Portuguese Institute for the Sea and Atmosphere (IPMA)
Centre of Mathematics (CMAT), University of Minho

Integrating fishery-dependent and independent data to model sardine distribution under environmental variability in Portuguese waters

Daniela Silva
daniela.dasilva@ipma.pt

[Joint work with: R. Menezes, G. Araújo, A. Machado, R. Rosa, A. Moreno, A. Silva, S. Garrido]

IPMA Instituto Português do Mar e da Atmosfera
MAR 2030
pnab programa nacional de amostragem biológica
Universidade do Minho Escola de Ciências

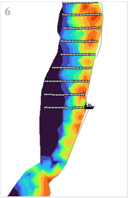
Nov 2025, 8-14

Fishery data sources & Challenges

Fishery-independent data (FID) - scientific surveys

- Standardized/Deterministic sampling.
- Wide spatial coverage.
- Short time span.
- Many zeros.

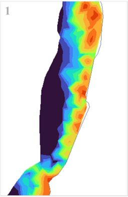
Our scientific eye - broad but limited in time.



Fishery-dependent data (FDD) - commercial fisheries

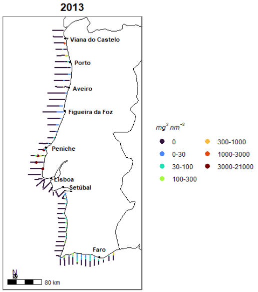
- Fishing activity (e.g., logbooks and AIS).
- Long time span.
- Short spatial coverage.
- Preferential sampling (PS).

Fishers' eye - detailed but biased toward fishing grounds.

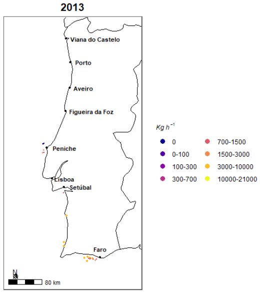


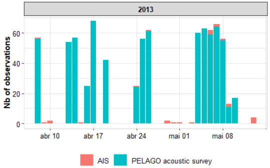
How can we combine broad but sparse surveys with dense but biased fishing data to better understand fish distributions?

PELAGO acoustic survey



AIS

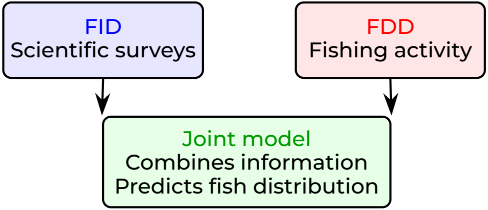




Goal: From Two Views of the Ocean to One Picture

Infer the spatio-temporal distribution of fish through a joint model that

- Combine standardized surveys and biased but detailed fishing data into a single modeling framework to obtain a coherent view of sardine distribution and abundance.



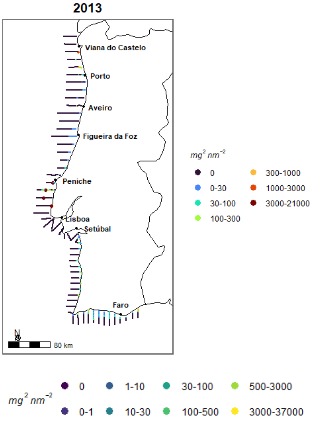
- Accounts for zero-inflation, PS and vessel catchability
- Deals with distinct scales of biomass index

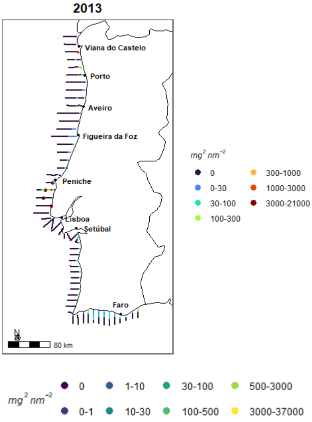
FID

- ▷ PELAGO 2013-2017 survey series.
- ▷ 2362 sardine NASC values ($mg\ m^{-2}$).

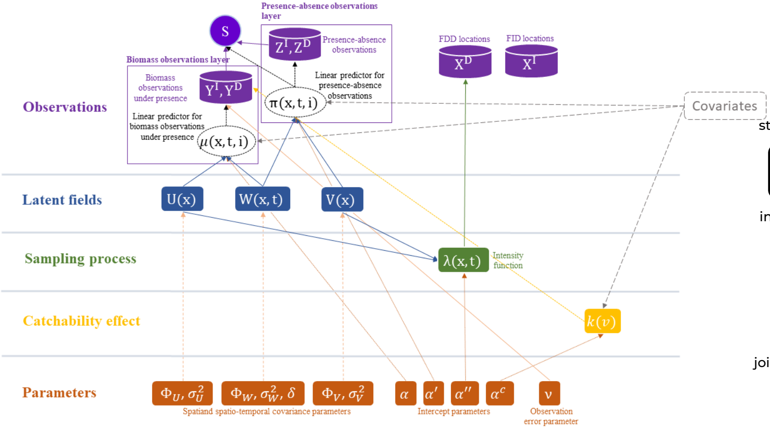
FDD

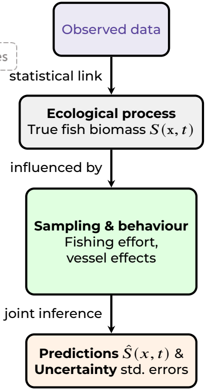
- ▷ Commercial data obtained through AIS.
- ▷ 1687 sardine biomass values ($Kg\ h^{-1}$).





Hierarchy of Processes





- Layers build from $data \rightarrow latent\,fields \rightarrow sampling \rightarrow catchability \rightarrow likelihood.$
- Allows us to integrate sources with **different biases and scales**.
- Ensures **realistic fish distribution estimates with uncertainty**.

Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:

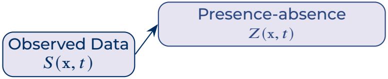
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:

Observed Data
$S(\mathbf{x}, t)$

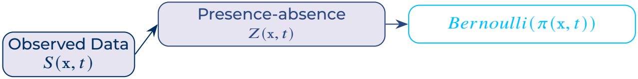
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



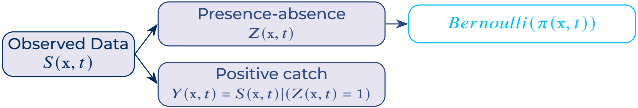
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



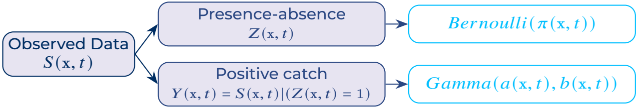
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



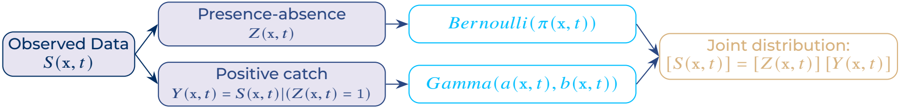
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



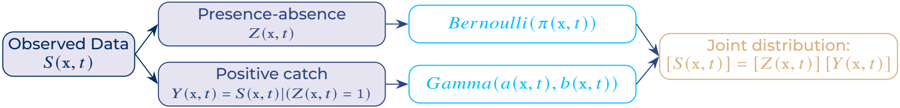
Observations layers

$S(\mathbf{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



Observations layers

$S(\mathrm{x}, t)$ denotes the **spatio-temporal process of interest** (biomass/abundance index) for location $\mathrm{x} \in \mathcal{A} \subset \mathbb{R}^2$. and time $t = \{t_1, \cdots, t_T\}$.
To handling zero-inflation, two sub-processes are generated:



Layers:

1 PAP observations:

$$logit(\pi(\mathbf{x}, t, i)) = \alpha' + \sum_{j=1}^{p'} f'(K(C'(j, \mathbf{x}, t, i), c, l)) + V(\mathbf{x}) + W(\mathbf{x}, t)$$

2 Biomass observations:

$$log(\zeta(\mathbf{x}, t, i)) = \alpha + \sum_{j=1}^{p} f(K(C(j, \mathbf{x}, t, i), c, l)) + U(\mathbf{x}) + W(\mathbf{x}, t)$$

- **i-th subperiod** (e.g., day) within time t (e.g, year).
- **intercepts**: α' and α.
- **smoother effects** $f(.)$ and $f'(.)$ of **covariates** $C(.)$ and $C'(.)$.
- **time-lagged effects** $K()$ of the covariates (*Silva et al., 2024*).

Latent fields layer

❶ **PAP observations:**

$$logit(\pi(\mathbf{x}, t, i)) = \alpha' + \sum_{j=1}^{p'} f'(K(C'(j, \mathbf{x}, t, i), c, l)) + V(\mathbf{x}) + W(\mathbf{x}, t)$$

❷ **Biomass observations:**

$$log(\zeta(\mathbf{x}, t, i)) = \alpha + \sum_{j=1}^{p} f(K(C(j, \mathbf{x}, t, i), c, l)) + U(\mathbf{x}) + W(\mathbf{x}, t)$$

Latent fields layer

❶ PAP observations:

$$logit(\pi(\mathbf{x},t,i)) = \alpha' + \sum_{j=1}^{p'} f'(K(C'(j,\mathbf{x},t,i),c,l)) + V(\mathbf{x}) + W(\mathbf{x},t)$$

❷ Biomass observations:

$$log(\zeta(\mathbf{x},t,i)) = \alpha + \sum_{j=1}^{p} f(K(C(j,\mathbf{x},t,i),c,l)) + U(\mathbf{x}) + W(\mathbf{x},t)$$

❸ Latent fields layer:

- $W(\mathbf{x},t)$: shared **spatio-temporal structure** based on a **first-order autoregressive process**:

  $$W(\mathbf{x},t) = \delta W(\mathbf{x},t-1) + \xi(\mathbf{x},t)$$

  Each year's spatial field depends on the previous one.

- $U(\mathbf{x}), V(\mathbf{x})$: **spatial structure associated**.
- $\xi, \mathbf{U}, \mathbf{V} \sim \mathrm{GRF}(0, \Sigma_F(\phi_F, \sigma_F^2))$ [or a Barrier model (*Bakka et al., 2019*) when the study region presents a peluliar shape or physical barriers].

  Spatial correlation

  $$\mathrm{Cov}[F(\mathbf{x}), F(\mathbf{x}')] = \sigma_F^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|}{\phi_F}\right)$$

  Nearby locations have similar abundances.

❹ Sampling process:

Data source:	FID	FDD*
Type:	Homogeneous Poisson	Inhomogeneous Poisson
Nature:	Random/Systematic	Preferential
Modeled by:	$\mathrm{X}^I(t) \sim HPP(\lambda^{HPP}(t))$	$\mathrm{X}^D(t) \sim IPP(\lambda(\mathrm{x}^D, t))$ $log(\lambda(\mathrm{x}^D, t)) = \alpha''(t) + \beta'(t)V(\mathrm{x}) + \beta(t)U(\mathrm{x})$
Drivers:	None	**Dependent** on U and V

*Notes:

- Following *Diggle et al. (2010)*.
- $\beta'(t)$ and $\beta(t)$ quantify the degree of spatial PS by scaling the relationship between the local fishing intensity and the local value of each process of interest $Z(., t)$ and $Y(., t)$ for time t.

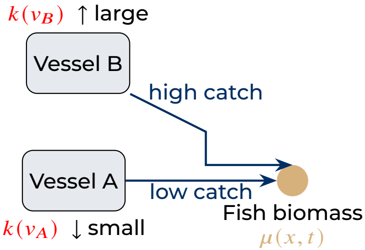
Catchability effects layer

Issue: **Different vessels have different efficiency** (e.g., gear, size, technology, crew).

⑤ Catchability effect:

$$\zeta(\mathbf{x}, t, i, v) = k(v) \times \mu(\mathbf{x}, t, i)$$

- with the **expected relative biomass** $\mu(\mathbf{x}, t, i)$ (where the relative biomass process $\mathbf{S}^* = \mathbf{Z} \cdot \mathbf{Y}^*$),
- and expected biomass $\zeta(\mathbf{x}, t, i, v)$



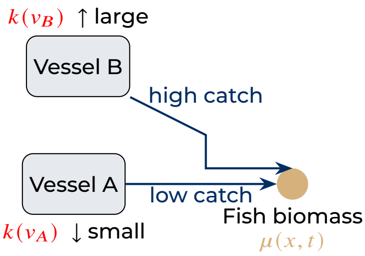
Catchability effects layer

Issue: **Different vessels have different efficiency** (e.g., gear, size, technology, crew).

5 Catchability effect:

$$\zeta(\mathbf{x}, t, i, v) = k(v) \times \mu(\mathbf{x}, t, i)$$

- with the **expected relative biomass** $\mu(\mathbf{x}, t, i)$ (where the relative biomass process $\mathbf{S}^* = \mathbf{Z} \cdot \mathbf{Y}^*$),
- and expected biomass $\zeta(\mathbf{x}, t, i, v)$



The catchability for vessel v is given by

$$k(v) = exp\{\alpha_c + \sum_{h=1}^{H} f_c(F(h, v)) + \gamma_c(v)\}$$

- ▷ α_c: intercept,
- ▷ f_c: smoother term for fixed effects F (vessel attributes),
- ▷ $\gamma_c \sim Normal(0, \sigma^2_{\gamma_c})$: vessel-specific random effect (i.i.d.)

Inference & Summary

Likelihood factorization:

$$\begin{aligned}\mathcal{L}(\Theta) =& \mathcal{L}(\mu, \upsilon; \mathbf{y}) \times \mathcal{L}(\pi; \mathbf{z}) \times \mathcal{L}(\lambda; \mathbf{x}) \\ &\times \mathcal{L}(\sigma_U, \phi_U) \times \mathcal{L}(\sigma_V, \phi_V) \\ &\times \mathcal{L}(\sigma_W, \phi_W, \delta)\end{aligned}$$

Likelihood factorization:

$$
\begin{aligned}
\mathcal{L}(\Theta) = & \mathcal{L}(\mu, \upsilon; \mathbf{y}) \times \mathcal{L}(\pi; \mathbf{z}) \times \mathcal{L}(\lambda; \mathbf{x}) \\
& \times \mathcal{L}(\sigma_U, \phi_U) \times \mathcal{L}(\sigma_V, \phi_V) \\
& \times \mathcal{L}(\sigma_W, \phi_W, \delta)
\end{aligned}
$$

Inference

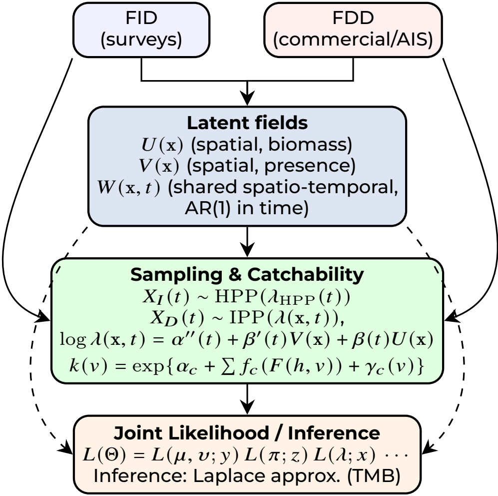
- Inference via Laplace approximation
- Implemented with Template Model Builder (TMB-**R package**)
- Likelihood coded in C++ **template functions** (*Kristensen et al., 2016*)

Likelihood factorization:

$$\mathcal{L}(\Theta) = \mathcal{L}(\mu, \upsilon; \mathbf{y}) \times \mathcal{L}(\pi; \mathbf{z}) \times \mathcal{L}(\lambda; \mathbf{x}) \times \mathcal{L}(\sigma_U, \phi_U) \times \mathcal{L}(\sigma_V, \phi_V) \times \mathcal{L}(\sigma_W, \phi_W, \delta)$$

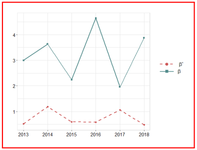
Inference

- Inference via Laplace approximation
- Implemented with Template Model Builder (TMB-**R package**)
- Likelihood coded in C++ **template functions** (*Kristensen et al., 2016*)



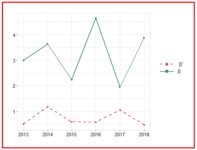
Results

Preferential effects

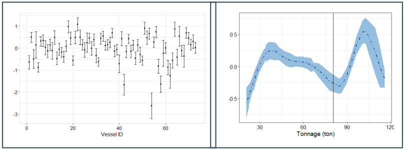


Preferential effect: Fishers tend to go where sardine are abundant.

Preferential effects



Catchability effects

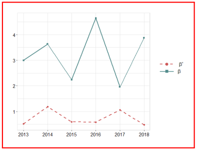


Preferential effect: Fishers tend to go where sardine are abundant.

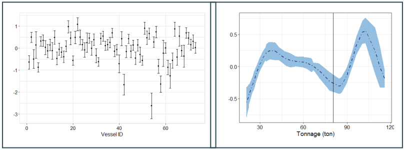
Catchability effect: Larger and better equipped vessels catch more fish.

Results

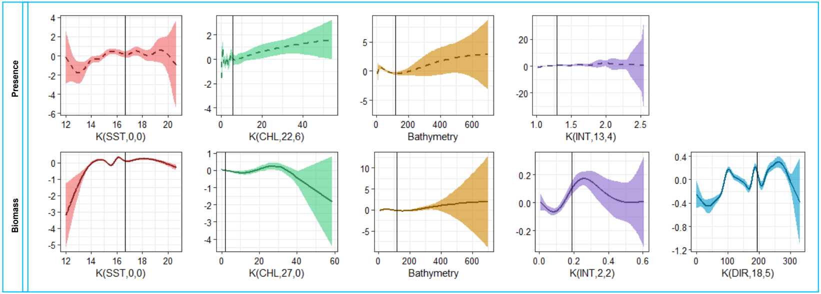
Preferential effects



Catchability effects



Environmental effects

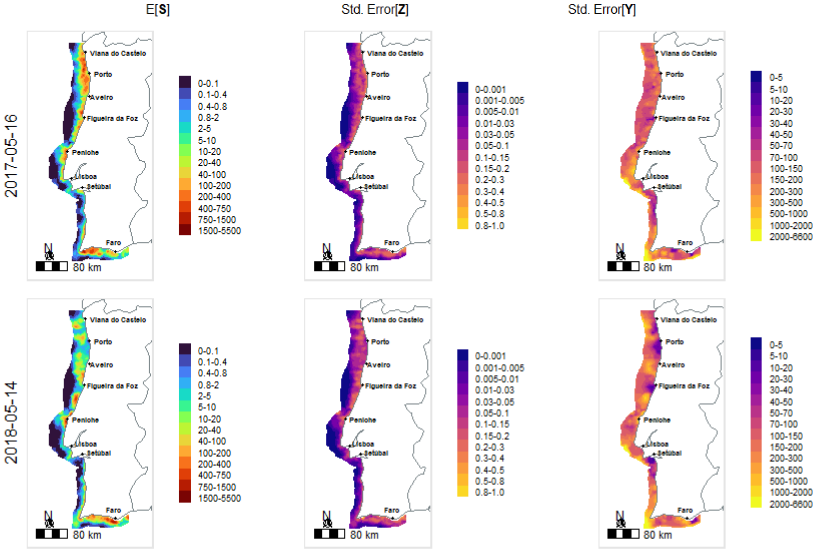


Preferential effect: Fishers tend to go where sardine are abundant.

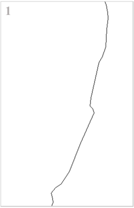
Catchability effect: Larger and better equipped vessels catch more fish.

Environmental effect: Temperature and plankton drive sardine distribution.

Results: Case study - Predictions and respective Std. error

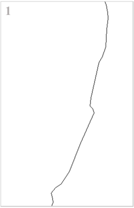


- We **integrate multiple data sources** to improve knowledge of sardine distribution.
- We **quantify uncertainty** and account for fisher behavior and gear efficiency.
- This approach can **support management decisions** and can be extended to other species.



Conclusions

- We **integrate multiple data sources** to improve knowledge of sardine distribution.
- We **quantify uncertainty** and account for fisher behavior and gear efficiency.
- This approach can **support management decisions** and can be extended to other species.



Future work

🚀 Expand the model to look at predator distribution jointly alongside SPF - **planned under Activity 1 of PICES-ICES WG53**.

Thank you for your attention

This study was supported by Portuguese funds through the **UID/00013: Centro de Matemática da Universidade do Minho (CMAT/UM)** and by the **European Maritime and Fisheries Fund** (MAR-01.04.02-FEAMP-0009) through the **SARDINHA2030 project**.

References

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics*, 29, 268-288.

Diggle, P., Menezes, R., Su, T. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Series C*, 59 (2), 191-232. https://doi.org/10.1111/j.1467-9876.2009.00701.x

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., Bell, B. (2016). TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70, 1-21. doi: 10.18637/jss.v070.i05

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. SpringerVerlag.

Silva, D., Menezes, R., Araújo, G., Rosa, R., Moreno, A., Silva, A., Garrido, S. (2025). Joint model for zero-inflated data combining fishery-dependent and fishery-independent sources. *Spatial Statistics*, page 100930. https://doi.org/10.1016/j.spasta.2025.100930

Silva, D., Menezes, R., Moreno, A., Teles-Machado, A., Garrido, S. (2024). Environmental effects on the spatiotemporal variability of sardine distribution along the Portuguese continental coast. *JABES*. https://doi.org/10.1007/s13253-023-00577-8

Silva, D., Menezes, R., Serra-Pereira, B., Azevedo, M., Figueiredo, I. (2023). Adapting the sampling design of research surveys to improve the biomass estimation of secondary species—the case study of *Raja clavata*. *ICES Journal of Marine Science*, fsad089, https://doi.org/10.1093/icesjms/fsad089