Ecological basis to embrace temporal assessment and spatial management of the European hake (*Merluccius merluccius*) in the northern Iberian Peninsula.

Working document to the Working Group for the Bay of Biscay and the Iberian Waters Ecoregion WGBIE – Lisbon 2-9 May 2019.

Francisco Izquierdo[1,2], Iosu Paradinas[2,3], Francisco Velasco[4], Maria Grazia Pennino[2,5], and Santiago Cerviño[5]

[1]Departament d'Estadística i Investigació Operativa. Universitat de València. C/ Dr. Moliner 50. Burjassot. 46100. Valencia, Spain.
[2]Statistical Modeling Ecology Group (SMEG). Departament d'Estadística i Investigació Operativa, Universitat de València, C/Dr. Moliner 50, Burjassot, 46100 Valencia, Spain.
[3]Ipar Perspective Asociación, Karabiondo Kalea. 48600 Sopela, Spain.
[4]Instituto Español de Oceanografía. Centro Oceanográfico de Santander. Promontorio San Martín s/n, 39004 Santander, Spain.
[5]Instituto Español de Oceanografía. Centro Oceanográfico de Vigo. Subida a Radio Faro, 50-52. 36390 Vigo (Pontevedra), Spain.

April 30, 2019

Contents

Background

Spatial management of commercial resources is becoming an effective measure to be broadly implemented in the European Seas. However, it is currently unconnected from the population dynamics and the official assessment. Indeed, it is known that species abundance can be influenced by the environmental features of its own habitat and/or by biotic process that are spatially structured (e.g. reproduction, predation, among others). Usually, this variability is assumed to be implicitly in the abundance trends used as inputs of the stock assessment models and it is not explicitly taken into account. Within this context, in this study we propose a novel methodological approach for an effective implementation of spatial and ecological knowledge that could help to embrace species spatial management in an operational way, providing a more holistic and ecosystem-based approach. As case study we used the European hake (*Merluccius merluccius*) in the northern continental shelf of the Iberian Peninsula. Hake data by length category collected during the scientific survey series "DEMERSALES" by the "Instituto Español de Oceanografía" (IEO) from 1992-to 2017 were analyzed using hierarchical Bayesian spatial-temporal models (H-BSTMs), considering as environmental variables Sea Bottom Temperature, Sea Bottom Salinity, bathymetry and rugosity of the seabed. H-BSTMs link spatially information on hake abundance to environmental variables to estimate and predict where (and how much of) this species is likely to be present in the studied area in a specific year.

Indices of abundance obtained as outputs from H-BSTMs, performed with the innovative integrated nested Laplace approximation (INLA) methodology and software, are then used as inputs for the GADGET (Globally applicable Area Disaggregated General Ecosystem Toolbox) stock assessment model (Figure 1). Finally, a comparative analysis of the results obtained with the GADGET model using the H-BSTMs abundance indexes and the ones commonly used in stock assessment evaluations is performed.

We argue that the analytical framework proposed in this study allowed to (1) assess which environmental factors influence the different life stages of the hake in the northern continental shelf of the Iberian Peninsula, (2) identity the areas in which the different life stages are more aggregated and their spatial-temporal fluctuations, and (3) could be a decisive step to improve habitat-based standardization abundance indexes and stocks' management in European Seas.

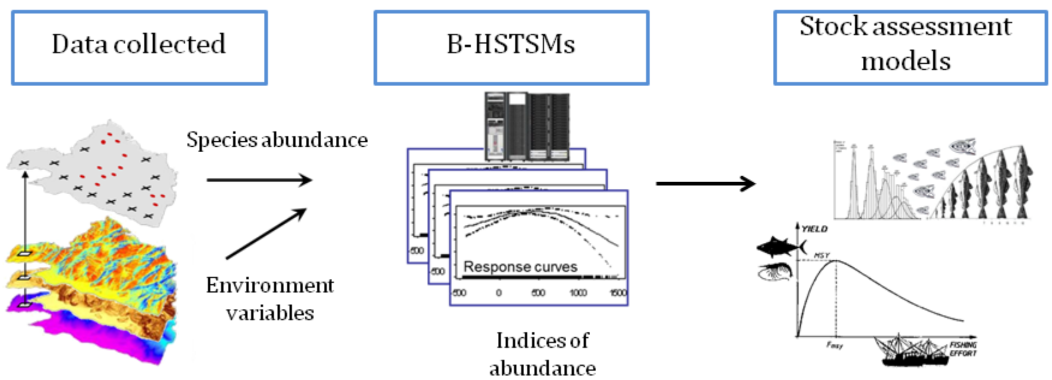

Figure 1: Working path representing how hierarchical Bayesian spatial-temporal models (H-BSTMs) will inform stock assessment models.

Material and methods

Data

The data used in this study were collected during the scientific survey series "DEMERSALES" by the "Instituto Español de Oceanografía" (IEO) carried out in autumn (September to October) from 1992 to 2017. The DEMERSALES survey makes use of a stratified sampling design based on depth with three bathymetric strata: 70–120 m, 121–200 m and 201–500 m. Sampling stations consisted of 30 min trawling hauls located randomly within each stratum at the beginning of the design. However, as a result of weather conditions or other external factors, station location varied slightly in some years and hauls were therefore not always performed at exactly the same latitude and longitude (Pennino et al., 2019). Approximately 128 hauls (minimum 119 and maximum 141) divided between the three bathymetric strata were performed every year in this zone (Figure 2), using the baka 44/60 gear (Sánchez and Gil, 2000).

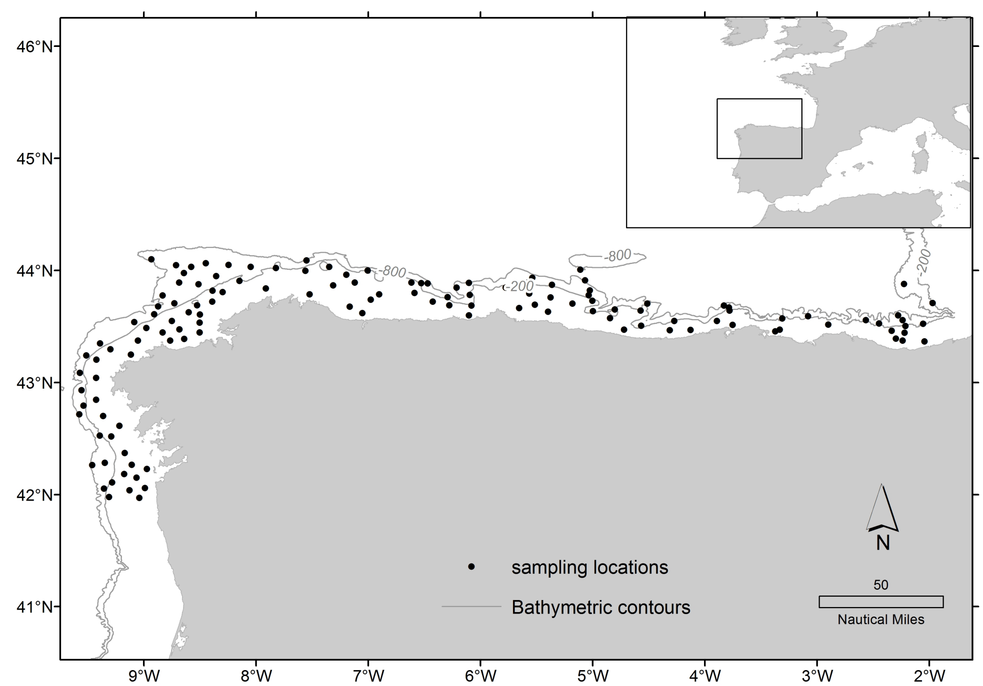

Figure 2: Study area and sampling locations (black dots) of the DEMERSALES surveys (1997-2016). Bathymetric contours indicate the 200 and 800 m isobatas.

With the European hake length distribution accessible to this gear three groups were created: recruits, which include all specimens with a length <21 cm; adults, individuals between 12 and 35 cm; and individuals larger than 35 were aggregated in a separated category.

For each one of this group two different variables were analyzed in order to describe the spatio-temporal behaviour of the European hake species. First, we considered the presence/absence variable to measure the occurrence of the species in each life stage. Secondly, we used a discrete variable, the total number of individuals per 30 minutes of trawling (i.e.

number per unit effort, NPUE), as an indicator of the conditional-to-presence abundance of the species.

Environmental variables

Three environmental variables were considered as potential or known predictors of the European hake life-stage distribution which may influence the habitat selection of this species. These include two oceanographic variables: Sea Bottom Temperature (SBT in C) and Sea Bottom Salinity (SBS in PSU), and the bathymetry (in metres).

SBT and SBS were added to the analysis as they are strongly related to marine system productivity, affecting nutrient availability and water stratification (Pennino et al., 2013). SBT and SBS values were collected during the survey with a sounding CTD (conductivity, temperature and depth) in different random sampling points of the study area. Monthly SBT and SBS maps of the entire area were obtained for each year of the studied period with the Radial basis functions (RBF) tool in ArcGIS 10.1.

The bathymetry map was retrieved from the European Marine Observation and Data Network (EMODnet, http://www.emodnet.eu/) with a spatial resolution of 0.02 x 0.02 decimal degrees.

In order to ensure the same spatial resolution, all environmental data were aggregated to the lower spatial resolutions using the *raster* package (Hijmans, 2018) in the R software (R Core Team, 2018). All covariates were explored for collinearity, outliers, and missing values before their use in the models following the approach of Zuur et al. (2010). In particular correlation among variables was tested using the Pearson's correlation, while the collinearity computing the Generalized variance-inflation factors (GVIF) (Fox and Weisberg, 2011).

Finally, to facilitate visualization and interpretation, the explanatory variables were standardized (difference from the mean divided by the corresponding standard deviation) (Gelman, 2008).

Characterizing the spatio-temporal behaviour of the European hake

This study used the spatio-temporal model structure comparison proposed by Paradinas et al. (2017) to categorize the spatio-temporal behaviour of the European hake in either opportunistic, persistent or progressive (see Table 1 and Figure 5). In particular, opportunistic structures indicate that species change their spatial pattern every year without following any specific pattern. Persistent structures imply that species have a spatial distribution that is common every year, while the progressive ones indicate that the spatial pattern of the process change from one year to another. The progressive structure contains a ρ_t parameter (see Table 1) that controls the degree of autocorrelation between consecutive years. This ρ_t parameter is bounded to $[0, 1]$, where parameter values close to 0 represent more opportunistic behaviors and parameter values close to 1 represent more persistent distributions.

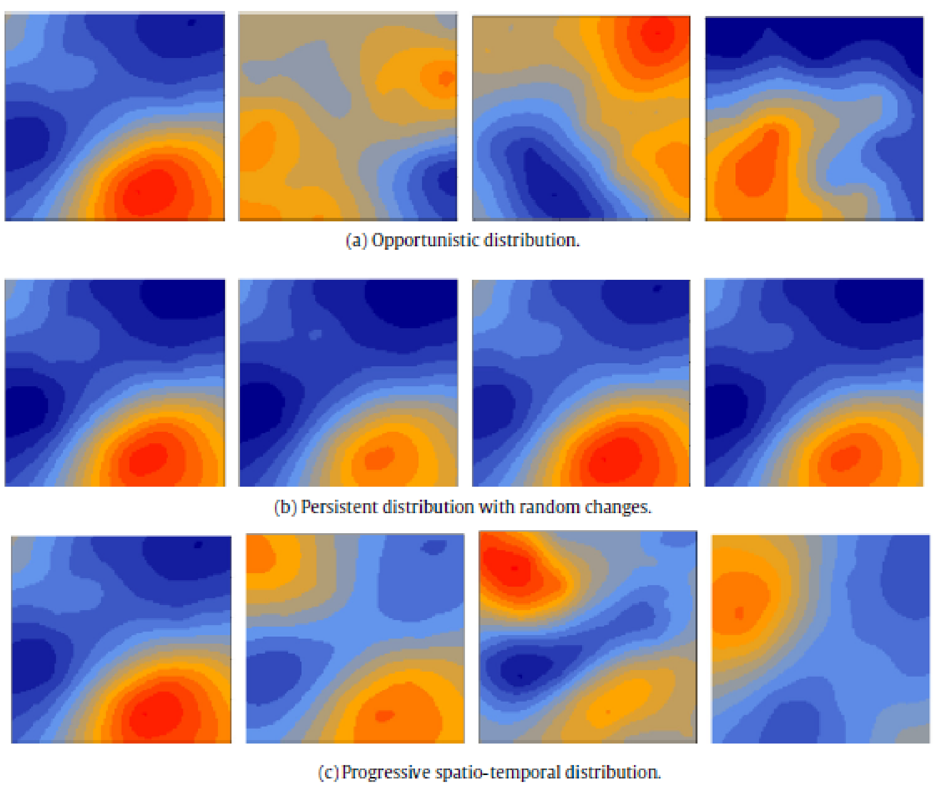

Figure 3: Simulated types of spatio-temporal scenarios. From Paradinas et al., (2017).

Modelling European hake occurrence and abundance distribution

Spatio-temporally fishery abundance data often result in observing large proportions of zeros, i.e. zero inflated data. These data are generally tackled using independent two-part models, also known as delta models. In these models, the occurrence and the conditional-to-presence abundances (NPUE) are modeled independently. However, abundance and detection probability are often related (Kéry et al., 2005), which violates the independence assumption of common delta models. This study incorporated the fact that both processes could be related by fitting shared environmental effects and/or spatio-temporal structures as described in Paradinas et al. (2017). In this way we combined information on the presence/absence of the species under study and its abundance.

In particular, Y_{st} and Z_{st} denote, respectively, the spatio-temporally distributed occurrence and the conditional-to-presence abundance (NPUE), where $s = 1, ..., n_t$ is the spatial location and $t = 1, ..., T$ the temporal index, being $i = 1, ..., I$ the environmental variable in location s. Then, as usual with this kind of variables, we modeled the occurrence, Y_{st}, using a Bernoulli distribution. In the case of the NPUE, Z_{st}, our selection to model it was a negative binomial distribution, a probability distribution that captures the overdispersion of the data. The mean of both variables was then related via the usual link functions (logit and log, respectively) to the bathymetric and spatio-temporal effects:

$$\begin{aligned}
Y_{st} &\sim \text{Ber}(\pi_{st}) \\
Z_{st} &\sim \text{NB}(\mu_{st}, \sigma_{st}) \\
\text{logit}(\pi_{st}) &= \alpha^{(Y)} + d_i + U_{st}^{(Y)} \\
\log(\mu_{st}) &= \alpha^{(Z)} + \theta_i d_i + U_{st}^{(Z)} \\
\Delta 2 d_i &= d_i - 2d_{i+1} + d_{i+2} \sim N(0, \rho_d)
\end{aligned} \tag{1}$$

where π_{st} represents the probability of occurrence at location s at time t and μ_{st} and σ_{st} are the mean and variance of the conditional-to-presence abundance. The linear predictors containing the effects to which these parameters π_{st} and μ_{st} are linked are formed with: $\alpha^{(Y)}$ and $\alpha^{(Z)}$, the terms representing the intercepts for each variable; d_i which stands for a second order Random Walk model that allows us to fit any possible non-linear relationship of the environmental variables (Fahrmeir and Lang, 2001); the final terms $U_{st}^{(Y)}$ and $U_{st}^{(Z)}$ refer to the spatio-temporal structure of the occurrence and conditional-to-presence abundance respectively and may follow any of the three spatio-temporal structures described in the previous section.

The spatial field (W_s) was modelled as a multivariate normal distribution with zero mean and a Matérn covariance function that depend on its range (r_w) and variance (σ_w). The temporal trend $f(t)$ could follow any suitable function, either a linear effect, a smooth effect, an unstructured random term, etc.

Vague prior distributions with a zero-mean and a standard deviation of 100 were implemented for all the fixed effects, the variance of the abundance process, and the scaling parameter of the shared effects. For the geostatistical terms and the ρ parameters of the second order Random Walks (RW2) PC priors (Simpson et al., 2017) were assigned fixing the probability of the range of the spatial effect at 0.15, the probability of the variance of the spatial effect at 0.20 and the probability that the precision of the RW2 effects at 0.01. A sensitivity analysis of the choice of priors was performed by verifying that the posterior distributions concentrated well within the support of the priors.

Model selection was performed testing all possible combinations among the possible spatio-temporal structures and variables and using the Watanabe Akaike Information Criterion (WAIC)(Watanabe, 2010) as criteria of the goodness of fit and the Log-Conditional Predictive Ordinates (LCPO) (Roos et al., 2011) as predictive quality measures. For both measures, the smaller the score the better the model. All these models and comparisons were fitted for all the European hake length groups.

Models were fitted using the integrated nested Laplace approximation (INLA) package (Rue et al., 2009) in the R environment.

Results and Discussion

Do the computational time at the moment we run these type of models only for the recruits group. The future steps will be do the same analysis for the others groups and use the derived abundance indices in the GADGET model to assess which kind of changes could have on the stock assessment of the European hake in this area.

European hake recruits

For the European hake recruits the best spatio-temporal structure was the progressive without shared spatio-temporal effects (Table 2). Concerning the spatio-temporal structures, shared components did not improve the progressive fitted model (Table 2), as also occurred in (Paradinas et al., 2017). This result could suggest that hake recruitment data is generated through two different processes; the probability of observing hake recruits and, if present, their abundance. However, the nature of the process under study induces to believe that this apparent independence is a consequence of the high sampling effort of the survey relative to the abundance of hake recruits, rather than being two different processes. The DEMERSALES survey trawls a relatively big areas, therefore the probability of observing at least one individual of an abundant fish species, such as hake, is quite high at environmentally not-too-challenging areas. Similarly, if effort was diminished, the detection probability would decrease proportionally and thus record a lot more zeros in our dataset.

No high correlation (Pearson's correlation lower than 0.60) and collinearity (Variance Inflation Factor, GVIF: values lower than 3) were found among the environmental variables. Consequently all variables were used in the models.

Bathymetry was the most important variable to define the occurrence and NPUE distribution of the hake recruits in the studies areas (Table 3). Indeed, although the best models, in terms of WAIC, were the one with the bathymetry and the SBS or SBS, the difference is negligible with the model that include only the bathymetry (i.e. lower than 5 units). For this reason, and following a parsimony principle, the selected model was the one with the bathymetry, fitted as shared smoothed effect between the two processes (i.e. occurrence and NPUE).

The selection of an autoregressive temporal term in the model suggests the presence of a certain degree of temporal continuity in the spatial distribution of hake recruits in the study area. These results were supported by the high temporal correlation parameters of the progressive spatio-temporal structures (0.99 and 0.96 for the occurrence and conditional-to-presence abundances respectively).

The smoothed bathymetric effect highlighted that abundance of hake recruits decreases gradually after the optimum 150–200 metre strata (Figure 4).

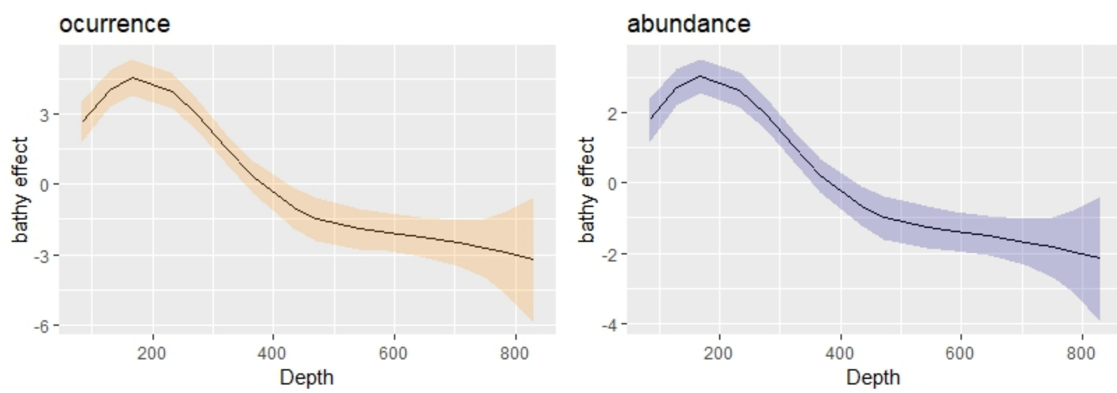

Figure 4: Bathymetric smoothed effect for both occurrence and abundance variables.

In addition, the posterior mean of the spatial effect maps in Figures 5 and 6 show a main persistent hot-spot along the continental shelf of the Artabrian gulf (off La Coruña). Although the recruitment of hake is mainly concentrated in this specific areas there have been smooth changes in the relative abundance and the spatial location from year to year.

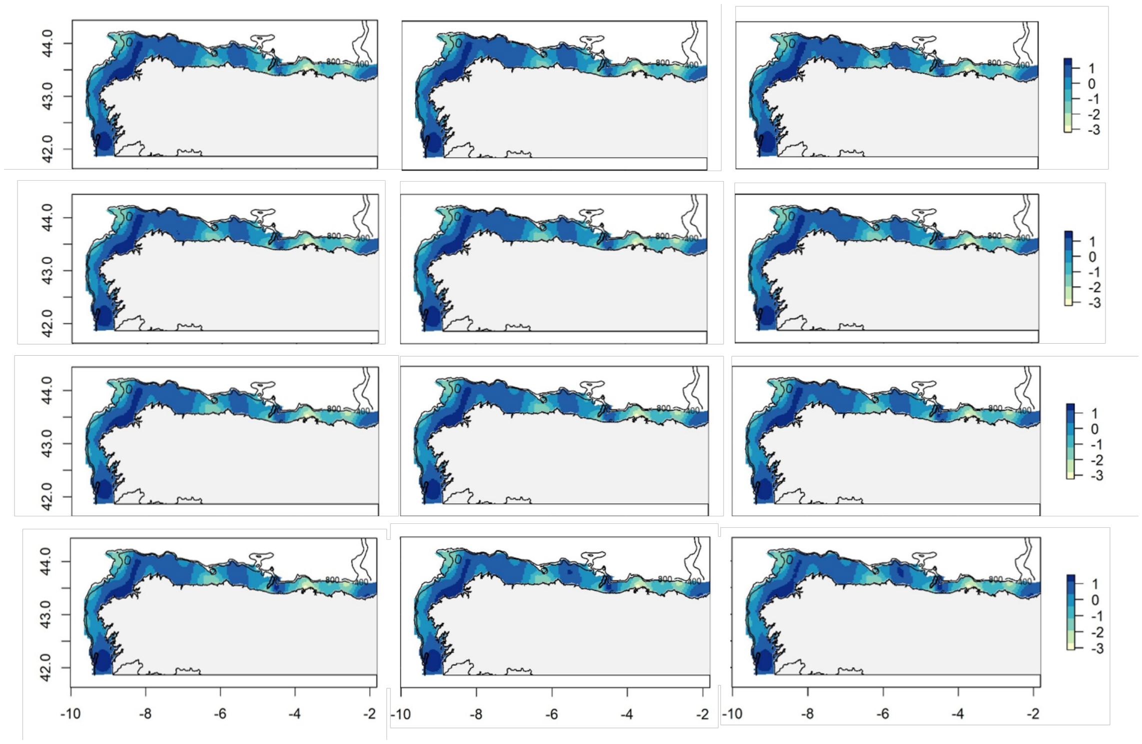

Figure 5: Posterior means of the spatial effect for the progressive model with the shared bathymetric smoothed effect for the occurrence pattern.

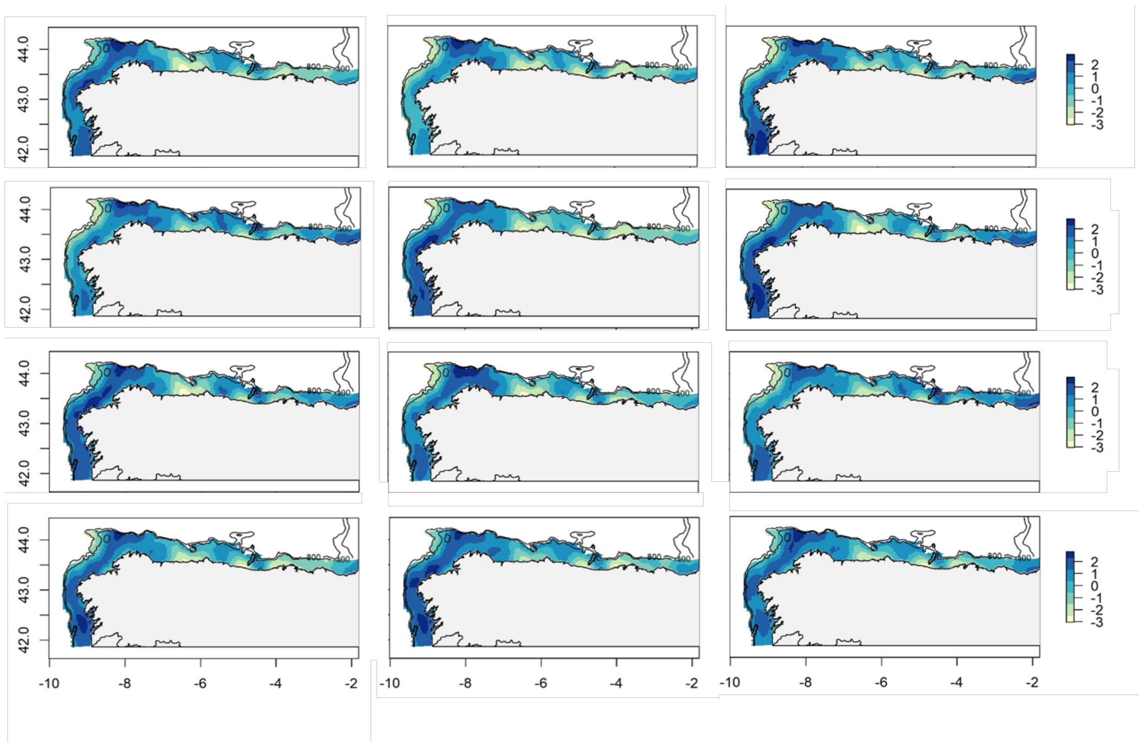

Figure 6: Posterior means of the spatial effect for the progressive model with the shared bathymetric smoothed effect for the abundance pattern.

Acknowledgments

The authors express their gratitude to all the people that work in the DEMERSALES surveys. DEMERSALES surveys were co-funded by the EU within the Spanish national program for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy.

References

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2):201–220.

Fox, J. and Weisberg, S. (2011). Multivariate linear models in r. *An R Companion to Applied Regression. Los Angeles: Thousand Oaks.*

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine*, 27(15):2865–2873.

Hijmans, R. J. (2018). *raster: Geographic Data Analysis and Modeling.* R package version 2.8-4.

Kéry, M., Royle, J. A., and Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological applications*, 15(4):1450–1461.

Paradinas, I., Conesa, D., López-Quílez, A., and Bellido, J. M. (2017). Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. *Spatial Statistics*, 22:434–450.

Pennino, M. G., Muñoz, F., Conesa, D., López-Quílez, A., and Bellido, J. M. (2013). Modeling sensitive elasmobranch habitats. *Journal of sea research*, 83:209–218.

Pennino, M. G., Vilela, R., Bellido, J. M., and Velasco, F. (2019). Balancing resource protection and fishing activity: The case of the european hake in the northern iberian peninsula. *Fisheries Oceanography*, 28(1):54–65.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Roos, M., Held, L., et al. (2011). Sensitivity analysis in bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, 6(2):259–278.

Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A., and Krainski, E. (2009). INLA: functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. *R package version 0.0*.

Sánchez, F. and Gil, J. (2000). Hydrographic mesoscale structures and poleward current as a determinant of hake (merluccius merluccius) recruitment in southern bay of biscay. *ICES journal of Marine Science*, 57(1):152–170.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594.

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, 1(1):3–14.

Tables

Model	Notation	Description
Opportunistic	$U_{st} = W_{s_t}$	Different and uncorrelated realizations of the spatial field every year.
Persistent	$U_{st} = W_s + f(t)$	A common realization of the spatial field for all years and an additive temporal trend
Progressive	$U_{st} = W_{st} + \rho_t U_{st-1}$	Spatial realizations change over time using a first order autoregressive model

Table 1: Explanation of the three different spatio-temporal structures compared in the models.

Model	WAIC	LCPO	Time (sec.)
Persistent Shared Effects	15879.45	2.90	80.91
Persistent Not Shared Effects	16001.28	2.92	118.08
Opportunistic Shared Effects	16095.17	2.95	59.82
Opportunistic Not Shared Effects	16231.99	2.95	79.56
Progressive Shared Effects	16774.70	3.05	401.62
Progressive Not Shared Effects	**15846.09**	**3.11**	**7138.10**

Table 2: Spatio-temporal structures comparison for the conditional-to-presence abundance distribution European hake recruits' model based on WAIC and LCPO scores. Time scores refer only to the estimation process of the model. The best model is highlighted in bold.

Model	WAIC	LCPO	Time
Progressive Bathymetry Shared Effects	**15659.88**	**3.02**	**13667.78**
Progressive SBS Shared Effects	15848.98	3.11	7168.39
Progressive SBT Shared Effects	15800.53	3.15	11032.17
Progressive Bathymetry SBS Shared Effects	15655.22	3.05	16488.46
Progressive Bathymetry SBT Shared Effects	15657.85	3.07	17097.45
Progressive SBS SBT Shared Effects	15804.95	3.16	11683.53
Progressive Bathymetry Not Shared Effects	15668.76	3.03	10143.00
Progressive SBS Not Shared Effects	15852.73	3.11	10662.15
Progressive SBT Not Shared Effects	15798.90	3.14	9416.98
Progressive Bathymetry SBS Not Shared Effects	15672.92	3.03	14104.07
Progressive Bathymetry SBT Not Shared Effects	15672.60	3.06	15135.95
Progressive SBS SBT Not Shared Effects	15805.43	3.14	11152.92

Table 3: Environmental effects comparison for the conditional-to-presence abundance distribution European hake recruits' model based on WAIC and LCPO scores. Time scores refer only to the estimation process of the model. The best model is highlighted in bold.