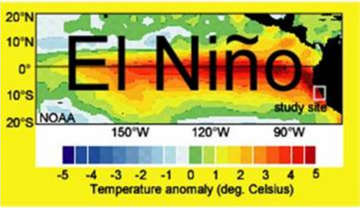
Bayesian spatio-temporal modelling of anchovy abundance through the SPDE Approach

Zaida Quiroz[1], *Marcos Prates*[2]
[1] Pontificia Universidad Católica del Perú
[2] Universidade Federal de Minas Gerais



Abbreviated abstract: The Peruvian anchovy is an important species from an ecological and economical perspective. In this context it is important to investigate the anchovy dynamics across years. Specifically, we propose flexible Bayesian hierarchical spatio-temporal models for zero-inflated positive continuous data. These models are able to capture the spatial and temporal distribution of the anchovies, to make spatial predictions. To make our modelling computationally feasible we use the stochastic partial differential equations (SPDE) approach combined with the integrated nested Laplace approximation (INLA) method.

Related publications:

– Quiroz Z. *et al*, *Bayesian spatio-temporal modelling of anchovy abundance through the SPDE Approach. Spatial Statistics* (28), 236-256 (2018)
– Quiroz Z. et al, *A Bayesian Approach to Estimate the Biomass of Anchovies Off the Coast of Peru.* Biometrics 71 (1), 208-217 (2015)

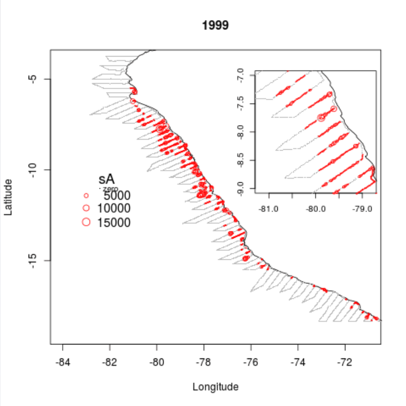
PUCP UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

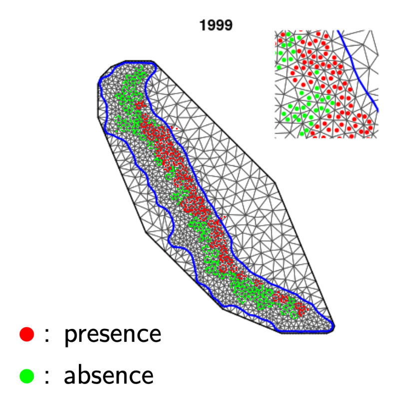
4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Problem and Data

The marine ecosystem of Peru is highly dominated by the anchovy.

- In order to preserve the species it is important to investigate the spatial anchovy abundance distribution across years.
- What is known about the anchovy distribution?
  - Nested aggregation structures: schools, clusters.
  - Fast response to environmental variability.





Data: collected by IMARPE during the summer season, from 1999 to 2007, —> T = 8.

Variable of interest:"anchovy abundance"

- Continuous and non-negative
- High proportion of zeros (gray dots)
- Spatio-temporal dynamics

The study area is divided using a Dealunay triangulation.

- N = 785 triangles with anchovy absence/ presence.
- Presence of anchovy ⇒ mean of anchovy abundance (>0).

PUCP UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Models

- $y(s_i,t)$: observed response representing the anchovy abundance in the location (centroid) $s_i \in D_s \subset R^2$ at time $t \in D_t \subset \aleph$, where $i = 1,\ldots,N$ and $t = 1,\ldots,T$. The distribution for $Y(s_i, t)$ has the finite mixture

$$\pi(y(s_i,t)|.) = p(s_i,t)\delta_0 + (1 - p(s_i,t))h(y(s_i,t)|\mu(s_i,t),\phi)I_{[y(s_i,t)>0]}$$

$p(s_i,t)$: probability of anchovy absence; δ_0: Dirac delta function; $Yit|Yit > 0 \sim \text{gamma}(\phi, \phi/\mu(s_i, t))$, mean $\mu(s_i, t)$ and precision parameter ϕ; h is the pdf of a gamma distribution.

- Let define the linear predictors:

$$\text{logit}(p(s_i,t)) = \eta(s_i,t)^{(1)} = Z^{(1)}\beta^{(1)} + f^{(1)}(t) + f_s^{(1)}(s_i,t)$$

$$\log(\mu(s_i,t)) = \eta(s_i,t)^{(2)} = Z^{(2)}\beta^{(2)} + f^{(2)}(t) + f_s^{(2)}(s_i,t)$$

For each linear predictor ($k = 1, 2$): $Z^{(k)}$ is a covariate matrix, $\beta^{(k)}$ is a coefficient vector (or regression parameters).

- $f^{(k)}(t)$ is a temporally structured effect:
  - Autoregressive dynamic AR(1)
  - Second-order random walk (rw2)
- $fs^{(k)}(.)$ is an spatio-temporal effect: spatio-temporal process changing in time with an autoregressive dinamic AR(1),

$$f_s^{(k)}(s_i,t) = a^{(k)} f_s^{(k)}(s_i,t-1) + w^{(k)}(s_i,t),$$

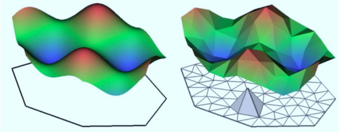
$$a^{(k)}| < 1, w^{(k)}(s\ ,t) \perp f^{(k)}(s\ ,1)$$

$$w_s^{(k)}(t) = (w^{(k)}(s_1,t),\ldots,w^{(k)}(s_N,t))' \sim GF(0,\Sigma_w^{(k)})$$

$ws^{(k)}(t)$ is a Gaussian field with Matérn covariance function with effective range $r^{(k)}$ and marginal variance $\sigma^{2(k)}$.

- Using the SPDE approach,

$w^{(k)}(s_i,t)$ is approximated by:



$$\widetilde{w}^{(k)}(s_i,t) = \sum_{j=1}^{n_v} \psi_j(s_i) g_j^{(k)},$$

for n_v vertices, $\{\psi_j\}$ are basis functions and $\{g_j\}$ are Gaussian weights with an sparse precision matrix (inverse of covariance matrix).

- Fast Bayesian Inference through INLA.

PUCP UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results

Table : The selection criteria for the models proposed

	Model S1	Model S2	Model S3	Model S4	Model ST1	Model ST2	Model ST3	Model ST4
$f^{(k)}(t)$	none	AR(1)	none(seasonal)	rw2	none	AR(1)	none(seasonal)	rw2
$\widetilde{f}_s^{(k)}(.)$	spatial	spatial	spatial	spatial	sp-temp	sp-temp	sp-temp	sp-temp
EAR	85.28	88.10	85.01	88.13	99.57	99.63	99.82	99.64
WAIC	134541.8	144472.8	132382	157928.1	57851.78	1300572	1223885	59135.91
LPML	−223454.6	−201825.6	−224057.3	−183475.9	−33240	−573887.6	−560865.3	−33999.51
PAR	86.11	86.24	85.98	75.66	83.31	83.82	84.71	85.22
RMSEE	1387.83	1379.31	1329.24	1325.89	429.03	459.29	600.20	450.00
RMSPE	1146.28	1134.05	1167.98	1356.89	1120.84	2846.36	4005.19	1320.29
Time	49 min	3 h	50 min	2 h	3 days	4 days 21 h	3 days 3 h	4 days 13 h

- Model ST1 is better in terms of goodness of fit and prediction capability.

Posterior mean (95% HPD CI) of the fixed parameters

Parameter	Model ST1
Binomial Part	
Intercept	8.9527 (−31.1589, 49.0949)
Distance to the coast	0.1725 (0.1344, 0.2115)
Latitude	10.9387 (3.0331, 18.8532)
Latitude^2	0.5010 (0.1354, 0.8670)
Depth	−2.2605 ($\star$) (−3.1458, −1.3909)
SST	1.2130 (0.6838, 1.7498)
Gamma Part	
Intercept	4.1838 (3.3823, 4.9847)
Distance to the coast	0.0015 (−0.0011, 0.0041)
Depth	0.4242 ($\star\star$) (0.3591, 0.4893)
SST	0.1079 (0.0668, 0.1489)

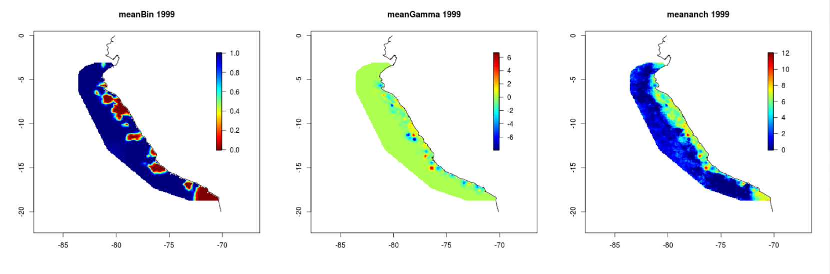
- Further the distance is, the higher the probability of anchovy absence.
- Higher probability of anchovy absence at the extremes.
- ($\star$) : >depth distance : >prob.abs.
- The higher SST, the higher probability of anchovy absence and anchovy abundance.
- ($\star\star$) : >depth distance: <abund.

Posterior mean (95% HPD CI) of the hyperparameters

Parameter	Model ST1	
Binomial Part		
$\sigma^{2(1)}$	343.8484	(341.5890, 348.0587)
$r^{(1)}$	2.2402	(2.2330, 2.2538)
$a^{(1)}$	0.7799	(0.7795, 0.7803)
Gamma Part		
ϕ	5.4724	(5.4453, 5.5189)
$\sigma^{2(2)}$	10.6815	(9.3656, 11.1250)
$r^{(2)}$	0.2032	(0.2020, 0.2055)
$a^{(2)}$	0.1041	(0.0948, 0.1192)

-Effective range of the probability of anchovy absence is 244 km, while for anchovy abundance is 22 km.

-The probability of absence/presence and abundance of anchovy for each site depends positively on the previous year .



Maps: Posterior estimation of probability of absence, anchovy of abundance given presence of anchovy and $E[y_p|y]$ of anchovy abundance from 1999 to 2006.

Conclusion: Our model provides a novel method to investigate the Peruvian anchovy dynamics across years, giving solid statistical support to many descriptive ecological studies.

PUCP

UFMG
UNIVERSIDADE FEDERAL DE MINAS GERAIS

zquiroz@pucp.edu.pe

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022