use opportunistically collected acoustic data and a set of readily available environmental variables to develop explanatory and predictive models for backscatter (as a proxy for MTLO abundance). The overarching goal was to provide predictions of acoustic backscatter to describe large-scale patterns of MTLOs in unsampled areas of the SO. The acoustic data used in these models were collected opportunistically from toothfish fishing vessels (TFVs) transiting across the SO between New

Zealand (NZ) and the Ross Sea (RS). Satellite-derived and geographical information (retrieved from online repositories), climatologies and temporal variables were used to develop the models. Boosted regression trees (BRTs) (Elith et al. 2008), and generalised additive mixed models (GAMMs) (Wood 2006) were used to model backscatter. The predictions from the resulting models were tested in the Pacific sector of the SO (training location) and also in an independent area in the Indian sector of the SO.

MATERIALS AND METHODS

Acoustic data

The acoustic dataset consisted of 28 acoustic transects conducted opportunistically in the Pacific sector of the SO between November and March from 2008 to 2014 (Table 1, Fig. 1; dataset available through the IMOS portal, https://portal.aodn.org.au/). This period corresponds to the Antarctic toothfish fishing season when the ice retreats close to the Antarctic continent, allowing access to toothfish fishing grounds located in the RS region. Transects were collected by 3 TFVs during their transits to and from the fishing grounds, and by RV 'Tangaroa' during its transits to and from the RS for research purposes (Table 1). The TFVs ('San Aotea II', 'San Aspiring' and 'Janas') collected 38 kHz single frequency data while the RV 'Tangaroa' collected multi-frequency data. Only the 38 kHz data collected on the RV 'Tangaroa' were used in this study. The National Institute of Water and Atmospheric Research (NIWA) calibrated all echosounders on a regular basis following procedures as per Demer et al. (2015) using a 38.1 mm tungsten carbide sphere. Transceiver settings used during data collection were typically 2 ms pulse length and 2000 W power.

The acoustic data were processed following the IMOS Bio-acoustic Ships of opportunity (BASOOP) sub-facility protocols (see Ryan et al. 2015), with modifications described by Escobar-Flores (2017). Acoustic backscatter, s_a, defined by MacLennan et al.

Table 1. Acoustic dataset breakdown indicating the number of transects by vessel and year

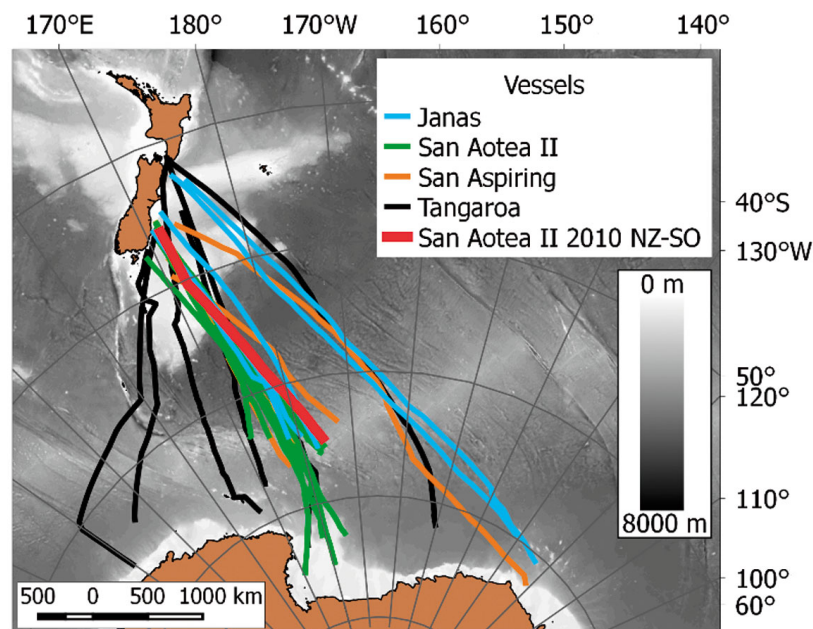| Vessel name	Transects (n)	2008	2009	2010	2011	2012	2013	2014
'San Aotea II'	11	1		2	2	2	2	2
'San Aspiring'	5			1	2	1	1	
'Janas'	6			1	2		1	2
'Tangaroa'	6	2		2			2	

Fig. 1. Acoustic transects (n = 28) conducted between 2008 and 2014 between New Zealand (NZ) and the Southern Ocean (SO) by the fishing vessels of opportunity 'San Aotea II', 'San Aspiring' and 'Janas', and the RV 'Tangaroa', used for developing explanatory and predictive models. The thick red line represents the 'San Aotea II' transect in 2010 used to compare prediction outputs (see 'Results' section)

(2002) in m^2 m^{-2} units was echo-integrated in 1 km long and 10 m vertical bins, and scaled by 1×10^6 for practicality (units in m^2 km^{-2}).

Explanatory variables

The relationships between mean s_a and 9 explanatory variables were analysed. Three surface oceanographic variables were obtained from satellite observations: sea surface temperature (SST), chlorophyll (chl a) and sea surface height (SSH). Salinity and potential temperature at depth for the epi- and mesopelagic zones were obtained from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas (CARS)

oceanographic climatologies. Three temporal variables were estimated from the acoustic data: fishing year (covering the toothfish fishing season, i.e. Nov 2009 to March 2010 defined as the 2010 fishing year); season (spring [November to December] or summer [January to March]); and time of the day (discretised as day or night). To classify day or night, sunrise and sunset times were estimated for each horizontal bin using the 'sunriset' function from the R package 'maptools' (Bivand & Lewin-Koh 2015), based on the civil twilight daytime extent definition (i.e. sun located 6° below the horizon). The mean projected latitude and longitude of each bin were calculated using the SpatialPoints function from the R package 'sp' (Pebesma & Bivand 2005, Bivand et al. 2013), and used in the 'sunriset' function along with the date of data collection. Bins with a time stamp between sunrise and sunset were flagged as day, and as night otherwise. Due to the earth's rotation and angle of inclination during summer within the Antarctic Circle, the light period extends up to 24 h. Bins collected within the 24 h of light period were flagged as day. Ocean floor depth was obtained from the General Bathymetric Chart of the Oceans 2014 dataset. Spatial and temporal resolutions of variables are shown in Table 2.

Satellite-derived variables, bathymetry (depth) and CARS salinity and potential temperature data compiled in NetCDF files were extracted using the statistical software R (R Core Team 2013) and the package 'ncdf' (Pierce 2014). A bilinear interpolation was performed for the mean latitude and longitude to assign the variables to each 1 km bin, using the interp.surface function of the 'fields' R package (Nychka et al. 2015).

Statistical modelling

Acoustic data integrated at 1 km horizontal and 10 m depth resolution were summed vertically to obtain mean s_a over the epi- and mesopelagic zones.

The epipelagic zone was defined between 10 m from the surface down to 200 m. The mesopelagic zone was from 200 to 1000 m, or to the maximum range where integration bins achieved criteria of 70% of accepted samples (Escobar-Flores 2017).

The original acoustic data (1 km horizontal bins) were averaged horizontally at 3 arbitrary coarser resolutions of 10 km (n = 5658), 25 km (n = 2444) and 50 km (n = 1301). Assuming an average vessel transit speed of 10 knots, the temporal resolutions of the 3 new horizontal bin sizes were approximately 30, 75 and 150 min. Modelling mean s_a using 3 different horizontal resolutions had 2 main objectives: first, to test modelling computation time; second, to perform sensitivity analysis relative to the bin size selected. The response variable, mean s_a, was $\log_{10}$ and cubic root-transformed for the epi- and mesopelagic zones, respectively, aiming to attain a normal distribution.

Six models were examined, 1 model for each pelagic zone (epi- and mesopelagic zones) at 10, 25 and 50 km horizontal bin sizes. These models are referred to hereinafter as the preliminary (explanatory) models.

We used 2 different modelling techniques, i.e. GAMMs and BRTs, to compare model performance and variable selection consistency across techniques. This information was used as one of the model validation tools. Both GAMMs and BRT models allow correlation structures to be specified using the approaches described below.

GAMMs

Generalised additive modelling (Hastie & Tibshirani 1990) is a semi-parametric regression technique that accommodates a variety of distributions for the response, whose relationship with the explanatory variables is modelled not only through coefficients, but through partial non-parametric

Table 2. Explanatory variables used for developing statistical models

Variable	Units	Spatial resolution	Temporal resolution	Product
Sea surface temperature (SST)	°C	4 km	Monthly	MODIS-Aqua day and MODIS-Aqua night
Chlorophyll-*a* (chl *a*)	mg m^{-3}	4 km	Monthly	MODIS-Aqua
Sea surface height (SSH)	m	0.25° grid	Daily	Global Ocean Gridded Sea Level Anomalies SSALTO/Duacs
Bathymetry	m	30 arc second (≈1 km)	-	General Bathymetric Chart of the Oceans
Potential temperature	°C	0.5° grid	-	CSIRO Atlas of Regional Seas (CARS)
Salinity	PSU	0.5° grid	-	CARS

smoothing curves (Hastie & Tibshirani 1990). Generalised additive models (GAMs) can be used as predictive models or as an exploratory method (O'Brien & Rago 1996), because of their ability to disentangle the underlying structure between variables, which is not feasible by other linear methods (Hastie & Tibshirani 1990, Guisan et al. 2002). By revealing the structure of the relationship, explanatory models aid comprehension of the ecological processes driving patterns observed in the data (Guisan et al. 2002). The principal strength of GAMs is their potential to model highly non-linear and non-monotonic relationships (Guisan et al. 2002).

GAMMs (see Wood 2006) combine GAMs and mixed effect models, allowing auto-correlation and multiple variances (Zuur et al. 2007). Mixed effect models are suitable for data generated from a hierarchical underlying procedure, which may have fixed or random coefficients and multiple error terms (Zuur et al. 2007). In mixed models, a positive definite matrix specifies the auto-correlation in the residuals through non-complex models, implying that these are no longer independent (Wood 2006). Models to specify the degree of auto-correlation between the residuals can be continuous autoregressive models such as those developed by Pinheiro & Bates (2000). In GAMMs, smoothers reveal the relationship between the response and the predictor variables incorporating the spatial dependency (Devictor et al. 2008).

We used the R package 'mgcv' (Wood 2015) to examine the relationship between predictor variables and backscatter using GAMMs. Cubic regression spline smoothers were used to model continuous variables.

BRTs

For a full description of BRTs, see Elith et al. (2008). BRT models combine a large number of simple tree models to fit underlying relationships within the data. This differs from a more traditional regression approach (e.g. GAMs) which attempts to fit the single 'best' model. The BRT fitting process uses 2 algorithms from 2 different modelling groups: regression trees and boosting (Elith et al. 2008).

The BRT models have 2 key parameters: tree complexity (tc) and learning rate (lr). Tree complexity defines whether interactions between the predictor variables will be allowed and if so, how many. The learning rate establishes the contribution of each single tree to the entire ongoing fitting process. The parameters lr and tc define the number of trees (nt) required for achieving the most favourable prediction (Elith et al. 2008).

A cross-validation (CV) method (Mosteller & Tukey 1968) was used to find optimal model settings and to prevent overfitting. Once an optimal nt was achieved on the training samples of the data, this was used to fit models using the full dataset with the required tc and lr (Elith et al. 2008). The deviance and standard error estimated using CV were used to assess model performance, with lower values indicating better model performance (Williams et al. 2010).

The relative influence of explanatory variables represents their contribution toward explaining the response variable (Elith et al. 2008). Partial dependence plots were used to visualise the fitted functions. Predictor variables were tested to evaluate their significance using bootstrapping (Efron & Tibshirani 1993). Here, 500 bootstrap replicates were used by implementing the gbm.bootstrap function in R.

The BRTs were fitted using the 'GBM' (Ridgeway 2015) and 'dismo' packages (Hijmans et al. 2015) in R.

Spatial autocorrelation

Because high-resolution acoustic data are collected continuously along a vessel's transit, we anticipated the presence of spatial autocorrelation in our data. Not accounting for spatial autocorrelation limits hypothesis testing and predictions (Lennon 2000, Dormann 2007), and can affect standard errors and p-values of the fixed effects (Zuur et al. 2007). Moran's I autocorrelation test (Cliff & Ord 1981) was used to measure spatial autocorrelation in model residuals using the function 'moran.test' (R package 'spdep', Bivand et al. 2015), and visualised using the 'correlog' function (R package 'ncf', Bjornstad 2013). Moran's I-test index ranges from −1 to +1 (from strong negative to strong positive spatial autocorrelation), with 0 indicating no spatial autocorrelation (Cliff & Ord 1981).

To account for spatial autocorrelation, an autoregressive correlation structure of order 1 (corAR1) (Pinheiro & Bates 2000) was used in GAMMs. In the BRT models, we used an extension of an auto-logistic approach to include an auto-covariate derived from the spatial autocorrelation in the model residuals (residual autocovariate, RAC) (Crase et al. 2012).
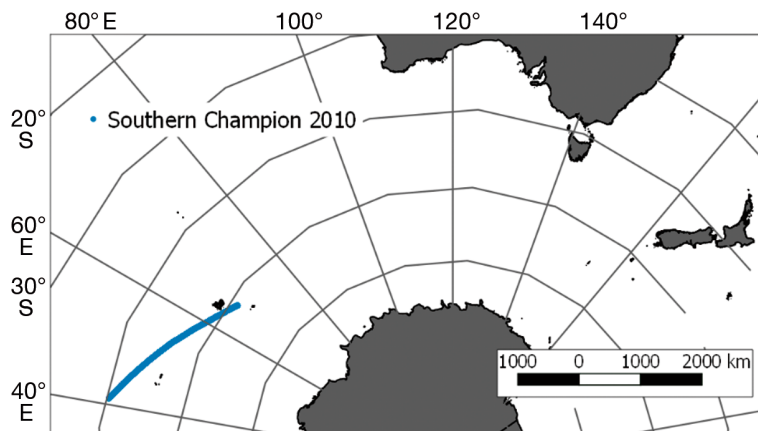
Fig. 2. Acoustic transect at 38 kHz frequency conducted by the fishing vessel 'Southern Champion' in January 2010 used for testing the predictive performance of the simplified models in an independent dataset from the Indian Ocean. The transect was retrieved from the Integrated Marine Observing System (IMOS) (https://imos.aodn.org.au/)

Variable selection

Preliminary models, fitted to the entire dataset except 1 randomly excluded transect, were used for variable selection. First, preliminary GAMM models (including all explanatory variables) were fitted, and non-significant (p < 0.5) variables were removed. Second, we used changes in Akaike's information criterion (AIC) of <10 points to remove variables from the preliminary GAMMs.

Unlike GAMMs, BRT models do not provide p-values to assess the significance of explanatory variables. Instead, we used changes in deviance explained by the model (DEBM, 2%) and 95% confidence intervals (CI) built using bootstrapping (500 samples). These methods allowed us to determine the best preliminary models for each modelling technique at each bin size and pelagic zone.

Model evaluation and selection

Model evaluation and selection were carried out on the best preliminary models by comparing R^2 values in GAMMs and percentage of DEBM in BRTs, which provide a measure of goodness of fit of the models. We aimed to find the best single model for each pelagic zone and modelling technique, assessing their predictive performance using an 'iterative' approach where each of the 28 transects was removed one at a time, while using the remaining transects as a training dataset to fit the models and to predict mean s_a in the excluded transect.

Models developed for each pelagic zone (epipelagic and mesopelagic) and modelling technique were compared, and those that achieved higher goodness of fit statistics were selected and referred to as candidate models. These models were used in the model simplification and validation steps.

Model simplification and validation

Candidate models were simplified to test whether more parsimonious models could attain similar predictive performances and are referred to as simplified models. These models were fitted including only those variables that were identified consistently by both modelling techniques as the most important for explaining mean s_a in each pelagic zone.

The validation of the candidate models was done by assessing their predictive performance using the same iterative approach described above for model evaluation and selection. The iterative validation approach proposed also enabled a test of the model sensitivity to changes in the training set provided (by iteratively excluding transects).

The iterative selection approach was also used to compare candidate and simplified models. Two statistical indicators of average model predictive performance regularly used in model evaluation studies (e.g. Zuur et al. 2009, Chai & Draxler 2014) were estimated: the root mean square error (RMSE) and the mean absolute error (MAE).

Testing model predictive power on an independent dataset

An acoustic transect conducted by the fishing vessel 'Southern Champion' at 38 kHz in January 2010 in the Indian Ocean sector of the SO retrieved from IMOS BASOOP (https://imos.aodn.org.au/) was used to test the simplified models' predictive performance on an independent transect from a different region (Fig. 2).

RESULTS

Spatial autocorrelation

Moran's I-test applied over the best preliminary models indicated that autocorrelation was lessened

in all models that included a correlation structure (Table 3). Adding a correlation structure to the GAMMs during the variable selection stage reduced the number of significant variables. Moreover, as judged by lower AIC (>10 points), these GAMMs outperformed models lacking a correlation structure. Including autocorrelation in the BRT models changed the relative contribution of variables, with some of this contribution subsequently assigned to the autocorrelation variable (RAC). BRT models which included spatial autocorrelation achieved higher percentages of DEBM.

A visual representation of the spatial autocorrelation in the residuals of GAMMs and BRT simplified models (50 km bin size) is shown by correlogram plots in Fig. 3. When accounting for spatial autocorrelation, comparatively lower and faster decaying Moran's I indices as a function of distance classes were obtained. Generally, spatial autocorrelation in the residuals was stronger at short distances (<150 km) in both pelagic zones. The spatial autocorrelation structure in the GAMM residuals for the epipelagic zone showed a transition between a positive and a negative Moran I index at ca. 150 km,

Table 3. Moran's I spatial autocorrelation indices estimated from the residuals of the best preliminary generalised additive mixed model (GAMM) and boosted regression tree (BRT) models fitted to mean acoustic backscatter (s_a) by bin size and pelagic zone when correlation structures (corAR1) and residual autocovariates (RAC) are absent and present. Δ GAMM AIC: difference in Akaike's information criterion (AIC) score, and Δ Dev. Exp.: difference in the deviance explained by the BRT models (%), showing gain in explanatory power

Zone	Bin size (km)	GAMM corAR1 absent[a] I index	GAMM corAR1 present I index	GAMM corAR1 present p	BRT RAC absent[a] I index	BRT RAC present I index	BRT RAC present p	Δ GAMM AIC	Δ Dev. Exp.
Epipelagic	10	0.08	0.008	<0.0001	0.07	−0.002	0.0160	3944	12.28
	25	0.10	0.013	<0.0001	0.07	−0.001	0.6610	1310	10.20
	50	0.10	0.013	0.0331	0.06	−0.005	0.3190	489	9.83
Mesopelagic	10	0.16	0.008	<0.0001	0.05	−0.001	0.6920	6733	12.87
	25	0.14	0.009	0.0056	0.03	−0.002	0.5600	2170	6.13
	50	0.16	0.017	0.0061	0.05	0.001	0.6850	707	7.16

[a]Moran's I index in GAMM and BRT model residuals with no correlation structure was highly significant (p < 0.0001)
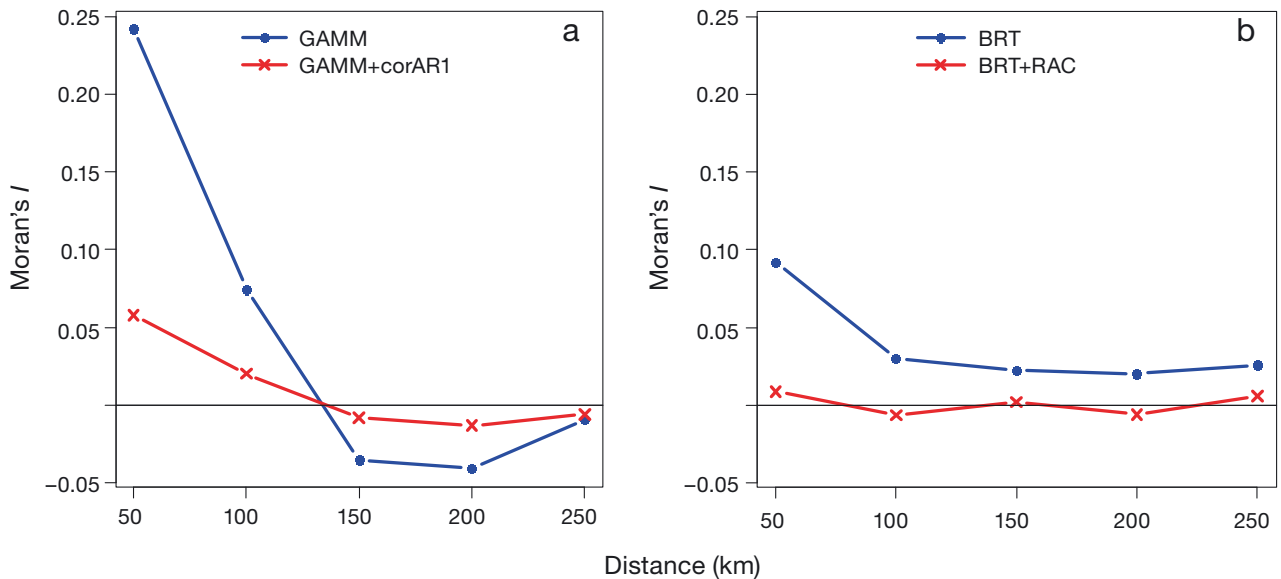

Fig. 3. Correlogram plots showing Moran's I indices of the (a) epipelagic generalised additive mixed model (GAMM) and (b) mesopelagic boosted regression tree (BRT) preliminary models' residuals fitted to mean acoustic backscatter (s_a) at 50 km bin size when correlation structures (corAR1) and residual autocovariate (RAC) were absent and present

which may be an effect of DVM by organisms into and out of the epipelagic zone. By contrast, spatial autocorrelation structure in the BRT model for the mesopelagic zone was positive, indicating that the deep scattering layers are more consistent over time and distance.

Variable selection

The number of significant variables varied according to bin size and zone amongst preliminary GAMMs. Models fitted for the epipelagic zone mean s_a at 10 km bin size had 5 significant variables (SST, depth, season, year, time of day), while the models at 25 and 50 km both had only 3 significant variables (SST, season and time of day). Mesopelagic models at 10 km had 4 variables (SST, depth, SSH and temperature at depth), and at 25 and 50 km resolution they had 5 significant variables (SST, depth, SSH, season and salinity).

Preliminary BRT models for the epipelagic zone mean s_a shared 2 variables across all bin sizes that were significant and contributed at least with a 2% of DEBM (SST and time of day), in addition to the always significant RAC. Mesopelagic mean s_a models at 10 km also had 2 variables (SST and depth), while models at 25 and 50 km had 3 variables (SST, depth and salinity) (Table 4).

The variable selection was relatively consistent between models fitted at different bin sizes, and showed that SST was always a main predictor variable for mean s_a in the epi- and mesopelagic zones (e.g. SST relative contribution in the BRT models was always >50%) (Table 4).

Model evaluation and selection

The model goodness of fit statistics obtained from the iterative model evaluation and selection approach showed that the best preliminary models fitted using coarser bin size data (25 and 50 km) performed better than models fitted at a finer bin size (10 km), with higher statistics of goodness of fit (Table 5).

Based on these results, the relative consistency of variable selection across modelling techniques and bin sizes, and practicality (computation time increased considerably at smaller bin sizes), the best preliminary models fitted at 10 km bin size for the epi- and mesopelagic zones mean s_a were dropped from further analysis.

The iterative model selection approach gave consistent results when modelling mean s_a at 25 and 50 km bin sizes in both pelagic zones using the best preliminary GAMMs and BRT models. Significance and relative contribution of variables varied only subtly, regardless of the training set (transects) used in the model fitting process. This demonstrated that the models were parsimonious and robust.

Since neither additional information nor predictive performance was gained by fitting finer-scale models, the best preliminary GAMMs and BRT models fitted at 50 km binned mean s_a were considered as performing the best for both pelagic zones.

Model simplification and validation

Two variables were consistently identified as important in all preliminary GAMMs and BRT models at all bin sizes in the epi- and mesopelagic zones: SST and time of day, and SST and depth, respectively. Consequently, these 2 pairs of variables were used for fitting simplified versions of the candidate models, referred to as the simplified models. Since the candidate BRT model for mean s_a in the epipelagic zone was already formulated using only SST and time of the day, this model was used as the simplified version.

The percentage of the dataset used for training and iteratively testing the predictive performance of the candidate and simplified models varied according to the length of transect (number of bins) excluded for testing (predicting). The mean percentage of the dataset used for training was 96.4% (min. = 93.5, and max. = 98.7).

As evaluated by the RMSE and MAE, simplified models (GAMM and BRT) fitted to mean s_a in the epi- and mesopelagic zones performed very similarly to more complex models (candidate models), and no significant differences between the means of their statistics were detected (95% CI, t-test, n = 28, p > 0.05; Table 6). These results confirmed key variables for producing reliable predictions of mean s_a in the epi- and mesopelagic zone, and highlighted the robustness of the simplified models. These models were adopted as the final models. In the case of the GAMMs, they can be expressed as

$$g(E[y_i]) = B_o + s(\text{SST}) + \text{factor(time of day)} + \varepsilon_i$$
$$\text{(epipelagic)} \quad (1)$$

$$g(E[y_i]) = B_o + s(\text{SST}) + s(\text{depth}) + \varepsilon_i$$
$$\text{(mesopelagic)} \quad (2)$$

where g corresponds to the link function (identity) of the expected response (y, mean s_a), B_o is the inter-

Table 4. Summary of preliminary generalised additive mixed model (GAMM) and boosted regression tree (BRT) model outputs and settings used for 3 bin sizes (10, 25, and 50 km; number of bins 5658, 2444 and 1301, respectively) and A) the epilagic and B) mesopelagic zone. RAC: residual autocovariate; CARS: CSIRO Atlas of Regional Seas; lr: learning rate; nt: number of trees; bf: bag fraction; tc: tree complexity; SST (SSH): sea surface temperature (height); sig: significance; rel. cont.: relative contribution; dev. expl.: deviance explained

A) Epipelagic zone

Model	10 km bin		25 km bin		50 km bin	
	Variable	Sig. level (p)	Variable	Sig. level (p)	Variable	Sig. level (p)
GAMM	s(SST)	<0.0001	s(SST)	<0.0001	s(SST)	<0.0001
	s(depth)	0.0024	f.season	<0.0001	f.season	0.0003
	f. season	<0.0001	Day/night	<0.0001	Day/night	<0.0001
	f.year	0.0005				
	Day/night	<0.0001				
	R^2	0.59	R^2	0.6	R^2	0.62
	Variable/parameter	Rel. cont. (%)	Variable/parameter	Rel. cont. (%)	Variable/parameter	Rel. cont. (%)
BRT	SST	54.29	SST	68.16	SST	73.06
	RAC	19.21	RAC	24.6	RAC	18.48
	Day/night	5.57	Day/night	7.24	Day/night	8.46
	lr	0.08	lr	0.03	lr	0.01
	nt	2750	nt	2250	nt	1800
	bf	0.5	bf	0.5	bf	0.5
	tc	5	tc	5	tc	5
	Dev. expl. (%)	75.6	Dev. expl. (%)	72.2	Dev. expl. (%)	73.3

B) Mesopelagic zone

Model	10 km bin		25 km bin		50 km bin	
	Variable	Sig. level (p)	Variable	Sig. level (p)	Variable	Sig. level (p)
GAMM	s(depth)	<0.0001	s(depth)	<0.0001	s(SST)	<0.0001
	s(SST)	<0.0001	s(SST)	<0.0001	s(depth)	<0.0001
	s(SSH)	<0.0001	s(SSH)	<0.0001	s(CARS salinity)	0.0004
	s(CARS salinity)	<0.0001	s(CARS salinity)	0.0002	s(SSH)	0.0003
			f.season	<0.0001	f.season	0.0002
	R^2	0.56	R^2	0.61	R^2	0.65
	Variable/parameter	Rel. cont. (%)	Variable/parameter	Rel. cont. (%)	Variable/parameter	Rel. cont. (%)
BRT	SST	51.48	SST	51.2	SST	54.26
	RAC	28.62	Depth	21.82	Depth	20.91
	Day/night	19.9	RAC	13.55	RAC	14.37
			CARS temp.	13.43	CARS temp.	10.45
	lr	0.1	lr	0.04	lr	0.02
	nt	2200	nt	1950	nt	2150
	bf	0.5	bf	0.5	bf	0.5
	tc	5	tc	5	tc	5
	Dev. expl. (%)	76.59	Dev. expl. (%)	79	Dev. expl. (%)	73.37

cept, s(SST) and s(depth) are the smooth functions of SST and depth, factor(time of day) is the corresponding categorical covariate, and ε_i represents the model residuals.

BRTs do not provide a single final model since this technique fits many simple models, which are combined for prediction (Elith et al. 2008).

In addition, GAMM smoothers and BRT partial dependence functions fitted to mean s_a in the epi- and mesopelagic zones across preliminary, candidate and simplified models showed only minor differences, conserving similar shapes and trends (Fig. 4). These functions for the epipelagic zone indicated an increase in mean s_a with warmer SSTs in the range of observation (approx. −2 to +19°C), and higher levels at night; and a similar relationship between SST and mean s_a in the mesopelagic zone, with the addition of a negative depth effect on mean s_a in shallow areas

Table 5. Mean of the generalised additive mixed model (GAMM) and boosted regression tree (BRT) models' goodness of fit statistics by bin size and pelagic zone, obtained through an iterative process where each transect (n = 28) was excluded one at a time

Bin size (km)	GAMM — R^2		BRT — deviance explained by the model (%)	
	Epipelagic	Mesopelagic	Epipelagic	Mesopelagic
10	0.59	0.57	72.9	76.1
25	0.6	0.61	71.7	78.8
50	0.62	0.64	73.5	78.4

(bottom depth <1000 m). A bottom depth effect on mesopelagic mean s_a occurs when transects cross over the continental shelf or oceanic plateaus south and southeast of NZ, where the mesopelagic zone is confined to a reduced range in comparison to open-ocean areas (Escobar-Flores et al. 2018, this Volume).

These results were in accord with backscatter-SST Pearson's correlation tests, which always showed strong highly significant positive correlations in the epi- and mesopelagic zones at all bin sizes used for modelling. For example, Pearson's correlation coefficient between backscatter and SST for the epi- and mesopelagic zones was 0.62 and 0.92 (n = 5658, p < 0.001). The relationship between backscatter and SST was also foreseeable, since both variables showed a south–north increasing pattern (Fig. 5).

Both modelling techniques predicted mean s_a more accurately in the epipelagic (lower RMSE and MAE) than in the mesopelagic zone (Table 6). The final GAMM performed better than the final BRT model when predicting mean s_a in the epipelagic zone for transect SAO 2010 NZ-SO used in the non-iterative model validation process (GAMM MAE = 0.955; BRT MAE = 0.972), but worse when predicting mean s_a in

the mesopelagic zone (BRT RMSE = 3.31 and MAE = 2.53; GAMM RMSE = 4.17 and MAE = 3.39).

Fig. 6 illustrates a comparison of the observed mean s_a along transect SAO 2010 NZ-SO and the predictions from the final GAMM and BRT models for the epi- and mesopelagic zones, respectively. Predictions for both pelagic zones were of the same order of magnitude. Epipelagic predictions of mean s_a followed the north–south decreasing trend in observed values reasonably well, but showed differences in the peaks and inter-bin variability. Time of day effect was evident by the periodic changes in both predicted and observed mean s_a. Mesopelagic predictions displayed the overall north–south decreasing trend and drastic changes in mean s_a around 55° S observed in the data, but small-scale variability along the transect was not modelled as accurately.

Validation from the independent dataset

Final models performed relatively well when tested on the independent dataset providing moderate measurements of predictive performance (Table 7, Fig. 7). The predicted values for the 'Southern Champion' transect were of the same order of magnitude as those observed, except at the northernmost end of the transect in the mesopelagic zone, where observed mean s_a increased rapidly (Fig. 7). This transect extended further north (40° S) than any of the training data from the Pacific sector. Both types of models had better predictive performance in the epi-

Table 6. Candidate versus simplified (final) model predictive performance statistics (root mean square error [RMSE] and mean absolute error [MAE]), and mean R^2 (generalised additive mixed model, GAMM) and percentage of deviance explained by the model (boosted regression tree, BRT) obtained from the iterative validation process. Note that no values are given for the simplified BRT model in the epipelagic zone, as this was the same as the candidate model

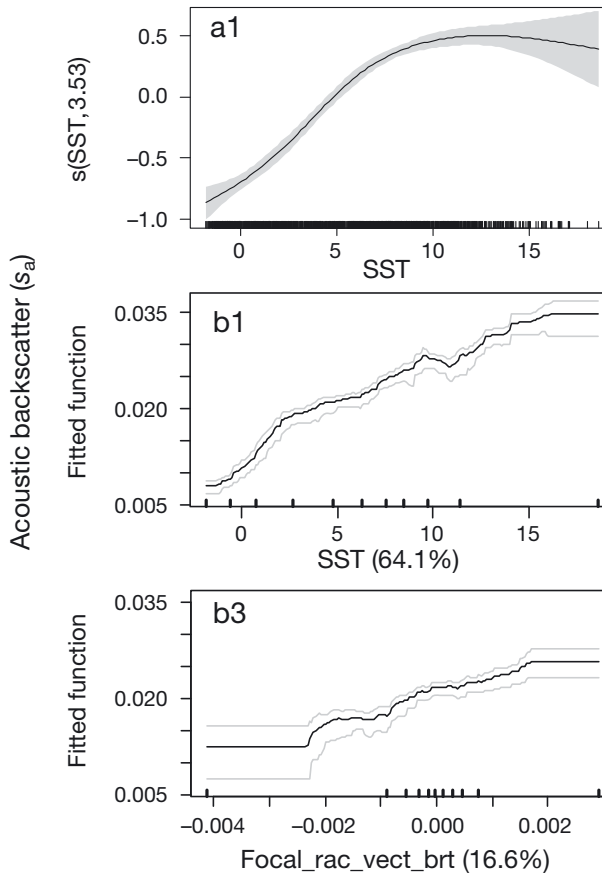| Statistic	Zone	Candidate models (50 km bin size)		Simplified models		t-test p	
		GAMM	BRT	GAMM	BRT	GAMM	BRT
RMSE	Epipelagic	3.85	3.66	3.89	–	0.94	–
	Mesopelagic	6.89	6.58	6.89	6.40	1.00	0.78
MAE	Epipelagic	2.13	2.03	2.13	–	1.00	–
	Mesopelagic	4.59	4.65	4.46	4.71	0.75	0.89
GAMM R^2	Epipelagic	0.643	–	0.616	–	–	–
	Mesopelagic	0.624	–	0.625	–	–	–
BRT deviance explained (%)	Epipelagic	–	73.5	–	–	–	–
	Mesopelagic	–	78.5	–	75.5	–	–

Fig. 4. Relationship between the $\log_{10}$-transformed mean acoustic backscatter (s_a) and (a1) sea surface temperature (SST), and (a2) the estimated effect (solid lines) of the factor levels day/night ('day' = reference level) on s_a, obtained by the simplified (final) generalised additive mixed model (GAMM) in the epipelagic zone. Partial dependency plots of the final boosted regression tree (BRT) model in the mesopelagic zone, show the relationship between the cubic-transformed mean s_a and (b1) SST and (b2) depth, and (b3) the residual autocorrelation covariate (Focal_rac_vect_brt) (used in the model for accounting for spatial autocorrelation). Percentage in brackets in the partial dependency plots show the variable relative contribution to the model. Shading, dashed lines (a1, a2; GAMM) and grey lines (b1–3; BRT) represent 95% confidence intervals. Thin tick marks on the *x*-axis in (a1) refer to observed data points (GAMM) and thick tick marks in (b1–3) refer to deciles of the predictor variable (BRT)
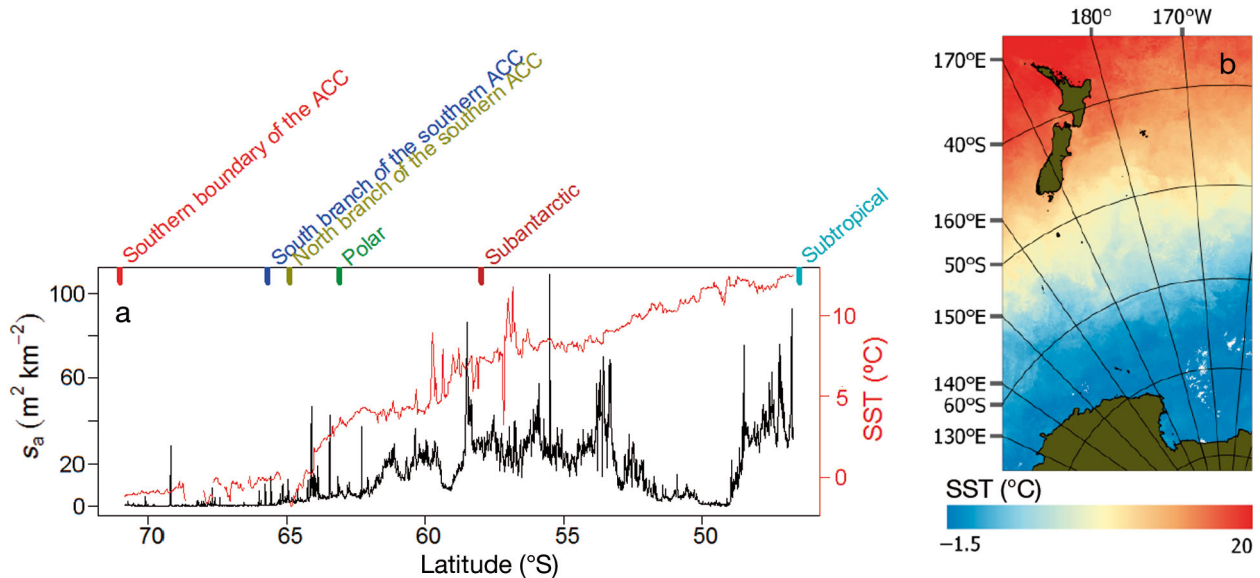

Fig. 5. (a) Mean acoustic backscatter (s_a, m^2 km^{-2}) and satellite-derived sea surface temperature (SST, °C) along a transect conducted by the vessel 'San Aotea II' (December 2010) across the Southern Ocean (SO). Indicated at the top of the panel are front positions as defined by Orsi et al. (1995) and Sokolov & Rintoul (2009) (ACC: Antarctic Circumpolar Current). (b) Monthly composite of SST shows a latitudinal gradient in the Pacific sector of the SO in December 2010
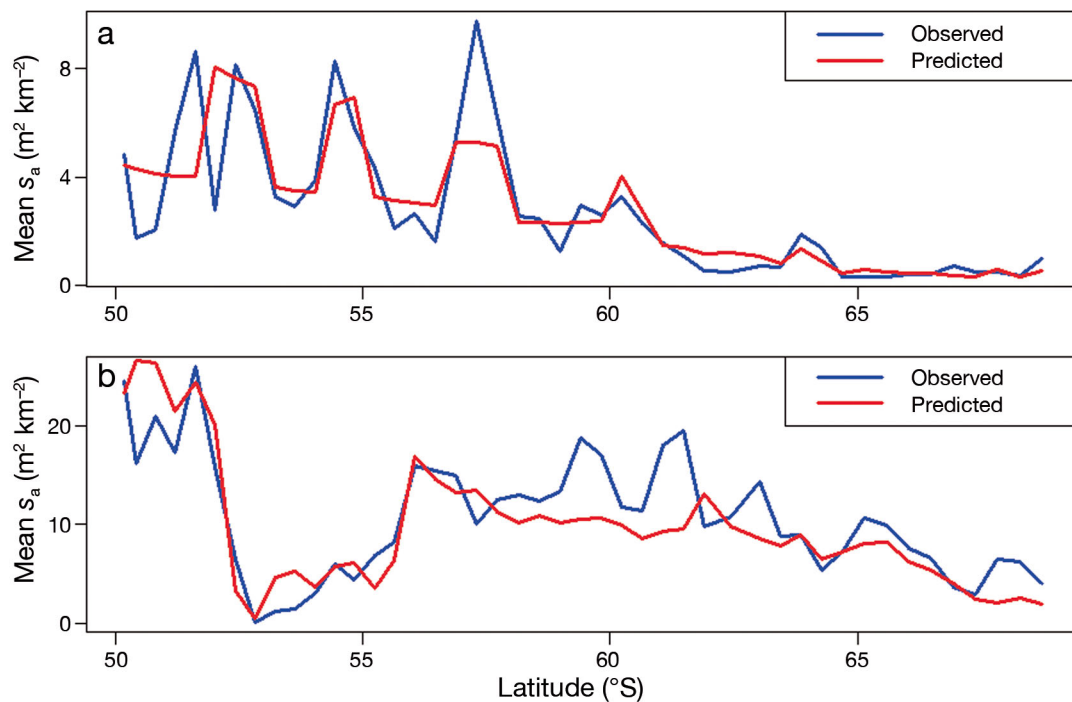
Fig. 6. Observed and predicted mean acoustic backscatter (s_a, m^2 km^{-2}) in the (a) epipelagic and (b) mesopelagic zones using the final generalised additive mixed model (GAMM) and boosted regression tree (BRT) models, respectively, for the transect conducted by the vessel 'San Aotea II' in 2010 across the Southern Ocean

Table 7. Final models' predictive performance statistics, root mean square error (RMSE) and mean absolute error (MAE), tested on an independent dataset in the Indian sector of the Southern Ocean. RMSE and MAE obtained from the final models' iterative validation approach (Pacific sector of the SO) were used to create 95% confidence intervals (CI) using Student's t-distribution for comparison. GAMM: generalised additive mixed model, BRT: boosted regression tree

| Statistic | Zone | Final models (Indian Ocean) | | 95% CI statistics (Pacific Ocean) | | | |
| | | | | GAMM | | BRT | |
		GAMM	BRT	Lower CI	Upper CI	Lower CI	Upper CI
RMSE	Epipelagic	1.86	2.43	3	4.79	2.82	4.52
	Mesopelagic	22.90	22.30	5.84	7.94	5.45	7.33
MAE	Epipelagic	1.46	1.87	1.69	2.56	1.65	2.42
	Mesopelagic	19.00	10.50	4.05	5.38	3.88	5.03

pelagic zone than in the mesopelagic zone (Table 7, Fig. 7). While the GAMM accomplished slightly better outcomes when predicting epipelagic mean s_a (lower RMSE and MAE), the BRT model obtained better predictions in the mesopelagic zone.

DISCUSSION

There is an ongoing demand for information on MTLOs for ecosystem model validation and parameterisation. The demand has been acknowledged, and acoustic data collection, gathering and processing endeavours are already underway. However, due to

the vast extent of the open ocean and the scarcity of observations, the development of models for explaining and predicting the abundance of MTLOs provides a valuable tool for filling current data gaps. This study demonstrated that opportunistically collected acoustic data and readily accessible explanatory variables can be used to successfully model and predict acoustic backscatter, s_a, a proxy for MTLO abundance within the observed order of magnitude in the epi- and mesopelagic zones in the Pacific and Indian sector of the SO. Although the models developed for the epipelagic zone represented general patterns well in both geographical regions of the SO tested, models for the mesopelagic zone did not per-
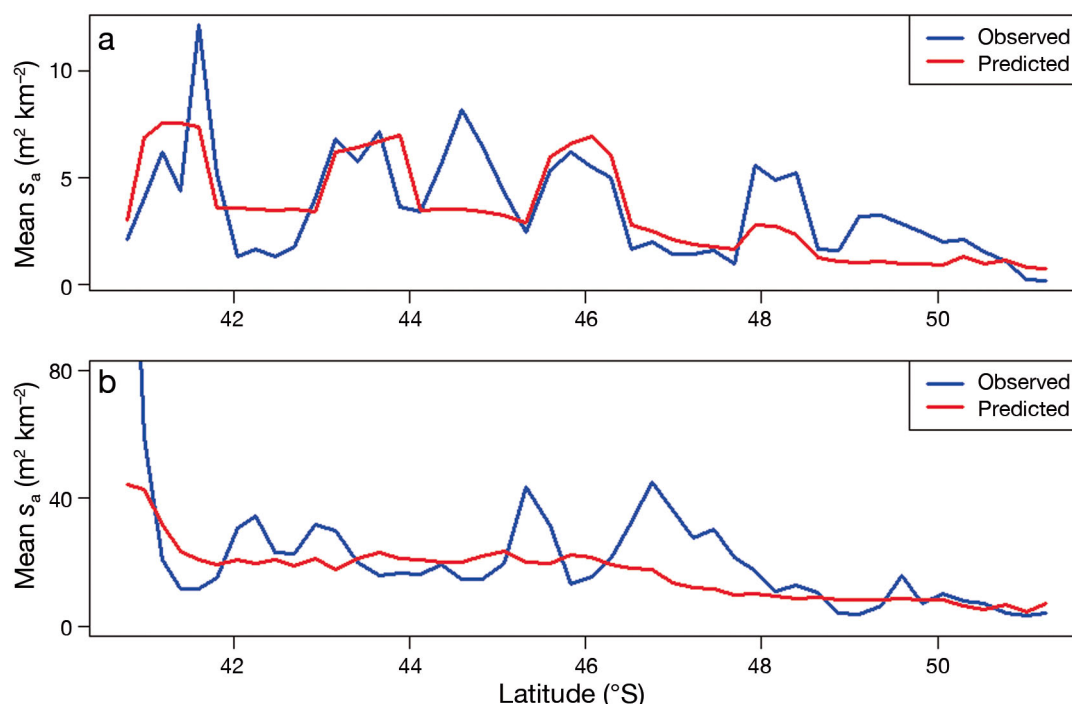
Fig. 7. Observed and predicted mean acoustic backscatter (s_a, m^2 km^{-2}) in the (a) epipelagic and (b) mesopelagic zones using final generalised additive mixed model (GAMM) and boosted regression tree (BRT) models, respectively, for an independent transect from the Indian sector of the Southern Ocean

form as well when applied to the Indian sector of the SO. The key finding was that SST and time of day were the main predictor variables for backscatter in the epipelagic zone, and SST and depth were the main predictors in the mesopelagic zone.

Role of temperature and depth

Epi- and mesopelagic final models included explanatory variables with latitudinal gradients: SST and depth. Physical and geographical variables have been used previously in ecological modelling studies to define ecosystems in the SO. For example, using clustering algorithms, Grant et al. (2006) partitioned the SO into bioregions, and Koubbi et al. (2011a) used dissimilarity models to define ecoregions of myctophids in the Indian Ocean sector of the SO. These studies depicted ecological spatial subdivisions following bands of latitudinal environmental gradients (e.g. SST) and geographical features (e.g. deep ocean, continental shelves and slopes) or chl *a* concentrations. Likewise, environmental variables from climatologies and satellites as well as depth and latitude were used to define myctophid zoogeography regions off eastern Australia (Flynn & Marshall 2013), to predict the distribution of 8 species of myctophids in the SO based

on presence/absence data from net catches using BRT (Duhamel et al. 2014), and to define potential habitats of *Electrona antarctica* (Myctophidae) based on abundance data around Kerguelen Island in the Indian Ocean (Loots et al. 2007). Other studies have correlated temperature with larval abundance of *E. antarctica*, *Notolepis coatsi* and the squid *Galiteuthis glacialis*, in the pelagic zone of the Cosmonaut Sea and Prydz Bay (Van de Putte et al. 2010), and identified SST and depth, in addition to krill, as the main factors driving *E. antarctica* postlarval density in the top 200 m in the Lazarev Sea (8° W–2° E) between 61 and 71° S (Flores et al. 2008). The only study exploring potential relationships between acoustic data for mesopelagic fish and environmental variables in the SO (Scotia Sea) also concluded that water temperature was positively correlated to backscatter in the top 200 m, and consequently could be used to determine its distribution (Fielding et al. 2012). Though only 1 of the studies mentioned above used acoustic data as a proxy for MTLO distribution and/or abundance, all underscore the usefulness of SST and depth in developing explanatory and predictive models for mesopelagic fish, in congruence with our models.

Temperature is among the most important abiotic factors for living organisms, affecting physiological functions and processes such as development, growth,

digestion and movement (Nikolsky 1963, Reynolds & Casterlin 1980, Withers 1992, Clarke 2009). Although temperature can be an important environmental factor for fish, as observed in the growth of young cod in the North Sea (Rindorf et al. 2008), it does not solely determine fish distributions (Johnson 1974, 1982, Barnett 1984). Despite the potential effects of temperature on marine organisms and their distributions, we considered that the relationship between SST and backscatter described in the models is correlational rather than functional. Existing bioregionalisation studies (e.g. Grant et al. 2006) in the SO have defined regions using temperature gradients, and they could incorporate backscatter as a new biological parameter for characterising the productivity of these bioregions.

Climate change effects in the SO marine environment through ocean warming might alter the relationships established in our models between SST and backscatter; however, if the structure of the mesopelagic communities remains stable, the use of spatially explicit variables could account for temporal variations through a 'space-for-time' substitution (Blois et al. 2013).

Models developed for the mesopelagic zone included depth as 1 of the 2 main explanatory variables. The depth effect relates to a decrease in s_a when bottom depth declines in the northern end of the transects as these approach the south and southeast side of the NZ continental shelf (Escobar-Flores et al. 2018). As expected in shallower waters, the 'mesopelagic' habitat is substantially reduced. As depth increases from neritic to open-ocean areas, more acoustic layers or functional groups, as described by Lehodey et al. (2010), become available for acoustic detection, e.g. migrant and non-migrant mesopelagic and bathypelagic. Reduction in s_a in shallow areas also may co-occur with changes in species composition (e.g. Koubbi et al. 2011a).

Effect of time of day in the epipelagic models

Time of day was consistently identified as a significant variable for explaining mean s_a in the epipelagic models. Mesopelagic fish perform extensive DVM between bathypelagic (>1000 m) and mesopelagic zones (200–1000 m) where they distribute during the day, and the epipelagic zone (<200 m) which they ascend to around sunset (Salvanes & Kristoffersen 2001). This behaviour, usually considered a strategy for feeding on zooplankton in the upper 200 m or for predator avoidance (Catul et al. 2011), is well docu-

mented in acoustics studies, where continuous sampling allows capture of the whole diel cycle over 24 h (Escobar-Flores et al. 2018). In terms of mean s_a, DVM is characterised by a substantial increase in the epipelagic zone at night, which contrasts with the rather subtle changes in the mesopelagic zone (Escobar-Flores et al. 2013, Béhagle et al. 2016), likely due to organisms migrating from deeper zones or changes in scattering properties (Escobar-Flores et al. 2018). Consequently, it seems reasonable that our epipelagic models account for the day/night differences in mean s_a.

Primary productivity and mesopelagic fish

Primary productivity (PP) is a function of phytoplankton biomass and growth rates. Although relationships between PP or chl a and backscatter have been described in the literature (e.g. Escobar-Flores et al. 2013, Irigoien et al. 2014), in our models we did not find a relationship. Chl a is patchily distributed in the SO, with major concentrations found in association with coastal areas, fronts and the marginal ice zone (Sakshaug & Holm-Hansen 1984, Hempel 1985, Smith & Sakshaug 1990). These conditions differ from typical oligotrophic conditions in open waters where our dataset was mainly collected, and therefore a lack of correlation between these variables seems reasonable. Other modelling studies in the SO are in agreement with our results and have not found significant relationships between chl a or PP for defining the zoogeography (Flynn & Marshall 2013) and determining community composition of lanternfish (Koubbi et al. 2011a).

Model resolution

Environmental variables selected for modelling need to be considered at the scale at which they affect species distributions (UNESCO 2009). We binned backscatter at 3 resolutions, the coarsest of which provided the best outputs. Although using a coarser resolution may lead to the loss of information, it reduces variability by averaging, allowing us to capture the most relevant variables driving the process or phenomenon under investigation. Using data at finer resolutions did not provide additional gains in explanatory power in the models or alter the relationship between the response and the explanatory variables. In fact, high-resolution models may add noise due to small-scale variability. Fulton (2010)

stated that in end-to-end ecosystem modelling, high resolution and complexity may not be the most useful approach, and advocated a constrained balance between performance and complexity. For example, backscatter at 50 km resolution to complement prey field studies of apex predators can still be considered fine enough, because it can describe spatial ecological processes, such as DVM, and can characterise biologically relevant oceanographic features, such as mesoscale eddies and meanders at spatial scales between 50 and 500 km (Haury et al. 1978, Morrow & Le Traon 2006); these features can drive mid-trophic level distribution (Godø et al. 2012, Béhagle et al. 2014). Our model resolution is also within the range used for predicting micronekton in other recent studies (1° grid, Lehodey et al. 2010; 36 km grid, Jennings et al. 2008).

Acoustic backscatter as a proxy for MTLOs

Here, acoustic backscatter was assumed to be a proxy for MTLO abundance. However, the lack of biological sampling to verify the taxonomic composition ('ground truthing') is a major limitation and source of uncertainty in this study. This is a common weakness of acoustically derived indices from vessels of opportunity (Handegard et al. 2013). We assumed that mesopelagic fishes were the main group responsible for s_a at 38 kHz. Most of the backscatter came from constant or intermittent sound scattering layers distributed throughout the water column, and a dense continuous deep scattering layer is consistently described as being dominated by mesopelagic fishes, particularly myctophids (e.g. Duhamel et al. 2000). However, other taxa may also contribute to acoustic backscatter at this frequency, particularly Antarctic krill (e.g. Hewitt et al. 2003) and siphonophores (e.g. Lavery et al. 2007). Antarctic krill contribution to total acoustic backscatter along some of the acoustic transects used in this research collected with multi-frequency systems (RV 'Tangaroa') was assessed by Escobar-Flores (2017). Using the difference between the mean volume backscattering strength measured at 120 and 38 kHz (2−16 dB window) (Watkins & Brierley 2002, Demer 2004), their contribution only represented between 0.6 and 1.4%. The 2−16 dB frequency difference window may also include fish species without a swimbladder (e.g. Antarctic silverfish, O'Driscoll et al. 2011), so the actual krill composition could be even lower.

Resonance scattering at various depths is a well described phenomenon in mesopelagic organisms (e.g. Kloser et al. 2016), which might positively bias estimates of biological density obtained from acoustic backscatter collected at 38 kHz (Davison et al. 2015). Changes in depth (i.e. pressure) experienced by gas-bearing organisms during DVM, may affect their swimbladder volume or the pressure of the gas contained in the swimbladder, causing them to resonate at the operating frequency (Godø et al. 2009). Although resonance for mesopelagic species is thought to occur at frequencies <38 kHz (e.g. Kloser et al. 2002, Godø et al. 2009), individuals of small size (<4 cm) or showing swimbladder atrophy in adult stages might still resonate at greater depths at this frequency, complicating the interpretation of backscatter as a proxy for mesopelagic fish biological density (Kloser et al. 2016).

Gelatinous zooplankton such as siphonophores, which are poorly sampled by trawls, can also contribute significantly in some regions to the acoustic energy measured at 38 kHz due to resonance scattering (Lavery et al. 2007, Davison et al. 2015, Kloser et al. 2016). Siphonophores are common organisms in the epi- and mesopelagic zones of high latitudes (Mapstone 2014). When these organisms are regionally dominant, a positive bias could be expected when backscatter is attributed only to fish (Davison et al. 2015). Consequently, we acknowledge that some of the backscatter assumed as a proxy for mesopelagic fish might correspond to gelatinous zooplankton. Uncertainties due to resonance scattering at 38 kHz from siphonophores as well as for some mesopelagic fish with a swimbladder at various depths can preclude reliable estimates of biological indices from bulk backscatter collected over large depth ranges (Handegard et al. 2013, Davison et al. 2015, Kloser et al. 2016).

More research is needed using broadband acoustics and scattering models to improve our understanding of scattering properties of gas-bearing organisms at different depths. Likewise, more biological sampling (i.e. nets and optics) will reduce the uncertainty regarding the composition of the micronektonic community and the potential contribution of different groups of organisms to s_a.

Mesopelagic fish community composition effect on backscatter

Patterns in acoustic backscatter may be due to changes in species composition of mesopelagic fish as well as abundance. Species composition affects backscatter levels as the acoustic target strength

varies between and within species (e.g. ontogenetic changes), particularly for myctophids, which display a variety of swimbladder shapes and content (Marshall 1960, Foote 2001, Davison 2011). Latitudinal changes in species composition or assemblages could indirectly explain backscatter correlation with SST. This hypothesis is worth considering, since the biogeography of lanternfishes proposed by McGinnis (1982) and ecoregions defined by Koubbi et al. (2011a) depict different lanternfish species assemblages organised in latitudinal bands. Species diversity decreases towards the southern regions of the SO (McGinnis 1982, Koubbi et al. 2011b), and the few species inhabiting these waters are endemic to high latitudes (Andriashev 1965, Gon & Heemstra 1990). These species mainly belong to the genera *Gymnoscopelus* and *Lampanyctus*, whose swimbladders are atrophied or absent as adults (Marshall 1960). A few species with swimbladders, e.g. *Krefftichthys anderssoni* and *E. antarctica* (whose adults display reduced swimbladders) (Marshall 1960), also belong to the southern groups. Conversely, towards lower latitudes there seems to be a gradual transition from less to more species-rich assemblages (McGinnis 1982, Macpherson 2002, Koubbi et al. 2011a), dominated by myctophids with swimbladders (e.g. *Protomyctophum normani*, *Diaphus danae* and *E. subaspera*), which will contribute more in terms of backscatter than their southern counterparts. This concurs with a north–south negative significant relationship between latitude and backscatter observed in the Pacific sector of the SO, although backscatter does not necessarily correlate with biological density (Escobar-Flores et al. 2018). Thus, the backscatter latitudinal pattern observed in this region could be related to abundance as well as latitudinal changes in species composition as shown in bioregionalisation studies and available trawl information (McGinnis 1982, Koubbi et al. 2011a, Escobar-Flores et al. 2018).

Ecological significance of the models

Backscatter can be converted into biological density using acoustic target strength and information from trawl samples about species composition and size distribution. When information from trawls is not available, backscatter may still be useful as a proxy for MTLO abundance (e.g. Irigoien et al. 2014). This information can be used to parameterise and validate ecosystem models such as the spatially explicit SEAPODYM model (Lehodey et al. 2010), by providing reference values for evaluating their per-

formance for predicting micronekton biomass. Day–night backscatter differences can also be used in the SEAPODYM model, as the model is vertically resolved (Handegard et al. 2013). Other models such as the ATLANTIS framework (Fulton et al. 2005), which is spatially and vertically structured, could benefit from the inclusion of backscatter indices. Likewise, the open-ocean ecosystem model APESCOSM (size-based) (Maury 2010) can use backscatter predictions and differences in vertical distribution to validate 3D spatially explicit (latitude, longitude and depth) predictions of open-ocean pelagic community biomass, reducing uncertainty in apex predator models.

Our models can also be used in ecological studies by characterising prey fields of apex predators in the SO. The increase in tracking studies of top predators has improved our capacity to identify the location of feeding grounds (Bost et al. 2009, Thiers et al. 2017). Combining predator tracking information and backscatter predictions will provide insights into the trophic conditions of preference (in terms of acoustic energy), and potentially help us to classify different predator species based on these preferences. Exploring these relationships could help to further validate the models as a tool for providing information on prey availability. Coupling mid-trophic level modelling studies and the dynamics of apex predators that rely heavily on mesopelagic fish as prey (e.g. Antarctic fur seals and king penguins) (Lea et al. 2002, Cherel et al. 2007) can be mutually beneficial to further validate the models, and to improve our understanding of predator trophic dynamics and behaviour, and spatial distribution.

Conclusions and future directions

Bearing in mind that 'nature is too complex and heterogeneous to be predicted accurately in every aspect of time and space from a single, although complex, model' (Guisan & Zimmermann 2000, p. 150), our study has shown that it is possible to develop models to predict mean s_a in the epi- and mesopelagic zones reasonably well, using opportunistically collected acoustic data and readily available satellite-derived oceanographic and bathymetry data. Our models predicted acoustic backscatter as a proxy for MTLOs within the observed order of magnitude in the epi- and mesopelagic zones in the Pacific and Indian sectors of the SO.

Programs like IMOS, MESOPP and the Southern Ocean Network of Acoustics will channel more acoustic data from existing and new data collection

initiatives. Future work on acoustic data collection and modelling of backscatter as a proxy for MTLO in the SO should concentrate on expanding the geographical extent of the existing training dataset, since further testing and validation across other sectors of the SO is still needed. At this stage, the use of the models outside the latitudinal geographical range of the training dataset is discouraged.

Although no temporal or seasonal variable other than time of day was identified as significant, the predictive power of the models in time is uncertain. The current and future spatial distribution of environmental gradients might differ (Guisan & Thuiller 2005); consequently, as the training dataset expands, is it important to examine the appearance of temporal-related changes over larger scales.

Our models were based on available explanatory variables derived from satellite or models with varied temporal and spatial resolutions. Although this can be seen as a limitation, we were still able to successfully capture useful relationships to predict acoustic backscatter reasonably well. However, improvements to our models may come with advances in the spatio-temporal resolution and coverage of satellite-derived variables, such as SST, in the SO.

LITERATURE CITED

Andriashev AP (1965) A general review of the Antarctic fish fauna. In: van Oye P, van Mieghem J (eds) Biogeography and ecology in Antarctica. Junk Publishers, The Hague, p 491–550

Barham EG (1966) Deep scattering layer migration and composition: observations from a diving saucer. Science 151: 1399–1403

Barnett MA (1984) Mesopelagic fish zoogeography in the central tropical and subtropical Pacific Ocean: species composition and structure at representative locations in three ecosystems. Mar Biol 82:199–208

Beamish RJ, Leask KD, Ivanov OA, Balanov AA, Orlov AM, Sinclair B (1999) The ecology, distribution, and abundance of midwater fishes of the Subarctic Pacific gyres. Prog Oceanogr 43:399–442

Béhagle N, Du Buisson L, Josse E, Lebourges-Dhaussy A, Roudaut G, Ménard F (2014) Mesoscale features and micronekton in the Mozambique Channel: an acoustic approach. Deep-Sea Res II 100:164–173

Béhagle N, Cotté C, Ryan TE, Gauthier O and others (2016) Acoustic micronektonic distribution is structured by macroscale oceanographic processes across 20–50° S latitudes in the South-Western Indian Ocean. Deep-Sea Res I 110:20–32

Bertrand A, Josse E, Bach P, Dagorn L (2003) Acoustics for ecosystem research: lessons and perspectives from a scientific programme focusing on tuna-environment relationships. Aquat Living Resour 16:197–203

Bianchi D, Galbraith ED, Carozza DA, Mislan KAS, Stock CA (2013) Intensification of open-ocean oxygen depletion by vertically migrating animals. Nat Geosci 6: 545–548

Bivand R, Lewin-Koh N (2015) maptools: tools for reading and handling spatial objects. R package version 0.8-34. https://CRAN.R-project.org/package=maptools

Bivand RS, Pebesma EJ, Gómez-Rubio V (2013) Applied spatial data analysis with R, 2nd edn. Springer, New York, NY

Bivand R, Altman M, Anselin L, Assunção R and others (2015) spdep: spatial dependence: weighting schemes, statistics and models. R package version 0.5-88. https://CRAN.R-project.org/package=spdep

Bjornstad ON (2013) ncf: spatial nonparametric covariance functions. R package version 1.1-5. https://CRAN.R-project.org/package=ncf

Blois JL, Williams JW, Fitzpatrick MC, Jackson ST, Ferrier S (2013) Space can substitute for time in predicting climate-change effects on biodiversity. Proc Natl Acad Sci USA 110:9374–9379

Bost CA, Cotté C, Bailleul F, Cherel Y and others (2009) The importance of oceanographic fronts to marine birds and mammals of the southern oceans. J Mar Syst 78:363–376

Catul V, Gauns M, Karuppasamy PK (2011) A review on mesopelagic fishes belonging to family Myctophidae. Rev Fish Biol Fish 21:339–354

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 7:1247–1250

Cherel Y, Hobson KA, Guinet C, Vanpe C (2007) Stable isotopes document seasonal changes in trophic niches and winter foraging individual specialization in diving predators from the Southern Ocean. J Anim Ecol 76:826–836

Clarke A (2009) Temperature and marine macroecology. In: Witman JD, Roy K (eds) Marine macroecology. The University of Chicago Press, Chicago, IL, p 251–278

Cliff AD, Ord JK (1981) Spatial processes: models and applications. Pion Limited, London

Collins MA, Stowasser G, Fielding S, Shreeve R and others (2012) Latitudinal and bathymetric patterns in the distribution and abundance of mesopelagic fish in the Scotia Sea. Deep-Sea Res II 59–60:189–198

Constable AJ, Melbourne-Thomas J, Corney SP, Arrigo KR and others (2014) Climate change and Southern Ocean ecosystems I: How changes in physical habitats directly affect marine biota. Glob Change Biol 20:3004–3025

Constable AJ, Costa DP, Schofield O, Newman L and others (2016) Developing priority variables ('ecosystem Essential Ocean Variables' — eEOVs) for observing dynamics and change in Southern Ocean ecosystems. J Mar Syst 161:26–41

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35:879–888

Davison P (2011) The specific gravity of mesopelagic fish from the northeastern Pacific Ocean and its implications for acoustic backscatter. ICES J Mar Sci 68:2064–2074

Davison PC, Checkley DM, Koslow JA, Barlow J (2013) Carbon export mediated by mesopelagic fishes in the northeast Pacific Ocean. Prog Oceanogr 116:14–30

Davison PC, Koslow JA, Kloser RJ (2015) Acoustic biomass estimation of mesopelagic fish: backscattering from individuals, populations, and communities. ICES J Mar Sci 72:1413–1424

Demer DA (2004) An estimate of error for the CCAMLR 2000 survey estimate of krill biomass. Deep-Sea Res II 51: 1237–1251

Demer DA, Berger L, Bernasconi M, Bethke E and others (2015) Calibration of acoustic instruments. Coop Res Rep 326. ICES, Copenhagen

Devictor V, Julliard R, Clavel J, Jiguet F, Lee A, Couvet D (2008) Functional biotic homogenization of bird communities in disturbed landscapes. Glob Ecol Biogeogr 17: 252–261

Dormann CF (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob Ecol Biogeogr 16:129–138

Duhamel G, Koubbi P, Ravier C (2000) Day and night mesopelagic fish assemblages off the Kerguelen Islands (Southern Ocean). Polar Biol 23:106–112

Duhamel G, Hulley PA, Causse R, Koubbi P and others (2014) Biogeographic patterns of fish. In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B and others (eds) Biogeographic atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge, p 328–362

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York, NY

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813

Escobar-Flores PC (2017) The use of acoustics to characterise mid-trophic levels of the Southern Ocean pelagic ecosystem. PhD thesis, University of Auckland

Escobar-Flores P, O'Driscoll RL, Montgomery JC (2013) Acoustic characterization of pelagic fish distribution across the South Pacific Ocean. Mar Ecol Prog Ser 490: 169–183

Escobar-Flores PC, O'Driscoll RL, Montgomery JC (2018) Spatial and temporal distribution patterns of acoustic backscatter in the New Zealand sector of the Southern Ocean. Mar Ecol Prog Ser 592:19–35

Fielding S, Watkins JL, Collins MA, Enderlein P, Venables HJ (2012) Acoustic determination of the distribution of fish and krill across the Scotia Sea in spring 2006, summer 2008 and autumn 2009. Deep-Sea Res II 59–60: 173–188

Flores H, Van de Putte AP, Siegel V, Pakhomov EA, van Franeker JA, Meesters EHWG, Volckaert FAM (2008) Distribution, abundance and ecological relevance of pelagic fishes in the Lazarev Sea, Southern Ocean. Mar Ecol Prog Ser 367:271–282

Flynn AJ, Marshall NJ (2013) Lanternfish (Myctophidae) zoogeography off eastern Australia: a comparison with physicochemical biogeography. PLOS ONE 8:e80950

Foote KG (2001) Acoustic scattering by marine organisms. In: Steele JH, Turekian KK, Thorpe SA (eds) Encyclopedia of ocean sciences. Academic Press, London, p 44–53

Fulton EA (2010) Approaches to end-to-end ecosystem models. J Mar Syst 81:171–183

Fulton EA, Fuller M, Smith ADM, Punt A (2004) Ecological indicators of the ecosystem effects of fishing: final report. Rep R99/1546. Australian Fisheries Management Authorities, Canberra

Fulton E, Smith A, Punt A (2005) Which ecological indicators can robustly detect effects of fishing? ICES J Mar Sci 62: 540–551

Gjøsaeter J, Kawaguchi K (1980) A review of the world resources of mesopelagic fish. FAO Fish Tech Pap 193. FAO, Rome

Godø OR, Patel R, Pedersen G (2009) Diel migration and swimbladder resonance of small fish: some implications for analyses of multifrequency echo data. ICES J Mar Sci 66:1143–1148

Godø OR, Samuelsen A, Macaulay GJ, Patel R and others (2012) Mesoscale eddies are oases for higher trophic marine life. PLOS ONE 7:e30161, oi:10.1371/journal. pone.0030161

Gon O, Heemstra PC (eds) (1990) Fishes of the Southern Ocean. JLB Smith Institute of Ichthyology, Grahamstown

Grant S, Constable A, Raymond B, Doust S (2006) Bioregionalisation of the Southern Ocean: report of Experts Workshop. WWF-Australia and Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC), Hobart

Griffiths G, Fielding S, Roe HSJ (2005) Biological-physical-acoustical interactions. In: Robinson AR, McCarthy JJ, Rothschild BJ (eds) The sea, Vol 12. Harvard University Press, New York, NY, p 441–474

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8: 993–1009

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157:89–100

Handegard NO, Demer DA, Kloser R, Lehodey P, Maury O, Simrad Y (2009) Toward a global ocean ecosystem mid-trophic automatic acoustic sampler (MAAS). In: Hall J, Harrison DE, Stammer D (eds) Proceedings of OceanObs'09: sustained ocean observations and information for society. European Space Agency, Venice, p 442–449

Handegard NO, Buisson L, Brehmer P, Chalmers SJ and others (2013) Towards an acoustic-based coupled observation and modelling system for monitoring and predicting ecosystem dynamics of the open ocean. Fish Fish 14: 605–615

Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall, New York, NY

Haury LR, McGowan JA, Wiebe PH (1978) Patterns and processes in the time–space scales of plankton distributions. In: Steele JH (ed) Spatial pattern in plankton communities. Plenum Press, New York, NY, p 277–327

Hempel G (1985) Antarctic marine food webs. In: Siegfried WR, Condy PR, Laws RM (eds) Antarctic nutrient cycles and food webs. Springer, Berlin, p 266–270

Hewitt RP, Demer DA, Emery JH (2003) An 8-year cycle in krill biomass density inferred from acoustic surveys conducted in the vicinity of the South Shetland Islands during the austral summers of 1991-1992 through 2001-2002. Aquat Living Resour 16:205–213

Hijmans RJ, Phillips S, Leathwick J, Elith J (2015) dismo: species distribution modeling. R package version 1.0-12. https://CRAN.R-project.org/package=dismo

Irigoien X, Klevjer TA, Røstad A, Martinez U and others (2014) Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nat Commun 5:3271

Jennings S, Mélin F, Blanchard J, Forster RM, Dulvy NK, Wilson RW (2008) Global-scale predictions of community and ecosystem properties from simple ecological theory. Proc R Soc B 275:1375−1383

Johnson RK (1974) A revision of the alepisauroid family Scopelarchidae. Field Museum of Natural History, Chicago, IL

Johnson RK (1982) Fishes of the families Evermannellidae and Scopelarchidae: systematics, morphology, interrelationships, and zoogeography. Fieldiana Zool 12:1−252

Kaartvedt S, Staby A, Aksnes DL (2012) Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass. Mar Ecol Prog Ser 456:1−6

Karp WA (Ed. (2007) Collection of acoustic data from fishing vessels. Coop Res Rep 287. ICES, Copenhagen

Kloser RJ, Ryan T, Sakov P, Williams A, Koslow JA (2002) Species identification in deep water using multiple acoustic frequencies. Can J Fish Aquat Sci 59:1065−1077

Kloser RJ, Ryan TE, Young JW, Lewis ME (2009) Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges. ICES J Mar Sci 66: 998−1006

Kloser RJ, Ryan TE, Keith G, Gershwim L (2016) Deep-scattering layer, gas-bladder density, and size estimates using a two-frequency acoustic and optical probe. ICES J Mar Sci 73:2037−2048

Koubbi P, Moteki M, Duhamel G, Goarant A and others (2011a) Ecoregionalization of myctophid fish in the Indian sector of the Southern Ocean: results from generalized dissimilarity models. Deep-Sea Res II 58:170−180

Koubbi P, Hulley PA, Pruvost P, Henri P and others (2011b) Size distribution of meso- and bathypelagic fish in the Dumont d'Urville Sea (East Antarctica) during the CEAMARC surveys. Polar Sci 5:195−210

Lavery AC, Wiebe PH, Stanton TK, Lawson GL, Benfield MC, Copley N (2007) Determining dominant scatterers of sound in mixed zooplankton populations. J Acoust Soc Am 122:3304−3326

Lea MA, Cherel Y, Guinet C, Nichols PD (2002) Antarctic fur seals foraging in the Polar Frontal Zone: inter-annual shifts in diet as shown from fecal and fatty acid analyses. Mar Ecol Prog Ser 245:281−297

Lehodey P, Senina I, Murtugudde R (2008) A spatial ecosystem and populations dynamics model (SEAPODYM)—modeling of tuna and tuna-like populations. Prog Oceanogr 78:304−318

Lehodey P, Murtugudde R, Senina I (2010) Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. Prog Oceanogr 84:69−84

Lehodey P, Conchon A, Senina I, Domokos R and others (2015) Optimization of a micronekton model with acoustic data. ICES J Mar Sci 72:1399−1412

Lennon JJ (2000) Red-shifts and red herrings in geographical ecology. Ecography 23:101−113

Loots C, Koubbi P, Duhamel G (2007) Habitat modelling of Electrona antarctica (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. Polar Biol 30:951−959

MacLennan DN, Fernandes PG, Dalen J (2002) A consistent approach to definitions and symbols in fisheries acoustics. ICES J Mar Sci 59:365−369

Macpherson E (2002) Large-scale species-richness gradients in the Atlantic Ocean. Proc R Soc B 269:1715−1720

Mann KH (1984) Fish production in open ocean ecosystems. In: Fasham MJR (ed) Flows of energy and materials in marine ecosystems. Plenum Press, New York, NY, p 435−458

Mapstone GM (2014) Global diversity and review of Siphonophorae (Cnidaria : Hydrozoa). PLOS ONE 9:e87737

Marshall NB (1960) Swimbladder structure of deep-sea fishes in relation to their systematics and biology. Discovery Reports, Vol 31. Cambridge University Press, Cambridge

Marshall NB (1971) Swimbladder development and the life of deepsea fishes. In: Farquhar GB (ed) Proceedings of an International Symposium on Biological Sound Scattering in the Ocean. Department of the Navy, Washington, DC, p 69−73

Maury O (2010) An overview of APECOSM, a spatialized mass balanced 'Apex Predators ECOSystem Model' to study physiologically structured tuna population dynamics in their ecosystem. Prog Oceanogr 84:113−117

Maury O, Faugeras B, Shin YJ, Poggiale JC, Ben Air T, Marsac F (2007) Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 1: The model. Prog Oceanogr 74:479−499

McGinnis RF (1982) Biogeography of the lanternfishes (Myctophidae) south of 30° S. American Geophysical Union, Washington, DC

Morrow R, Le Traon PY (2006) 15 Years of satellite altimetry and mesoscale ocean dynamics. Special Publication. European Space Agency, Paris

Mosteller F, Tukey J (1968) Data analysis, including statistics. In: Lindzey G, Aronson E (eds) Handbook of social psychology, 2nd edn. Addison−Wesley, Reading, MA, p 80−203

Newman L, Constable AJ, Swart S, Schofield O, Wåhlin A (2015) Southern Ocean Observing System 3-Year Progress Report 2012−2014. http://www.soos.aq/images/soos/products/attachments/SOOS-Progress%20Report 2015.pdf

Nikolsky GV (1963) The ecology of fishes. Academic Press, London

Nychka D, Furrer R, Sain S (2015) fields: tools for spatial data. R package version 8.2-1. https://CRAN.R-project.org/package=fields

O'Brien L, Rago P (1996) An application of the generalized additive model to groundfish survey data with Atlantic cod off the northeast coast of the United States as an example. Northwest Atl Fish Organ Sci Council Stud 28: 79−95

O'Driscoll RL, Macaulay GJ, Gauthier S, Pinkerton M, Hanchet S (2011) Distribution, abundance and acoustic properties of Antarctic silverfish (Pleuragramma antarcticum) in the Ross Sea. Deep-Sea Res II 58:181−195

Orsi AH, Whitworth T III, Nowlin WD Jr (1995) On the meridional extent and fronts of the Antarctic Circumpolar Current. Deep-Sea Res I 42:641−673

Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. https://CRAN.R-project.org/package=sp

Peña M, Olivar MP, Balbín R, López-Jurado JL, Iglesias M, Miquel J (2014) Acoustic detection of mesopelagic fishes in scattering layers of the Balearic Sea (western Mediterranean). Can J Fish Aquat Sci 71:1186−1197

Pierce D (2014) ncdf: Interface to Unidata netCDF data files. R package version 1.6.8. https://CRAN.R-project.org/package=ncdf4

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-Plus. Springer-Verlag, New York, NY

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.r-project.org

Reynolds WW, Casterlin ME (1980) The role of temperature in the environmental physiology of fishes. In: Ali MA (ed) Environmental physiology of fishes. NATO Advanced Study Institutes Series (Series A: Life Science), Vol 35. Springer US, Boston, MA, p 497–518

Ridgeway G (2015) gbm: generalized boosted regression models. R package version 2.1.1. https://CRAN.R-project.org/package=gbm

Rindorf A, Jensen H, Schrum C (2008) Growth, temperature, and density relationships of North Sea cod (*Gadus morhua*). Can J Fish Aquat Sci 65:456–470

Rintoul SR, Meredith MP, Schofield O, Newman L and others (2012) The Southern Ocean observing system: initial science and implementation strategy. Scientific Committee on Antarctic Research, Cambridge

Ryan TE, Downie RA, Kloser RJ, Keith G (2015) Reducing bias due to noise and attenuation in open-ocean echo integration data. ICES J Mar Sci 72:2482–2493

Sakshaug E, Holm-Hansen O (1984) Factors governing pelagic production in polar oceans. In: Holm-Hansen O, Bolis OL, Gilles R (eds) Marine phytoplankton and productivity. Springer-Verlag, Berlin, p 1–18

Salvanes AGV, Kristoffersen JB (2001) Mesopelagic fishes. In: Steele JH, Thorpe SA, Turekian KK (eds) Encyclopedia of ocean sciences. Academic Press, San Diego, CA, p 1711–1717

Simmonds J, MacLennan DN (2005) Fisheries acoustics: theory and practice, 2nd edn. Wiley-Blackwell, Oxford

Smith WO, Sakshaug E (1990) Polar phytoplankton. In: Smith WO (ed) Polar oceanography. Part B: chemistry, biology and geology. Academic Press, New York, NY, p 477–525

Sokolov S, Rintoul SR (2009) Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 2. Variability and relationship to sea surface height. J Geophys Res 114:C11019

Thiers L, Delord K, Bost CA, Guinet C, Weimerskirch H (2017) Important marine sectors for the top predator community around Kerguelen Archipelago. Polar Biol 40: 365–378

Trenkel VM, Ressler PH, Jech M, Giannoulaki M, Taylor C (2011) Underwater acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. Mar Ecol Prog Ser 442:285–301

Turner J, Barrand NE, Bracegirdle TJ, Convey P and others (2014) Antarctic climate change and the environment: an update. Polar Rec 50:237–259

UNESCO (2009) Global open oceans and deep seabed (GOODS) - biogeographic classification. UNESCO, Paris

Van de Putte AP, Jackson GD, Pakhomov E, Flores H, Volckaert FAM (2010) Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign. Deep-Sea Res II 57: 956–967

Watkins J, Brierley A (2002) Verification of the acoustic techniques used to identify Antarctic krill. ICES J Mar Sci 59: 1326–1336

Williams GJ, Aeby GS, Cowie ROM, Davy SK (2010) Predictive modeling of coral disease distribution within a reef system. PLOS ONE 5:e9264

Withers PC (1992) Comparative animal physiology. Saunders College Publishing, Philadelphia, PA

Wood SN (2006) Generalized additive models: an introduction with R. Texts in statistical science. Chapman & Hall/CRC, Boca Raton, FL

Wood S (2015) mgcv: mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.8-23. https://CRAN.R-project.org/package=mgcv

Yasuma H, Takao Y, Sawada K, Miyashita K, Aoki I (2006) Target strength of the lanternfish, *Stenobrachius leucopsarus* (family Myctophidae), a fish without an airbladder, measured in the Bering Sea. ICES J Mar Sci 63:683–692

Zuur AF, Ieno EN, Smith GM (2007) Analyzing ecological data. Springer, New York, NY

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York, NY