South Pacific Regional Fisheries Management Organisation

8th MEETING OF THE SCIENTIFIC COMMITTEE

New Zealand, 3 to 8 October 2020

SC8-HM07

Spatial occurrence of Jack mackerel using Bayesian Hierarchical spatial models

Chile

PO Box 3797, Wellington 6140, New Zealand
P: +64 4 499 9889 – F: +64 4 473 9579 – E: secretariat@sprfmo.int
www.sprfmo.int

Estimation and prediction of the spatial occurrence of jack mackerel (*Trachurus murphyi*) using Bayesian Hierarchical spatial models

Sebastián Vásquez[1], Cristian Salas, Aquiles Sepúlveda[1] & Maria Grazia Pennino[2]

[1] Instituto de Investigación Pesquera, Chile.
[2] Instituto Español de Oceanografía, España.

Abstract

A methodological approach is presented for modelling the occurrence patterns of jack mackerel (*Trachurus murphyi*) with the aim i) to describe the spatial distribution of the species; ii) to determine the environmental variables that drives the spatial distribution; iii) to provide insights of the spatial structure of the resource for fisheries management purposes. Information from the commercial catches of jack mackerel is used to implement the model. This information comes from different fleets that operate in the southeastern Pacific: Ecuador and Peru, northern Chile, south central Chile and the international fleet from the high sea. The presence/absence of jack mackerel is modelled with a hierarchical Bayesian spatial model using the geographical and environmental characteristics of each fishing location. Maps of predicted probabilities of presence are generated using Bayesian kriging. Bayesian inference on the parameters and prediction of presence/absence in new locations (Bayesian kriging) are made by considering the model as a latent Gaussian model. This allows the use of the integrated nested Laplace approximation (INLA) which has been seen to be quite a bit faster than the well-known MCMC methods. In particular, the spatial effect has been implemented with the Stochastic Partial Differential Equation (SPDE) approach. The analysis shows that environmental and geographical factors can play an important role in directing local distribution and variability in the occurrence of jack mackerel. Although this approach is used to recognize the habitat of adult jack mackerel, it could also be for other different life stages in order to improve knowledge regarding the species population structure.

1. Introduction

Modelling patterns of the presence/absence of the species using local environmental factors has been extensively used to address several issues, including identifying essential fish habitats and predicting the response of species to environmental features (Giannoulaki *et al*., 2013). Species distribution models (SDMs) are an essential tool for science and management, as they provide a clear picture of the distribution dynamics and extent of marine resources (Pennino *et al*., 2014; Martínez-Minaya *et al*., 2018). In general, information to build marine SDMs can be obtained from two main sources: catch-independent and catch-dependent data.

As has been pointed out by Muñoz *et al*. (2013), different approaches and methodologies have been proposed for SDMs: species envelope models such as BIOCLIM (Busby, 1991), Generalized Linear Models (GLM), Generalizes Additive Models (GAM), neural networks (Zhang *et al*., 2008) and the multivariate adaptive regression splines (MARS) (Leathwick *et al*., 2005). Most of these methods are explanatory models that assess the presence of a species in relation to a suite of one or more explanatory variables (e.g. sea temperature, bathymetry, etc.). The theoretical basis of these methods consider that the observations are independent, while the fisheries data are often characterized by presenting spatial autocorrelation under the consideration that the species display a preferential distribution due to environmental conditions or features (Hurlbert, 1984). In addition, one of the main concerns associated with fisheries data is that target-species samples are collected by preferential sampling as the fishing fleets are commercially driven (Conn *et al*., 2017).

Chilean jack mackerel (CHJM; *Trachurus murphyi*, Nichols) a highly migratory pelagic species that is widely distributed in the southeastern Pacific from Ecuador to Chile, reaching across the Pacific to New Zealand and Tasmania (Bailey, 1989; Gretchina *et al*., 1998). The wide distribution of this species in the region and its highly migratory behavior make it difficult to gather evidence supporting specific hypotheses about its spatial dynamics and population structure. Furthermore, due to its transboundary nature, jack mackerel is caught by several fleets that operate in different areas of its global distribution. Since fishing data are the most abundant information related to the distribution of the species, it is important to gather databases from all these fleets to study the spatial distribution and its relationship with environmental variables.

In this contribution, the use of a hierarchical Bayesian model is presented to predict the occurrence of jack mackerel incorporating the environmental and spatial characteristics of each fishing location to a specific fishing year. In particular, this approach uses the geographical characteristics, such as latitude, longitude of each fishing location from different fleets that operate in different areas: i) Ecuador and Peru; ii) central and northern Chile; iii) central and southern Chile; iv) international high sea fleet. The Bayesian approach is appropriate to spatial hierarchical model analysis because it allows both the observed data and model parameters to be random variables (Banerjee *et al*., 2004), resulting in a more realistic and accurate estimation of uncertainty. In particular, the integrated nested Laplace approximations (INLA) methodology (Rue *et al*., 2009) and software (http://www.r-inla.org) are used as an alternative to Markov chain Monte Carlo (MCMC) methods. Another advantage of this approach it is their generality, which makes it possible to perform Bayesian analysis in a straightforward way and to compute model comparison criteria and various predictive measures so that models can be compared easily (Rue *et al*., 2009).

2. Methods

2.1. Study area and jack mackerel data base

The study area is bounded by longitude 65ºE to 120ºW and latitude 0º to 50ºS in the southeastern Pacific, where jack mackerel fishery operates throughout the year. Jack mackerel occurrence data were obtained from fisheries data available by the EU (Dutch fleet), Peruvian and Chilean fishing fleets (Table 1). In particular, for the application of this methodological approach, the database of the 2009 fishing season was used, when information was available from all the fleets. Due to the fisheries nature of the data, only records positive samples (i.e. presences) are available in the original dataset. To generate a presence/absence dataset, pseudo-absences were generated randomly for the study area. A set of 1000 pseudo-absences were generated times using the "randomPoints" function (simple random sampling without replacement) from the "dismo" package (Hijsman *et al*., 2017) of the R software (R Development Core Team, 2019).

Table 1. Jack mackerel fisheries presence data used to construct the species distribution model for jack mackerel distribution.

Fleet	Source	Number of records	Temporal resolution
Perú	PRODUCE	35	Monthly
central-northern Chile	IFOP	670	Daily
central-southern Chile	INPESCA-IFOP	925	Daily
International high seas	European Union	624	Monthly

2.1. Environmental data

The environmental variables were obtained from different sources including satellite information and regional biogeochemical models. Five abiotic and biotic variables were processed and analyzed based on previous studies analyzing jack mackerel environmental preferences (Nuñez *et al*., 2009): i) monthly sea surface temperature; ii) monthly chlorophyll-a (https://oceancolor.gsfc.nasa.gov/); iii) geostrophic currents derived eddy kinetic energy (https://www.aviso.altimetry.fr/); iv) wind induced turbulence (https://climatedataguide.ucar.edu/); v) dissolved oxygen at a depth of 50 meters (ROMS-PISCES model). From the dissolved oxygen information (DO) a derived dummy variable (ZMO) was obtained: the presence of minimum oxygen (if DO <= 48 µmol L^{-1}, ZMO = 1; if OD > 50 µmol L^{-1}, ZMO = 0). The environmental variables were extracted for each jack mackerel occurrence record, according to their position and date. All environmental variables were tested for correlation, collinearity, outliers and missing values considered to be use in the models. As expected, dissolved oxygen was highly correlated to ZMO (Pearson correlation, $r > 0.8$; p-value < 0.001) and was eliminated. The rest of the variables presented low correlation and collinearity values. Moreover, variables were standardized using the function "decostand" in the "vegan" package (Oksanen *et al*., 2013) of the R software, in order to facilitate interpretation and to enable comparison of relative weights between variables (Orúe *et al*., 2020).

2.2. Species distribution model

Hierarchical Bayesian models were used to predict the probability of jack mackerel presence with respect to the selected environmental variables using catch-dependent data. Following Orue *et al* (2020), for modelling purpose the response variable was a binary variable that represents the presence (1) or absence (0) of jack mackerel (Yi) in each location i, and then the occurrence was modelled as:

$$Y_i \sim Bernoulli(\pi_i) i = 1, \dots, p$$

$$\log(\pi_i) = X_i \beta + W_i$$

$$\beta \sim N(\mu_\beta, q_\beta)$$

$$W_i \sim N(0, Q(k, t))$$

where π_i represents the probability of the species presence for a given location (i), $Xi\beta$ represents the matrix of the fixed effects for the linear predictor and Wi represents the spatially structured random effect at the location i. Gaussian distributions with a zero mean and covariance matrix (Q) was assumed for the spatial component, which depend on the hyperparameters k and τ, and determine the range of the effect and the total variance, respectively. Hyperpriors for k and τ are centered in values such that the range is about 20% of the diameter of the region and the variance is equal to 1 (Lindgren *et al*., 2011).

Integrated Nested Laplace Approximations (INLA) approach (Rue *et al*., 2009) and the package INLA (http:\\www.r-inla.org) that is implemented in the R software were used to obtain Bayesian parameter estimates and predictions. The spatial effects (W) were computed using the Stochastic Partial Differential Equations (SPDE) approach implemented in INLA (Lindgren *et al*., 2011), which ensures that the continuous spatial domain (also known as Gaussian Random Field; GRF) is discretized into smaller spatial units (known as Gaussian Markov Random Field; GMRF). Default zero-mean Gaussian non-informative prior distributions with a variance of 100 were used for all of the parameters involved in the fixed effects as recommended by Held *et al*. (2010).

As pointed out by Orue *et al*. (2020), the selection of explanatory variables was conducted by comparing all possible interactions, but only the best combination of variables was chosen based on the Watanabe-Akaike (WAIC) information criterion (Watanabe, 2010) and Log-Conditional Predictive Ordinations (LCPO) (Roos & Held, 2011). Specifically, lower WAIC values indicate a better fit, while lower LCPO scores represent better predictive quality. The best compromise between fit, parsimony and predictive quality occurs when smaller values the WAIC and LCPO are obtained. Thus, the best models were selected based on the mentioned compromise between low WAIC and LCPO values, containing only relevant predictors (i.e. those with 95% credibility intervals excluding zeros).

Once the inference was carried out, we predicted the probability of jack mackerel presence in the area of interest using Bayesian kriging (Muñoz *et al*., 2013). The prediction in INLA was performed simultaneously with the inference, considering the prediction locations as points where the response is missing. The INLA SPDE module allows the construction of a Delaunay triangulation covering the region of interest for the prediction (Figure 1). Once the prediction is performed in the observed location, there are additional functions that linearly interpolate the results within each triangle into a finer regular grid.

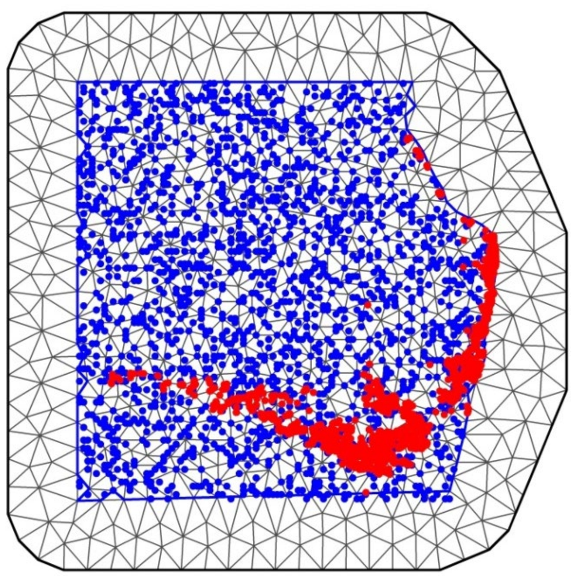

Figure 1. Map of the study area with jack mackerel observations (red dots). Delaunay triangulation used to calculate the Gaussian Markov random field for the SPDE approach. Observations (presence) are shown in red and pseudo-absences in blue.

3. Results

In order to demonstrate the applicability of this method to evaluate the spatial structure of jack mackerel incorporating environmental information, a multi-fleet approach was used covering an extensive area in the South Eastern Pacific (Figure 1). The distribution of fishing sets shows a characteristic distribution of jack mackerel that extends from the coastal region off Peru and northern Chile to the high seas off central-southern Chile. Regarding environmental variables, those that were available and considered potentially relevant for migratory pelagic species such as jack mackerel were included. In particular, sea surface temperature and chlorophyll have been recognized as important covariates of jack mackerel distribution, as well as dissolved oxygen and the presence of minimun oxygen zone (Bertrand *et al*., 2016). Furthermore, wind-induced turbulence and eddy kinetic energy were incorporated as potential modulators of the water column stability (Figure 2).

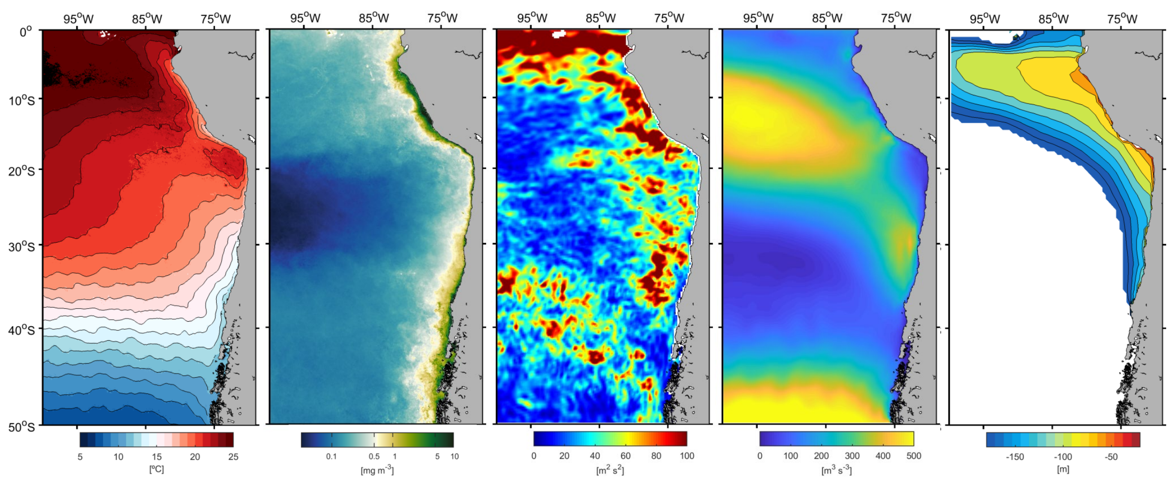

Figure 2. Maps of the covariates considered in the modelling of jack mackerel distribution in the southeastern Pacific: a) sea surface temperature; b) chlorophyll-a; c) eddy kinetic energy; d) wind induced turbulence; f) minimum oxygen depth.

All the resulting models obtained from combining those five covariates were fitted and compared. WAIC was used as a measure for goodness-of-fit, while the logarithmic score (LCPO) measure the predictive quality of the models. As shown in Table 2 and Figure 3, both measures agree on the same model, with a reasonable predictive quality. In particular, the model comparison indicates that (apart from the spatial effect) the sea surface temperature, chlorophyll-a concentration, wind induced turbulence and the presence of minimal oxygen concentrations (less statistical significance) play a determining role in jack mackerel distribution.

Table 2. Comparison of the most relevant models for the jack mackerel B-HSMs selected using catch data. Statistics acronyms are: Watanabe Akaike Information Criterion (WAIC) and Logarithmic Cross Validated Score (LCPO). Predictor acronyms are: W= spatial effect, SST = Sea Surface Temperature, SSH = Sea Surface Height, TUR = Wind Induced Turbulence, ZMO = presence of minimum oxygen concentration (50 m), EKE= Kinetic Energy, CHL= Chlorophyll.

ID	Model	WAIC	LCPO
1	beta0 + sst + clo + tur + zmo + W	1876.02996	0.56423327
2	beta0 + sst + clo + tur + W	1876.10477	0.55674368
3	beta0 + sst + clo + eke + tur + W	1876.42977	0.53756811
4	beta0 + sst + clo + eke + tur + zmo + W	1878.62428	0.54303407
5	beta0 + sst + tur + zmo + W	1883.76859	0.6119793
6	beta0 + sst + eke + tur + zmo + W	1886.46385	0.61817065
7	beta0 + sst + tur + W	1896.30608	0.71220925
8	beta0 + clo + tur + W	1896.82713	0.48600369
9	beta0 + sst + eke + tur + W	1899.20714	0.71779017
10	beta0 + clo + eke + tur + W	1899.50565	0.49008257

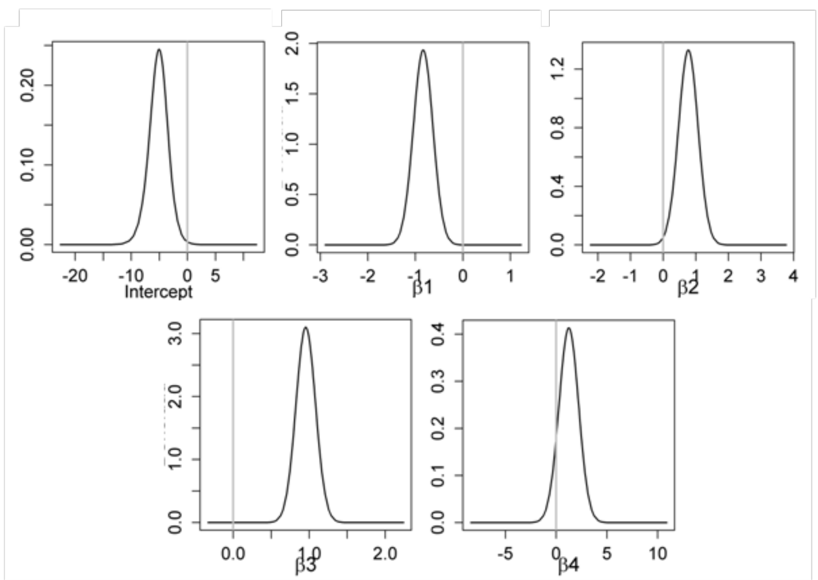

Figure 3. Posterior distributions of the fixed effects for the jack mackerel B-HSMs. β1=sea surface temperature; β2=chlorophyll-a concentration; β3=wind-induced turbulence, and; β4= presence of minimum oxygen concentration (50 m);

As can be seen in Table 2 and Figure 3, most of covariates have a significant influence on driving the jack mackerel distribution. Table 3 shows a numerical summary of the posterior distribution of the effects shown in Figure 3. This results show that sea surface temperature affects the distribution of the species studied negatively, while the chlorophyll-a concentration, the wind induced turbulence and the presence of minimum oxygen concentration have a positive relationship. Results therefore indicate that during the study period occurrence of jack mackerel was is greater in coastal and transition waters where due to the presence of active upwelling centers, the waters are colder and the concentration of the chlorophyll-a is higher with to respect to deeper waters. Unexpectedly, the highest occurrence of jack mackerel occurred where the minimum oxygen zone is shallower, which is associated with the presence of subsurface equatorial waters in the coastal sector from Peru to the central zone of Chile. The latter suggests that an oxygenated column of water of 50 meters is appropriate for the common occurrence of jack mackerel.

Table 3. Numerical summary of the marginal posterior distribution of the fixed effects for the best jack mackerel B-HSMs selected using catch data. For each variable, the mean, standard deviation, and a 95% credible central interval ($Q_{0.025}$ - $Q_{0.975}$) is provided, containing 95% of the probability under the posterior distribution.

	Mean	SD	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$
Parameters					
beta0	-5.16	1.75	-8.76	-5.13	-1.78
sst	-0.84	0.21	-1.24	-0.84	-0.43
clo	0.77	0.30	0.18	0.77	1.36
tur	0.96	0.13	0.70	0.95	1.21
zmo	1.24	0.96	-0.65	1.25	3.14
Hyperparameters					
r	13.69	2.17	9.96	13.51	18.48
σ	5.37	0.62	4.28	5.33	6.69

Figure 4 displays the spatial effect that indicates the intrinsic variability of the distribution of jack mackerel after excluding environmental variables. This component shows a strong effect, with positive values from the northern coast of Peru (~ 8°S) to the southern coast of Chile (~ 41°S) which extend offshore in a band that reaches 115° W in around 40°S. The highest values are observed in the coastal region of southern Peru and northern Chile, in addition to the oceanic zone off south-central Chile where the Chilean industrial fleet operates, as well as the international high seas fleet. Moreover, the mean of the range of the spatial effect of the normal area was about ~13 geographical degrees, resulting in a wide spatial correlation field for jack mackerel occurrence. The physical meaning of this value is that jack mackerel records are this distance or greater apart are not spatially correlated.

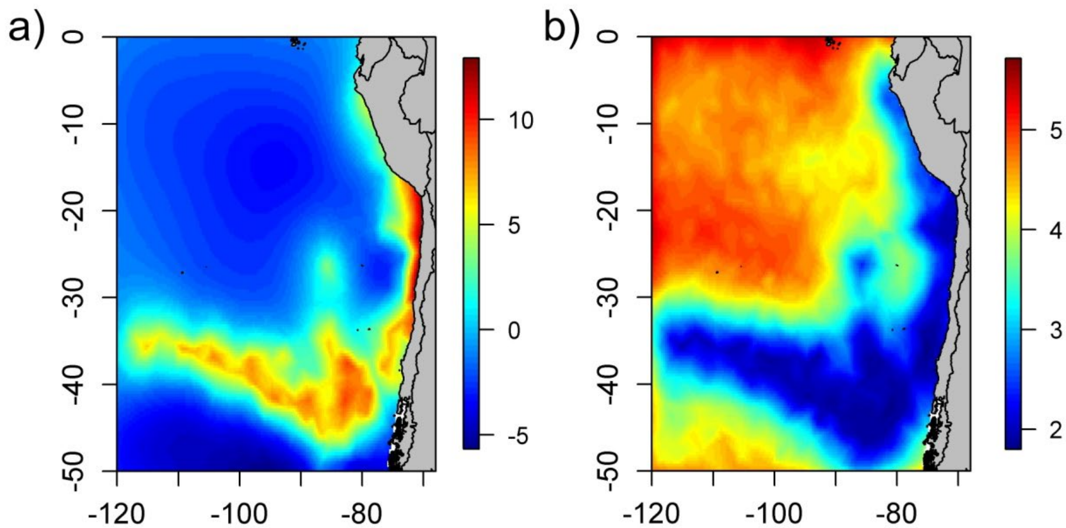

Figure 4. a) Mean and b) standard deviation for posterior distribution of the spatial effect W for jack mackerel distribution.

The predicted probability map of jack mackerel occurrence using catch data shows high aggregation and spatial continuity in the hotspots. A narrow coastal band is observed between 8ºS and 30ºS where the probability drops sharply offshore in response to warmer and less productive waters in the oceanic region. In the offshore extension around 40ºS a small discontinuity was obtained, however the predicted probability expands towards a vast oceanic area off central-southern Chile (Figure 5a). The standard deviation map (Figure 5b) show very low values in the area where data were collected, while the error increases along the edges and off the modelled domain. Finally, one of the most important results is the continuity in the area of high probability for jack mackerel ocurrence by incorporating information from all fleets in a spatial modelling procedure, which suggests the need to advance in the use of joint databases in order to study the spatial structure of the jack mackerel population in the Southeast Pacific and its relationship with habitat variability.

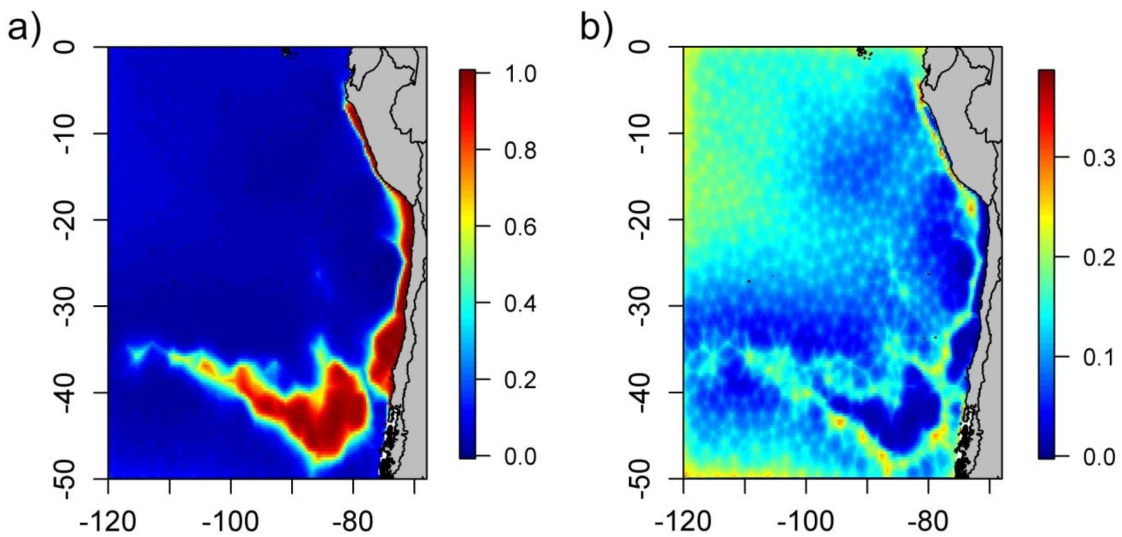

Figure 5. Posterior a) mean and b) standard deviation for predictive distribution of the probability of jack mackerel presence in southeastern Pacific.

4. Concluding remarks

The use of hierarchical Bayesian spatial models (hBSMs) is presented for the study of jack mackerel distribution and its relationship with variables that characterize its habitat. One of the main advantages of this approach is that simultaneously deals with spatial autocorrelation issues and different sources of uncertainties. Multiple sources of uncertainty associated with both the observed data and the ecological process can be included in hBSMs, resulting in a stronger statistical inference. Moreover, the posterior predictive distribution of the probability of finding the species turns out to be a very suitable tool that allows us to express our uncertainties associated with the entire species habitat prediction. Another advantage is the computational gain of using the INLA approach, which allows us to easily make inferences and predictions within a highly structured model.

Our results suggest that sea surface temperature affects the distribution of the jack mackerel negatively, while the chlorophyll-a concentration, the wind induced turbulence have a positive relationship. In addition, the results suggest the presence of jack mackerel in areas where the minimum oxygen zone reaches depths of 50 meters. These results confirm that jack mackerel is characterised by a high plasticity since it tolerates a large range of abiotic conditions. However, in its northern edge of distribution (5ºS - 25ºS) the jack mackerel habitat is restricted to a narrower coastal band due to the presence of warmer and less productive waters in the oceanic zone, while at the south of 30ºS, jack mackerel habitat expands offshore covering a larger area.

Regarding the database used in this study, it is worth mentioning that through the use of georeferenced information from different fleets that catch jack mackerel, it is possible to establish more robust statistical models to make inferences regarding the biophysical processes that drive the jack mackerel ocurrence throughout its distribution area. The latter is highly relevant to strengthen our understanding of the spatial dynamics of the species, its implications in the definition of its population structure and finally towards integrated management in the south Pacific.

5. References

Bailey, K. 1989. Description and surface distribution of juvenile Peruvian jack mackerel, *Trachurus murphyi*, Nichols from the Subtropical Convergence Zone of Central South Pacific. Fishery Bulletin US 87: 273–278.

Banerjee, S., Carlin, B., Gelfand, A. 2004. Hierarchical modeling and analysis for spatial data. CRC, London.

Bertrand A., Habasque, J., Hattab, T., Hintzen, N.T., Oliveros-Ramos, R., Gutiérrez, M., Demarcq, H., Gerlotto, F. 2016. 3-D habitat suitability of jack mackerel *Trachurus murphyi* in the Southeastern Pacific, a comprehensive study. Progress in Oceanography, 146: 199-211.

Busby JR. 1991. BIOCLIM: A bioclimatic analysis and predictive system. In: Margules C, Austin M (eds) Nature conservation: cost effective biological surveys and data analysis, CSIRO, Canberra , pp 64–68.

Conn, P.B., Thorson, J.T., Johnson, D.S. 2017. Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. Methods Ecol Evol. 8, 1535-1546.

Giannoulaki, M., Iglesias, M., Tugores, M. P., Bonanno, A., Patti, B., De Felice, A., et al. (2013). Characterizing the potential habitat of European anchovy Engraulis encrasicolus in the Mediterranean Sea, at different life stages. Fish. Oceanogr. 22, 69–89. doi: 10.1111/fog.12005.

Gretchina, A., Núñez, S. & D. Arcos. 1998. Biología reproductiva del jurel (*Trachurus symmetricus murphyi*) en el Pacífico sur. In: Arcos, D. (Ed.), Biología y ecología del jurel en aguas chilenas. Instituto de Investigación Pesquera, Talcahuano, pp. 77–79.

Held, L., Schrödle, B., Rue, H., 2010. Posterior and cross-validatory predictive checks: a comparison 545 of MCMC and INLA. Statistical modelling and regression structures, 91-110.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54:187–211.

Kinas, P.G., Andrade, H.A., 2017. Introdução à análise bayesiana (com R). Consultor Editorial.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., Hastie, T. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshwater Biol 50:2034–2052.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. 73, 579 423-498.

Martínez-Minaya, J., Cameletti, M., Conesa, D. & Pennino, M.G., 2018. Species distribution modeling: a statistical review with focus in spatio-temporal issues. Stoch. Env. Res. Risk A, 1-18.

Muñoz, F., Pennino, M.G., Conesa, D., López-Quílez, A., Bellido, J.M., 2013. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stoch. Env. Res. Risk A. 27, 1171-1180.

Núñez Elías, S., Correa Ramirez, M., Vasquez Pastene, S., 2009. Variability of the Chilean jack mackerel fishing habitat in the southern Pacific Ocean. In: EighthScience Working Group of SPRFMO, 2–6 November 2009, Auckland, NewZealand.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Package 'vegan'. Community ecology package, version 2.

Orúe, B., Pennino, M.G., Lopez, J., Moreno, G., Santiago, J., Ramos, L. & H. Murua. 2020. Seasonal Distribution of Tuna and Non-tuna Species Associated With Drifting Fish Aggregating Devices (DFADs) in the Western Indian Ocean Using Fishery-Independent Data. Front. Mar. Sci., 17 June 2020 | https://doi.org/10.3389/fmars.2020.00441

Pennino, M.G., Muñoz, F., Conesa, D., López-Quílez, A., Bellido, J.M., 2014. Bayesian spatio temporal discard model in a demersal trawl fishery. J. Sea Res. 90, 44-53.

R Development Core Team, 2019. R: A language and environment for statistical computing. R.Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roos, M., Held, L., 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Anal. 6, 259-278.

Rue, H., Martino, S., Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society, Series B 71(2):319–392.

Zhang W., Zhong X., Liu G. 2008. Recognizing spatial distribution patterns of grassland insects: neural network approaches. Stochastic Environmental Research and Risk Assessment 22(2):207–216.