INLA Bayesian spatio-temporal Models in Action: Fisheries Science Applications

Marta Cousido-Rocha

Instituto Español de Oceanografía (IEO, CSIC)

Why are complex regression models needed?

Spatial dependency

The **First Law of Geography**, according to Waldo Tobler, is "*everything is related to everything else, but near things are more related than distant things*."

This first law is the foundation of the fundamental concepts of **spatial dependence** and **spatial autocorrelation**.

Temporal dependency

The impact of previous behavior on current behavior.

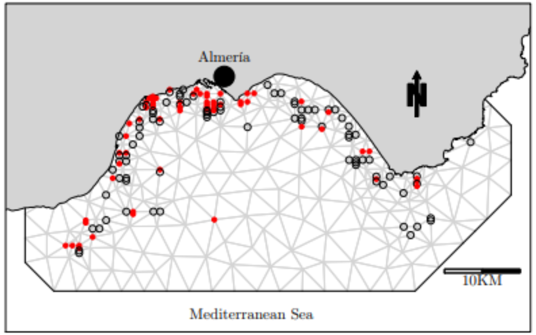
Other sources of dependency in data

E.g. Hierarchical Dependency.

Spatial dependency

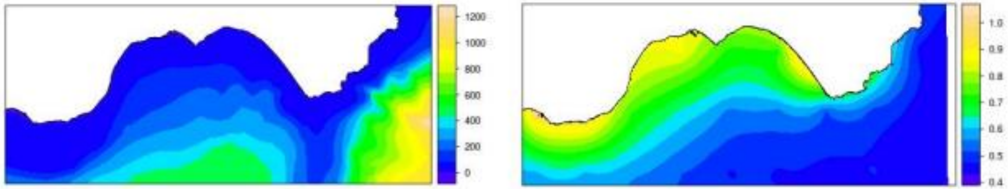
The property of random variables taking values, **at pairs of locations a certain distance apart, that are more similar** (*positive autocorrelation*) or less similar (negative autocorrelation) than expected for randomly associated pairs of random observations.

Example



Sampling locations for the presence (·) and the absence (◦) of the Mediterranean horse Mackerel in the bay of Almería

In the case of **fish species distributions**, spatial autocorrelation occurs mostly because of *habitat heterogeneity*, or because of biotic processes such as conspecific attraction or competition with another species.



(a) Bathymetry (b) Chlorophyll-a

Munoz, 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.Stochastic Environmental Research and Risk Assessment,27(5), 1171-1180.

Formulation

Intercept

Spatial effect

$$Y_s = \alpha + \sum_{j=1}^{k} f_j\left(X_s^j\right) + U_s$$

Response variable

Covariables effect

The spatial effect represents the intrinsic spatial variability of the data after excluding other covariables (e.g. environmental covariables).

Why go Bayesian?

1

Allow to incorporate prior information

Ecological Modelling 366 (2017) 1–14

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

A Bayesian model of fisheries discards with flexible structure and priors defined by experts

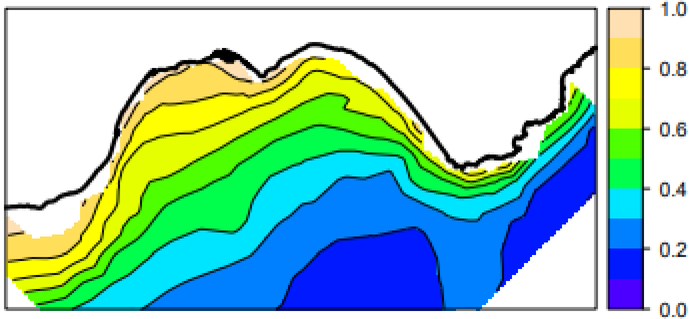
Eduardo Eiji Maeda[a,*], Samu Mäntyniemi[a], Smaragda Despoti[b,d], Claudia Musumeci[c], Vassiliki Vassilopoulou[b], Konstantinos I. Stergiou[b,d], Marianna Giannoulaki[b], Alessandro Ligas[c], Sakari Kuikka[a]

[a] Fisheries and Environmental Management Group, Department of Environmental Sciences, University of Helsinki, P.O. Box 68, FI-00014, Helsinki, Finland
[b] Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, P.O. BOX 2214, Iraklion, Crete, Greece
[c] Consorzio per il Centro Interuniversitario di Biologia Marina ed Ecologia Applicata G. Bacci, Viale N. Sauro 4, I-57128 Livorno, Italy
[d] Department of Zoology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

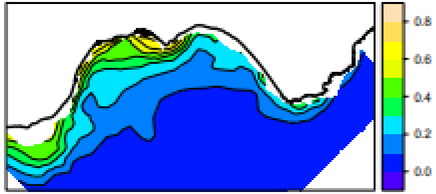
Why go Bayesian?

2

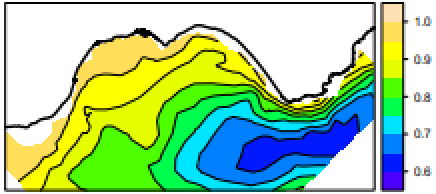
Better estimation of the uncertainty
(posterior probability distributions)



Median for probability of occurrence.



First quartile.



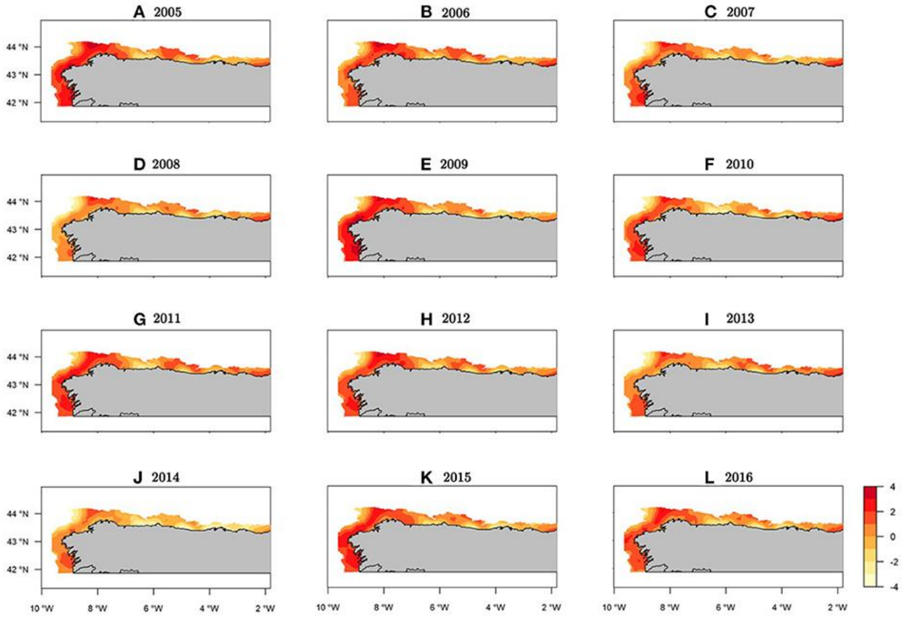
Third quartile.

Munoz, 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.Stochastic Environmental Research and Risk Assessment,27(5), 1171-1180.

Why go Bayesian?

3

Possibility to account for spatial autocorrelation and temporal correlation.



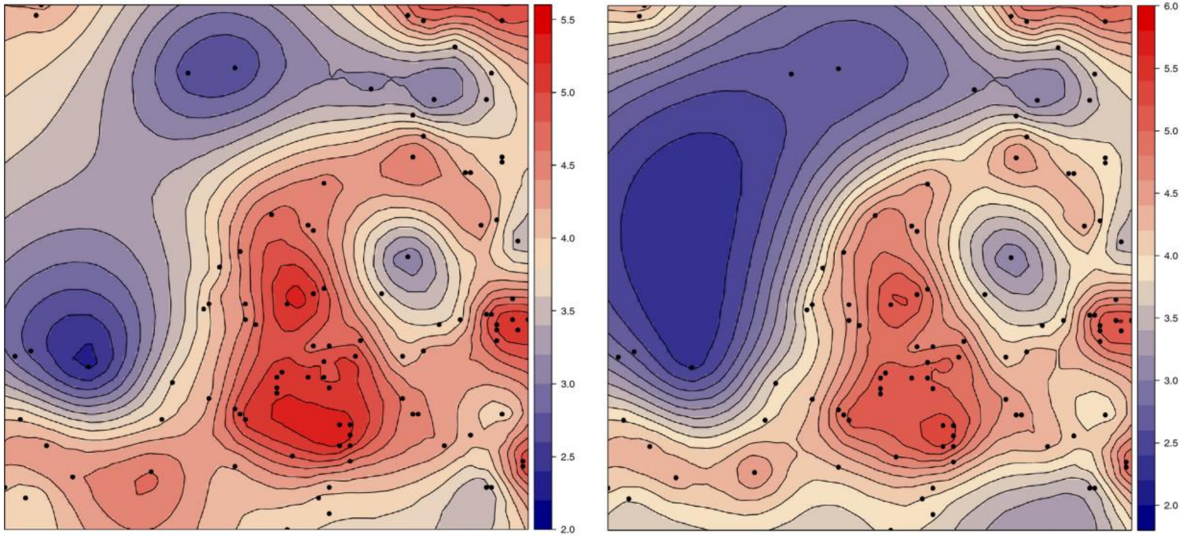
Spatio-temporal abundance maps for *Merluccius merluccius* recruits in the northeast Atlantic

Izquierdo, F., Paradinas, I., Cerviño, S., Conesa, D., Alonso-Fernández, A., Velasco, F., ... & Pennino, M. G. (2021). Spatio-temporal assessment of the European hake (Merluccius merluccius) recruits in the northern Iberian Peninsula.Frontiers in Marine Science,8, 1.

Why go Bayesian?

4

Allow to correct sampling error (e .g. preferential sampling)



(a) Non-preferential sampling model

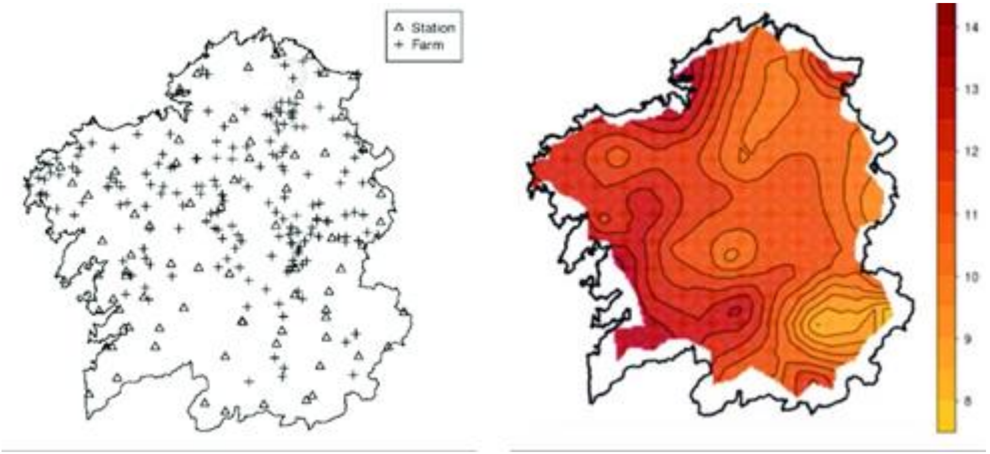
(b) Preferential sampling model

Pennino, M. G., Paradinas, I., Illian, J. B., Muñoz, F., Bellido, J. M., López-Quílez, A., & Conesa, D. (2019). Accounting for preferential sampling in species distribution models.Ecology and evolution,9(1), 653-663.

Why go Bayesian?

5

Allow to model more complicated data and situations



Misalignment problem: the sixty-seven official weather stations in Galicia do not coincide with the farms where data were observed.

Barber, X., Conesa, D., Lladosa, S., & López-Quílez, A. (2016). Modelling the presence of disease under spatial misalignment using Bayesian latent Gaussian models.Geospatial health,11(1).

Which tool should be used?

The **Integrated Nested Laplace Approximation (INLA)** is a method for approximate Bayesian inference.

In the last years it has established itself as an alternative to other methods such as Markov chain Monte Carlo because of its speed and ease of use via the **R INLA package**.

Which tool should be used?

The spatial component is a continuous Gaussian field (GF) with Matérn covariance and is approximated by a Gaussian Markov random fields (GMRF) which is a solution of a stochastic partial differential equation (SPDE).

A very popular correlation function is the Matérn correlation function. It has a scale parameter $\kappa > 0$ and a smoothness parameter $\nu > 0$. For two locations $\mathbf{s}_i$ and $\mathbf{s}_j$, the stationary and isotropic Matérn correlation function is:

$$Cor_M(U(\mathbf{s}_i), U(\mathbf{s}_j)) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel)^\nu K_\nu(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel)$$

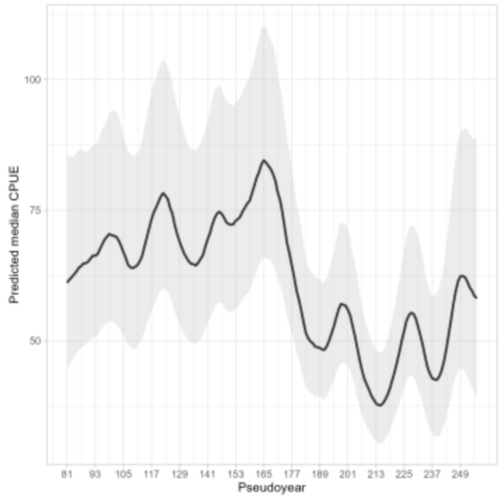
where $\|.\|$ denotes the Euclidean distance and K_ν is the modified Bessel function of the second kind. The Matérn covariance function is $\sigma_u^2 Cor_M(U(\mathbf{s}_i), U(\mathbf{s}_j))$, where σ_u^2 is the marginal variance of the process.

INLA

The benefit is that the GMRF representation of the GF, which can be computed explicitly, provides a sparse representation of the spatial effect through a sparse covariance matrix.

Fisheries science applications

Catch per unit effort (CPUE) standarization

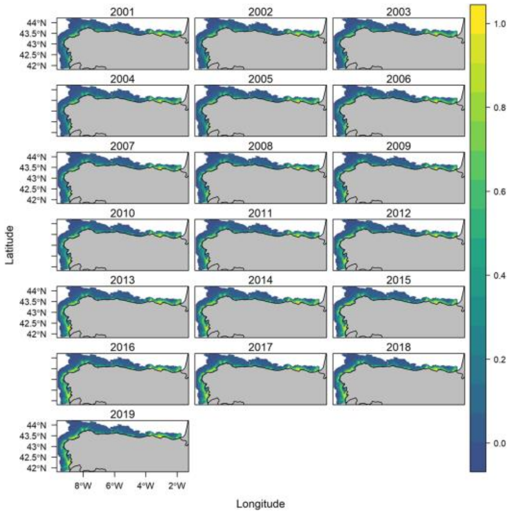


CPUE's are crucial for estimating fish population biomass.

CPUE's can be influenced by several factors as vessel size, engine power, bathymetry or location/area (*standarization process*).

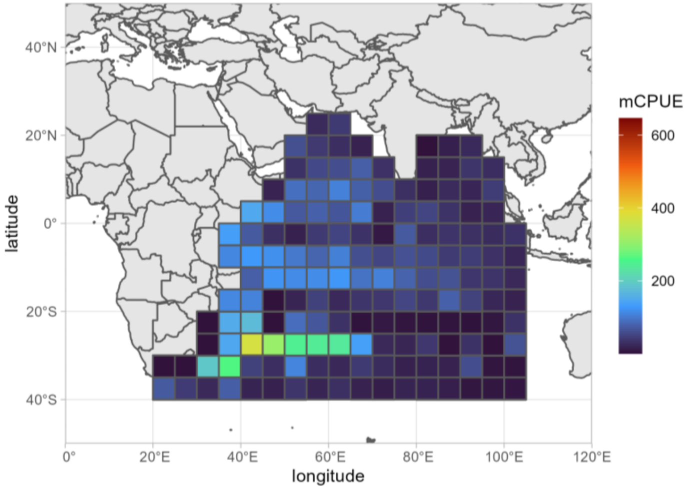
Species distribution models (SDMs)

SDMs link species occurrence and abundance to unobserved spatiotemporal autocorrelation effects and to habitat characteristics.



Catch per unit effort (CPUE) standarization

Simulated CPUE



"The unstandardized CPUE corresponded to the number of fish caught divided by the number of hooks in each cell (N/hooks)"

Spatiotemporal data

- Continuous response variable: non standardized CPUE (no 0's)
- Regular lattice 221 cells (5x5 degrees)
- Pseudoyear time variable (year and season).

Catch per unit effort (CPUE) standarization

In order to improve the index we fitted Bayesian spatiotemporal models for lattice data with R INLA

$$Z_{st} \sim Gamma(\mu_{st}, \phi)$$

Z_{st} represents the CPUE at cell s and time t, and μ_{st} and, ϕ the mean and variance of its gamma distribution.

$$\log(\mu_{st}) = \alpha + g(t) + U_{st}$$

$g(t)$ corresponds to the temporal trend fitted through a RW2 effect over the years.

$$U_{st} = W_{st} + \rho U_{s,t-1}, \; W_{st} \sim N(0, \Sigma)$$

U_{st} refers to spatiotemporal structure.

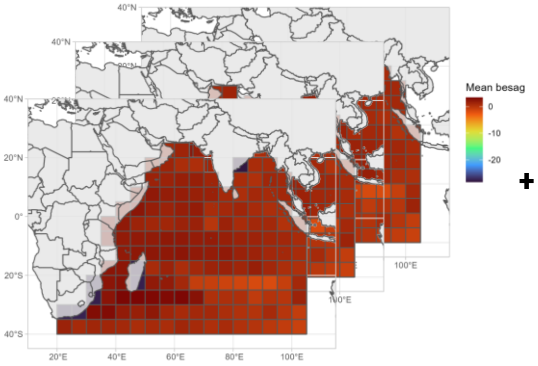
Σ is the mátern variance-covariance matrix.

Catch per unit effort (CPUE) standarization

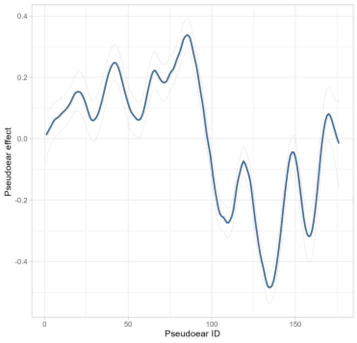
We took 200 samples of the posterior distribution for all model components

Mean spatiotemporal effect

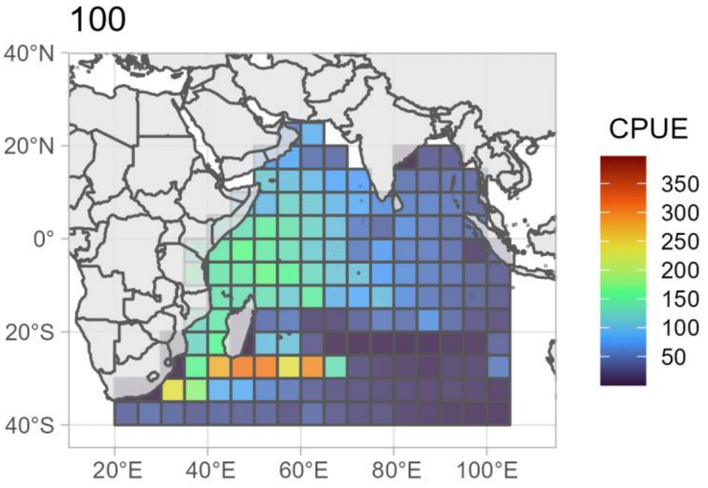
Time smoothed (RW2) trend



+



=

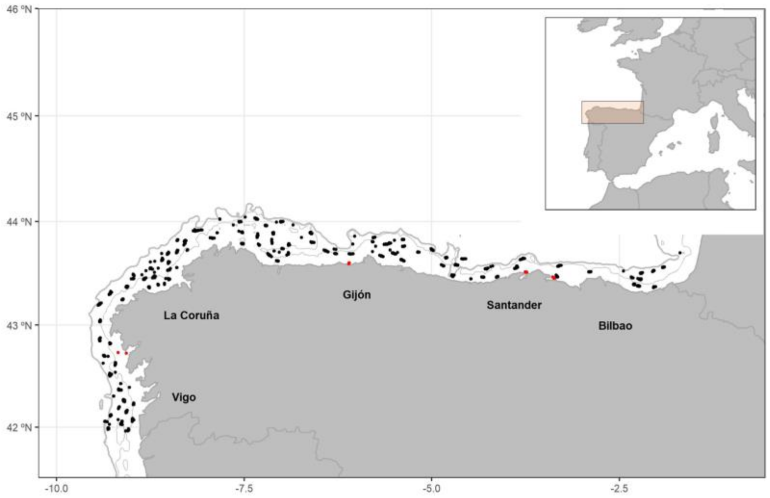


Species distribution models (SDMs)

Study of **the spatiotemporal distribution of common sole** (*Solea solea*) in the northern Iberian waters

Data

- Common sole data was collected during the **scientific survey** series performed by IEO between 2001 and 2019.
- **Presence/absence variable** was considered to measure the occurrence probability of the species.
- **The weight by haul (kg)** was used as an indicator of the conditional-to-presence biomass of the species.



Species distribution models (SDMs)

$$Y_{st} \sim Ber(\pi_{st})$$

$$Z_{st} \sim Gamma(\mu_{st}, \phi)$$

$$logit(\pi_{st}) = \alpha^{(Y)} + g(t) + \sum_{j=1}^{k} f_j(X_j) + U_{st}^{(Y)}$$

$$log(\mu_{st}) = \alpha^{(Z)} + \theta g(t) + \sum_{j=1}^{k} \eta_j f_j(X_j) + U_{st}^{(Z)}$$

Y_{st} occurence and Z_{st} conditional-to-presence biomass.

π_{st} probabity of occurrence at location s and time t.

μ_{st} and ϕ the mean and dispersion conditional-to-presence biomass.

Species distribution models (SDMs)

$$Y_{st} \sim Ber(\pi_{st})$$

$$Z_{st} \sim Gamma(\mu_{st}, \phi)$$

$$logit(\pi_{st}) = \alpha^{(Y)} + g(t) + \sum_{j=1}^{k} f_j(X_j) + U_{st}^{(Y)}$$

$$logit(\mu_{st}) = \alpha^{(Z)} + \theta g(t) + \sum_{j=1}^{k} \eta_j f_j(X_j) + U_{st}^{(Z)}$$

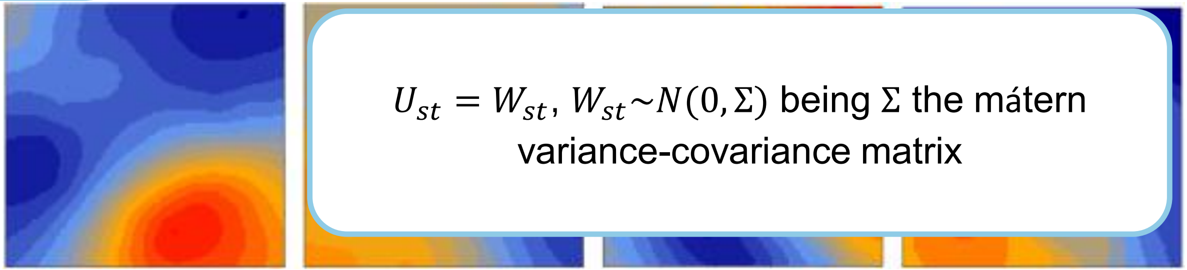
$g(t)$ the temporal trend as RW2 effect over the years.

$f_j(X_j), j = 1, \dots, k,$ functions of extra covariables X_j.

$U_{st}^{(Y)}$ and $U_{st}^{(Z)}$ the spatiotemporal structures.

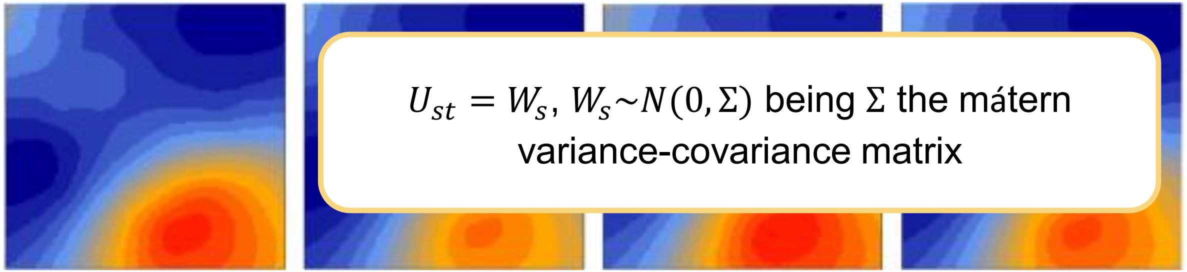
Species distribution models (SDMs)

Opportunistic



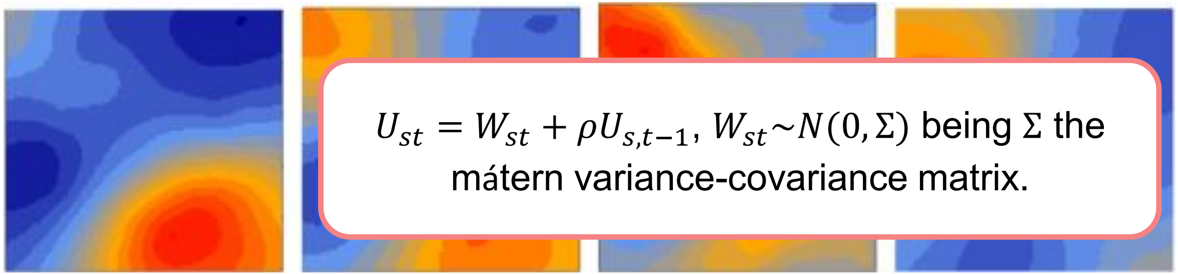
$U_{st} = W_{st}$, $W_{st} \sim N(0, \Sigma)$ being Σ the mátern variance-covariance matrix

Persistent



$U_{st} = W_s$, $W_s \sim N(0, \Sigma)$ being Σ the mátern variance-covariance matrix

Progressive



$U_{st} = W_{st} + \rho U_{s,t-1}$, $W_{st} \sim N(0, \Sigma)$ being Σ the mátern variance-covariance matrix.

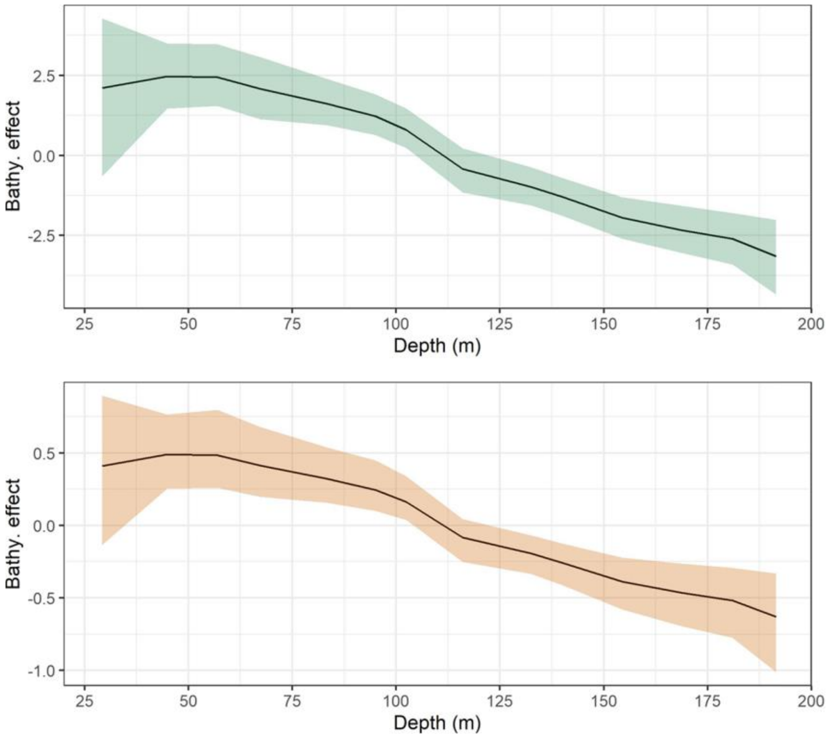
Species distribution models (SDMs)

Model	WAIC	LCPO
Persistent	1732	0.52
Opportunistic	1770	0.54
Progressive	1728	0.61

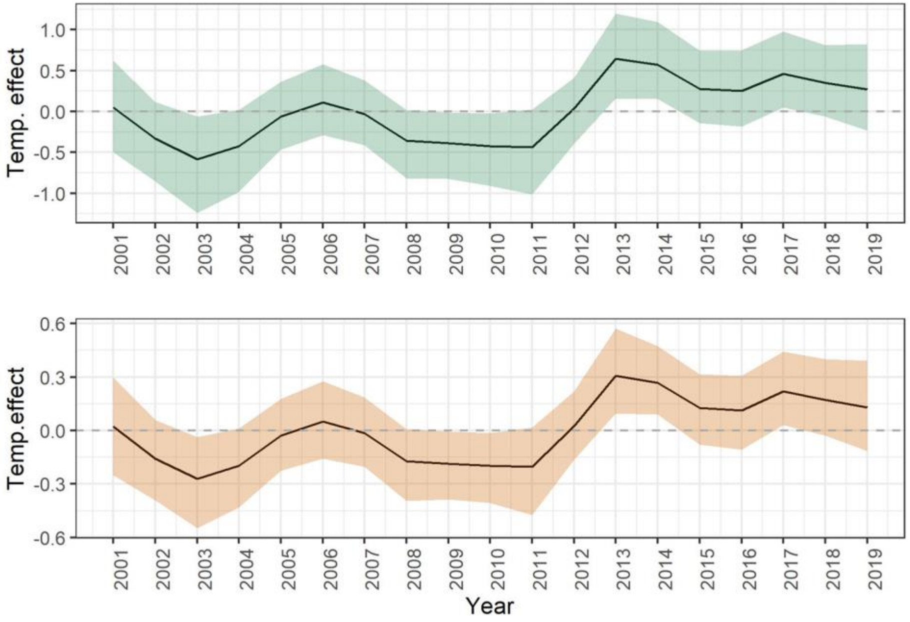
$*$ Watanabe Akaike information criterion (WAIC) and log-conditional predictive ordinates (LCPO).

Autoregressive parameter values: ρ = 0.98 and 0.96 for the occurrence and conditional-to presence-biomass processes, respectively.

Species distribution models (SDMs)

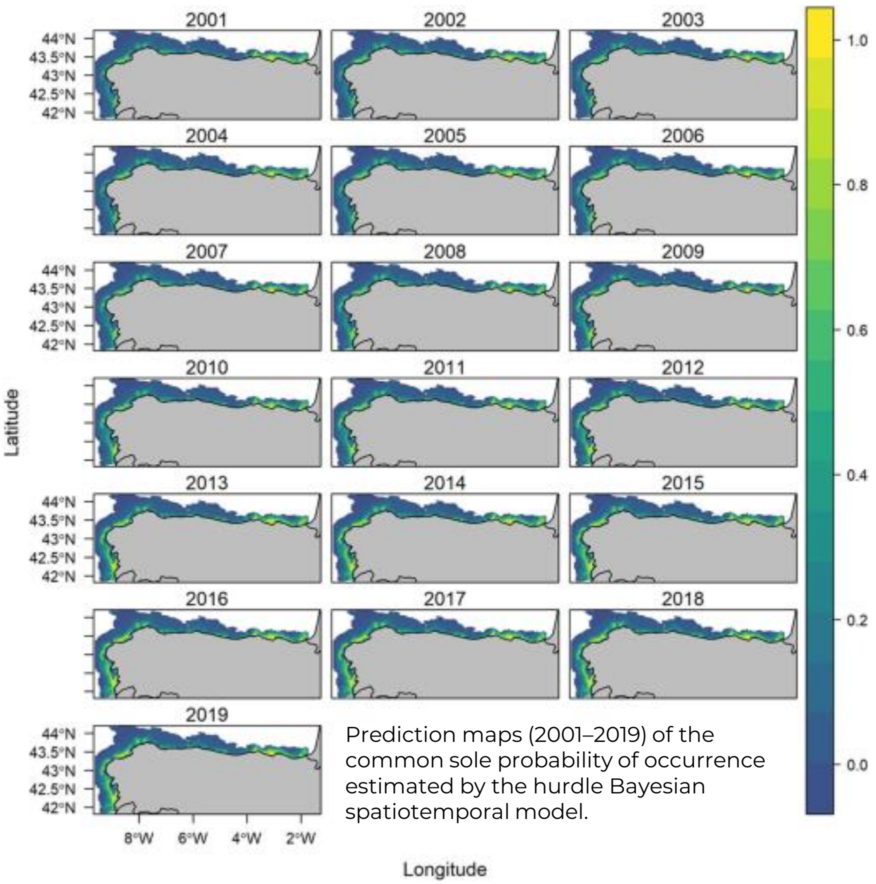


Predicted occurrence (top) and conditional-presence-abundance (bottom) for **the bathymetry effect**. Shaded regions represent the approximate 95% credibility interval.

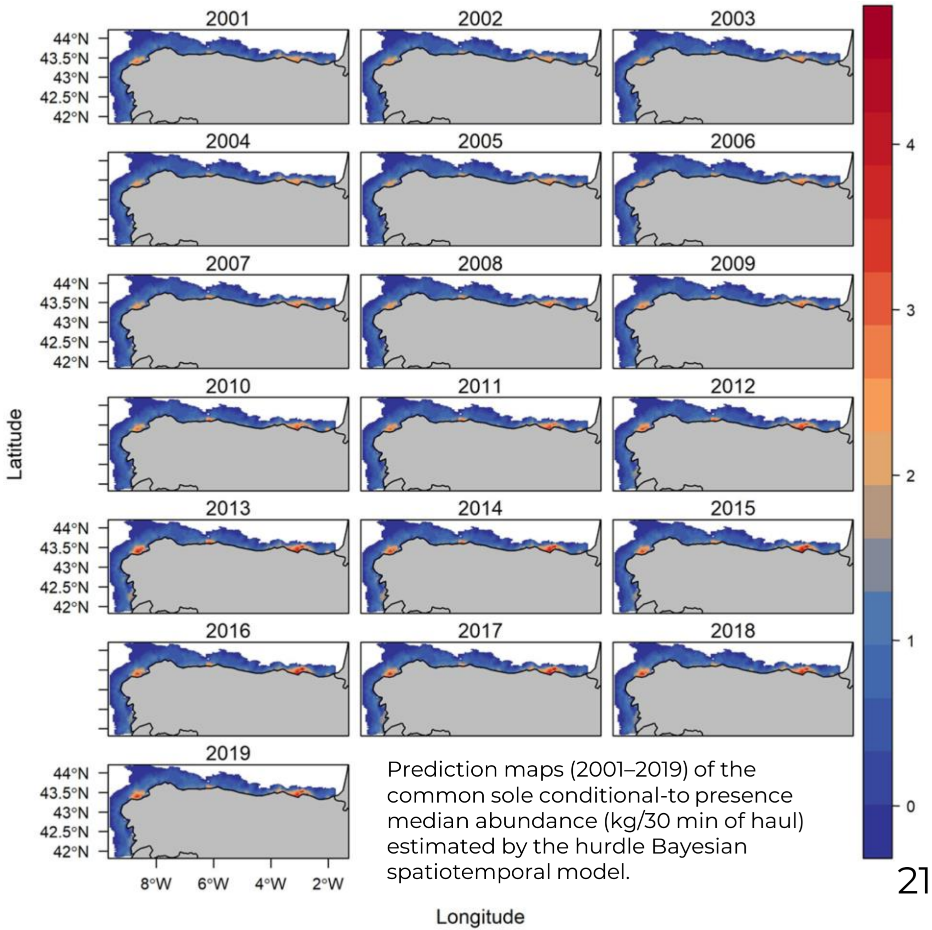


Marginal temporal shared effect in the linear predictor scale (logarithmic link) of common sole occurrence (higher panel) and conditional-to-presence-abundance processes (lower panel

Species distribution models (SDMs)



Prediction maps (2001–2019) of the common sole probability of occurrence estimated by the hurdle Bayesian spatiotemporal model.



Prediction maps (2001–2019) of the common sole conditional-to presence median abundance (kg/30 min of haul) estimated by the hurdle Bayesian spatiotemporal model.

These examples serve as an evidence that INLA Bayesian spatiotemporal models are a powerful tool for addressing complex spatio-temporal challenges in a wide range of fields.

Thank you for your attention!

Instituto Español de Oceanografía (IEO, CSIC).
Centro Oceanográfico de Vigo.
Subida a Radio Faro, 50-52, 36390, Vigo, Pontevedra, Spain.

marta.cousido@ieo.csic.es

Pennino, M. G., Izquierdo, F., Paradinas, I., Cousido-Rocha, M., Velasco, F., Cerviño, S. (2022). Identifying persistent biomass areas: The case study of the common sole in the northern Iberian waters, Fisheries Research, 248, 106196.

https://www.r-inla.org/what-is-inla