
Small Pelagic Fish:
New Frontiers in Science and Sustainable Management
November 7 - 11, 2022
Lisbon, Portugal
ICES CIEM
Food and Agriculture Organization of the United Nations
ENDORSED BY
2021 2030 United Nations Decade of Ocean Science for Sustainable Development

Centre of Mathematics (CMAT), University of Minho

Spatio-temporal variability of the distribution and abundance of small pelagic fish off the Portuguese continental coast and relationship with environmental drivers

Daniela Silva
danyelasylva2@gmail.com

with
Raquel Menezes, University of Minho
Ana Moreno, IPMA
Ana Teles-Machado, IPMA
Susana Garrido, IPMA

CMAT Centro de Matemática UNIVERSIDADE DO MINHO
Universidade do Minho Escola de Ciências
FCT Fundação para a Ciência e a Tecnologia
IPMA
pnab programa nacional de amostragem biológica

November, 10

Scope of the work

Loss of marine biodiversity

Improve the knowledge about population dynamics of main fishing resources

Restoration of marine biodiversity

Spatio-temporal analysis of abundance indexes

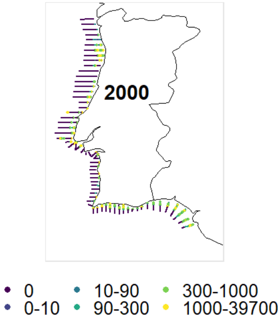
Motivating data

PELAGO surveys

Estimate the abundance of sardine inhabiting the Portuguese shelf.

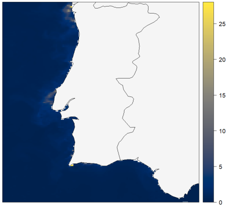
Database

Time:	20 annual surveys from 2000 to 2020, except 2012.
Study region:	Portuguese continental coast and Gulf of Cadiz.
Interest:	Biomass index (NASC, $n^2\ nm^{-2}$).
Geographic information:	Latitude and longitude.
Additional information:	Sectors and geographical areas.

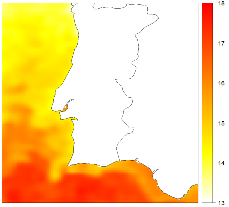


Environmental data

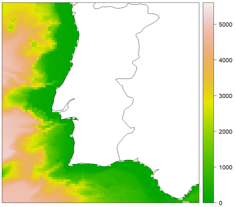
Chlorophyll



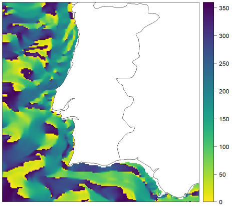
Temperature



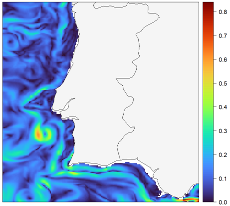
Bathymetry



Direction of ocean currents



Intensity of ocean currents



Problem and Objective

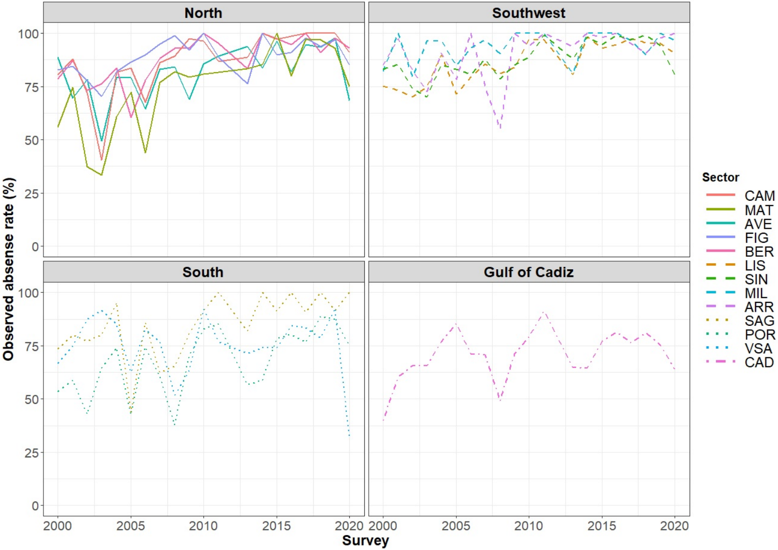
Species Distribution Data often implies residual **spatial autocorrelation**

- **Non-consideration** of important environmental conditions.
- **Intrinsic factors**: competition, dispersal, aggregation, etc.

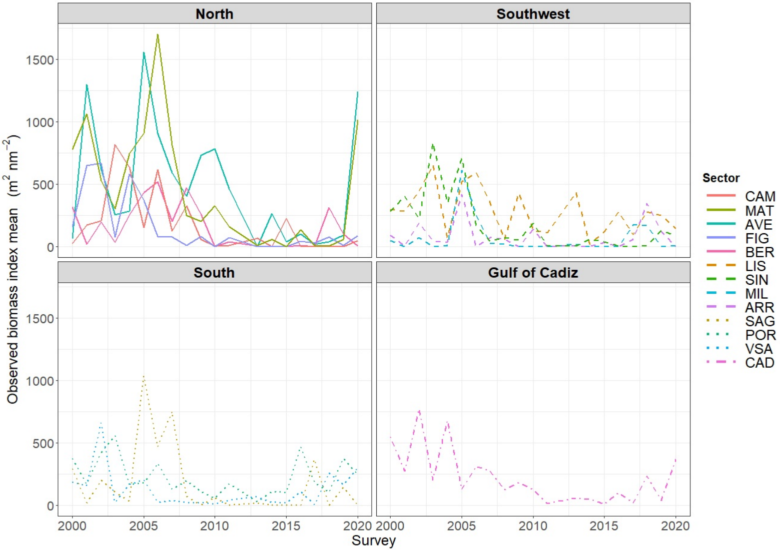
↓

Geostatistics

Problem and Objective



Problem and Objective



Problem and Objective

Main aims

- Estimate the **spatio-temporal distribution** of sardine in western and southern Iberian waters.
- Understand sardine dynamics over time and space.
- Identify the **main drivers** of sardine spatial dynamics.

To consider:

- Complex spatio-temporal dynamics.
- Excess of zeros.
- Difference between occurrence process and biomass process under occurrence.
- Relationship between response and **environmental conditions with a time lag**.

Two-part model

Hierarchical model

Series of levels linked by probability functions.

Y_{st} - **Biomass process** at location s and time t
Z_{st} - **Occurrence sub-process**

Distribution of biomass index

$$
\begin{aligned}
[Y_{st}] &= [Z_{st}]\,[Y_{st}|(Z_{st}=1)] \\
&= \begin{cases} 1-\pi_{st}, & y_{st}=0 \\ \pi_{st}\,[Y_{st}|(Z_{st}=1)], & y_{st}>0 \end{cases}
\end{aligned}
$$

such that:

$$
\begin{aligned}
Z_{st} &\sim Bernoulli(\pi_{st}) \\
Y_{st}|(Z_{st}=1) &\sim Gamma(a_{st}, b_{st})
\end{aligned}
$$

Two-part model

Two-part model can be defined by:

$$log(\mu_{sti}) = \alpha + \sum_{j=1}^{p} \mathbf{f}(K(X_{jsti}, c, l)) + \gamma_t + W_{st}$$

$$logit(\pi_{sti}) = \alpha' + \sum_{j=1}^{p'} \mathbf{f}'(K(X'_{jsti}, c, l)) + \gamma'_t + kW_{st}$$

time lag $c + l$ in days from i^{th} day of the survey in year t,
smoother function $\mathbf{f}$ of the j^{th} covariate X_{jsti},
spatio-temporal structure W_{st},
unstructured temporal effects γ_t and γ'_t.

Spatio-temporal structure with AR(1) process

$$W_{\mathbf{s}t} = \delta W_{\mathbf{s}(t-1)} + \xi_{\mathbf{s}t}$$

1. $|\delta| < 1$
2. $\xi_{\mathbf{s}t}$ is is a zero-mean GF with spatio-temporal covariance:

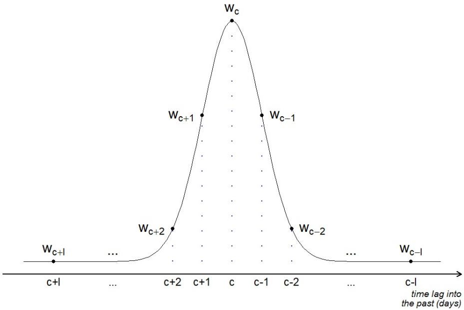
$$Cov(\xi_{\mathbf{s}t}, \xi_{\mathbf{u}j}) = \begin{cases} 0 & if \quad t \neq j \\ Cov(\xi_{\mathbf{s}}, \xi_{\mathbf{u}}) & if \quad t = j \end{cases}$$

such that $Cov(\xi_{\mathbf{s}}, \xi_{\mathbf{u}})$ is given by **Matérn spatial covariance** with partial variance σ^2 and range ϕ.

Kernel application $K(.,.,.)$

$$K(X_{jsti}, c, l) = \sum_{q=-l}^{l} w_{c-q} X_{jst(i-(c-q))}$$

- $w_{c-q} = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{q^2}{2h^2}\right\}$
- $X_{jst(i-(c-q))}$ - jth covariate observed in day $i-(c-q)$ of year t
- On the i^{th} day, the maximum effect of X_j occurs for lag c

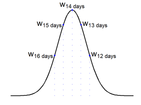


Examples of Kernel application

Presence/absence modelling:

$$logit(\pi_{sti}) = \alpha_1 + \sum_{j=1}^{p'} \mathbf{f}(K(X'_{jsti}, c, l)) + \gamma'_t + kW_{st}$$

1. If the biomass **is only affected by chlorophyll on the same day**, then $K(CHL_{sti}, 0, 0) = CHL_{sti}$
2. If the biomass **is affected by chlorophyll 14 days ago**, then $K(CHL_{sti}, 14, 0) = CHL_{st(i-14)}$
3. If the biomass **is affected by chlorophyll 14 days earlier and 2 days before and after**, then $K(CHL_{sti}, 14, 2) = w_{16}CHL_{st(i-16)} + w_{15}CHL_{st(i-15)} + w_{14}CHL_{st(i-14)} + w_{13}CHL_{st(i-13)} + w_{12}CHL_{st(i-12)}$



Bayesian perspective

Advantages

- It allows to **incorporate prior information**.
- **Information and uncertainty** about all the unknown can be better (and easily) **expressed in terms of probability distributions**.
- It might more **easily handle with inference and prediction** (Banerjee, Carlin, and Gelfand 2004).

Inference method

INLA approach was used to approximating the posterior marginals of the latent GF (Rue, Martino, and Chopin 2009).

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

Some highlights

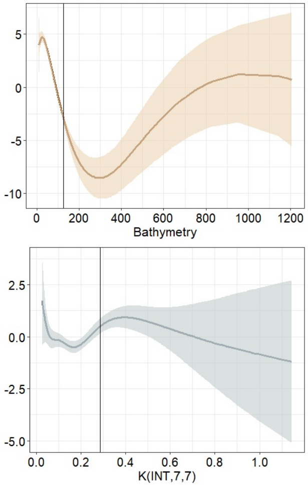
- Various **combinations of c and l** were tested.

- Data from the west and south Iberian coasts are **studied separately**.

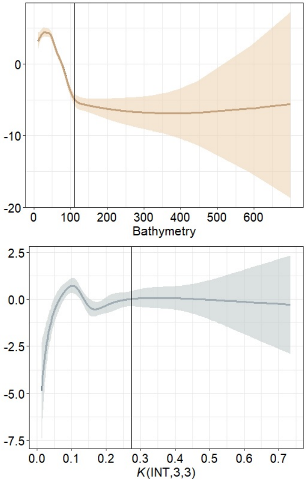
- Spatial predictions over the entire study region were obtained **for a "representative day"** of each survey for the total 21 years.

Results: Environmental effects for the presence

Effects on west coast



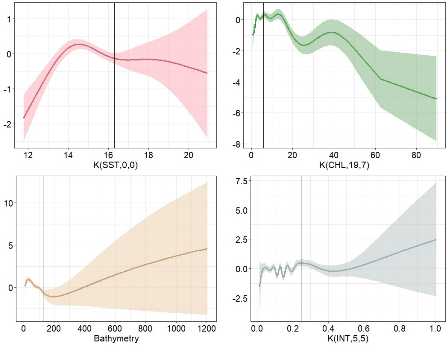
Effects on south coast



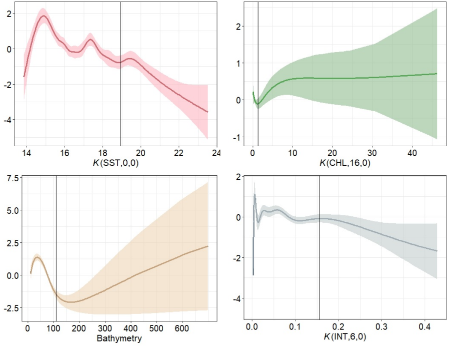
**vertical line - quantile 80%*

Results: Environmental effects for the biomass

Effects on west coast

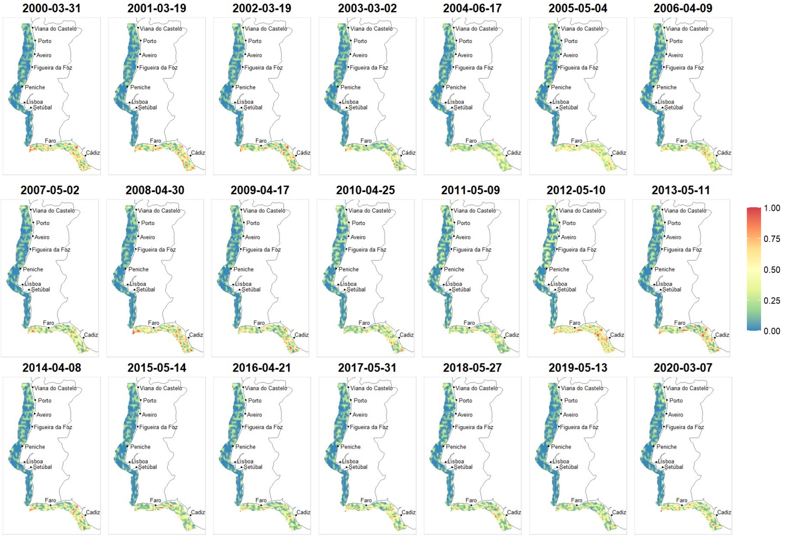


Effects on south coast

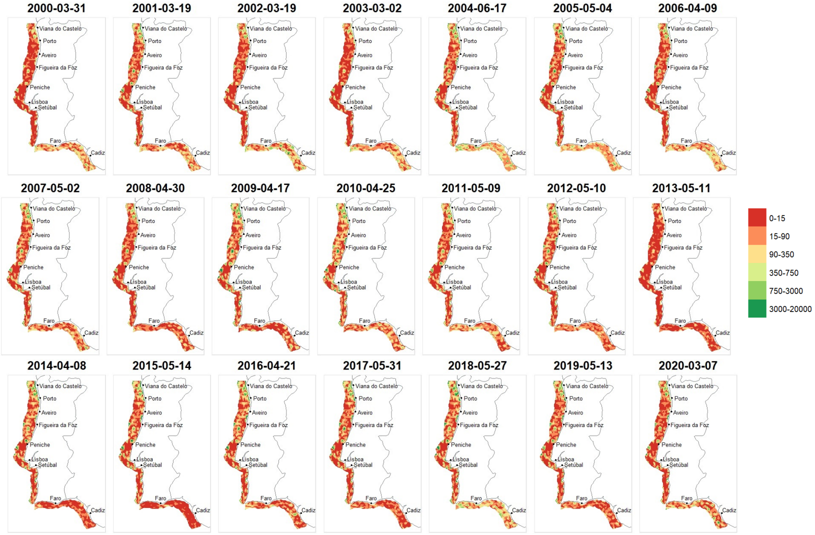


**vertical line - quantile 80%*

Results: Predicted occurrence



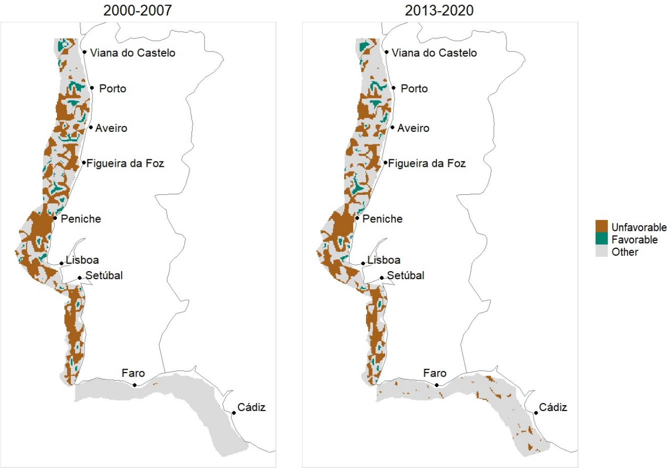
Results: Predicted biomass



Results: Occupancy areas

Results: Favourable and unfavourable zones

Persistent rare (unfavourable) and preferred (favourable) zones



On going and Future work

On going work

- Apply this methodology to the **anchovy data**.

Future work

- Model the spatio-temporal distribution of sardine from data obtained from commercial fisheries, taking into account **preferential sampling**.
- **Joint modelling** fishery-dependent and fishery-independent data.

Thank you for your attention!

This study was supported by Portuguese funds through the Centre of Mathematics and the Portuguese Foundation for Science and Technology (FCT), within Individual Scholarship PhD PD/BD/150535/2019 and projects UIDB/00013/2020, UIDP/ 00013/2020 and PTDC/MAT-STA/28243/2017.

References I

Sudipto Banerjee, Bradley Carlin, and Alan Gelfand (Jan. 2004). *Hierarchical Modeling and Analysis of Spatial Data*. Vol. 101. Chapman & Hall/CRC Monographs on Statistical and Applied Probability. DOI: 10.1201/9780203487808.

James S. Clark and Alan Gelfand (2006). *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*. USA: Oxford University Press, Inc. ISBN: 0198569661.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed (1998). "Model-based geostatistics. (With discussion).". In: *J. R. Stat. Soc., Ser. C, Appl. Stat.* 47.3, pp. 299–350. ISSN: 0035-9254; 1467-9876/e.

Joaquín Martínez-Minaya et al. (2018). "Species distribution modeling: a statistical review with focus in spatio-temporal issues". In: *Stochastic Environmental Research and Risk Assessment*. DOI: 10.1007/s00477-018-1548-7.

References II

Håvard Rue, Sara Martino, and Nicolas Chopin (Apr. 2009). "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations". In: *Journal of the Royal Statistical Society Series B* 71, pp. 319–392. DOI: 10.1111/j.1467-9868.2008.00700.x.