

Bayesian spatio-temporal analysis in the fishery context.

David Conesa



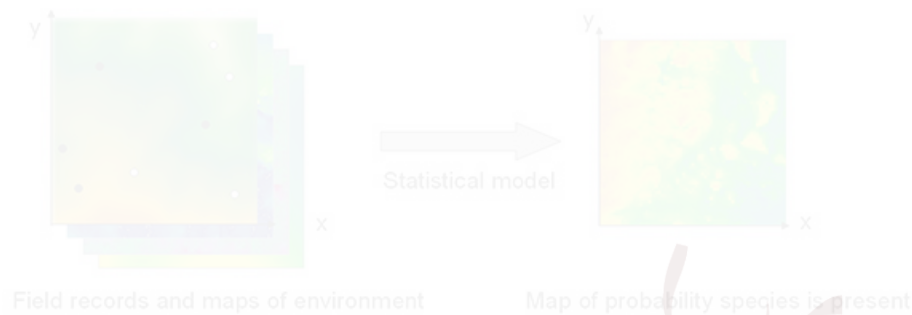
Universitat de València

*conjuntamente con A. López-Quílez, F. Muñoz, M. G. Pennino,
y otros.*

1 | Introducción

Distribución de especies

- ¿Cómo se distribuyen las especies?



- **Ejemplos:** peces, plantas, animales, enfermedades, etc.
- **Variables ambientales:** elevación, clima, vegetación, temperatura, clorofil-a, etc.
- **Aplicaciones:** cambio climático, conservación de especies, prevalencia de enfermedades, etc.

Distribución de especies (2)

Conservación de las especies

- **Objetivo:** identificar aquellas áreas que **contribuyen más a sostener** la viabilidad a largo plazo de la población.
- **¿Cómo?**
 - ▶ con los **mapas de probabilidad** conocemos **las áreas de mayor probabilidad de presencia y las de mayor abundancia**,
 - ▶ podemos **clasificar y gestionar** esas áreas como **Hábitats Esenciales**.

Distribución de especies (3)

Datos disponibles

Programas de monitorización de la biodiversidad: diseños experimentales de campo.

¿Aleatoriedad?

Modelo

- Dos tipos de **variable respuesta** (la que queremos explicar):
 - ▶ **Presencia/ausencia** de la especie, o
 - ▶ **Abundancia**.
- Dos situaciones:
 - ▶ **especies que no son objetivo**: podemos asumir independencia entre las localizaciones de las observaciones y la propia variable respuesta.
 - ▶ **especies objetivo**: sólo se muestrea donde hay un conocimiento acerca de la presencia de las especies (**muestreo preferencial**).

Ecología en Pesquerías

- Desde la antigüedad, **la pesca** ha constituido una fuente de **alimento**, de **trabajo** y **beneficio económico**.
- Los **sistemas de gestión pesquera están fallando** seriamente.
- Los gobiernos quieren recuperar **stocks de pesca** y prevenir mayor desgaste de las zonas de pesca mediante una gestión sostenible.
- ¿Cómo? **Gestión de pesquerías basada en un enfoque ecosistémico**.
 - ▶ **Identificar hábitats esenciales** y zonas de **mayor presencia de juveniles**;
 - ▶ **Reducir** los niveles excesivos de **capturas accidentales** y **descartes**;
 - ▶ **Desarrollar** técnicas efectivas y flexibles que funcionen a pesar de la **incertidumbre** y la **limitación en la información**.
- **Colaboración interdisciplinaria**: Instituto Español de Oceanografía + Departamento de Estadística e I.O.

Modelización Estadística

- **Objetivo:** modelar la **presencia/ausencia** (o la abundancia) de una especie no objetivo. **Se han pescado sin querer.**
- Los datos están Geo-referenciados → **geo-estadística.**
- El efecto espacial se puede analizar a través de un **modelo jerárquico bayesiano espacial. Ventajas:**
 - ① **Estadística Bayesiana** ⇒ Nos permite **combinar modelos complejos** junto con la **opinión experta previa** que se tenga del problema.
 - ② **Jerarquías** ⇒ Facilitan una **mejor percepción de la modelización.**
- **Estadística Bayesiana** es una manera más natural de entender la Estadística. De hecho ...
- “**Todo el mundo es bayesiano. Lo que pasa es que algunos lo saben y otros no**”: T. Raghunathan.

Para modelizar la distribución de la **presencia/ausencia** (abundancia de manera similar), definimos una **variable aleatoria binaria** como variable respuesta:

$$Z_i \sim \text{Bernoulli}(\pi_i),$$

(presencia (1) o ausencia (0) de la especie).

- Probabilidad de ocurrencia

$$\text{logit}(\pi_i) = X_i\beta + W_i \quad i = 1, \dots, n$$

- Predictor lineal observación i
- Efecto espacial
- Además incorporamos **Información previa** disponible (bayesiano).
- No hay expresión analítica para las distribuciones a posteriori: **resultados** de la inferencia y la predicción mediante **computación**.

La idea para abordar la computación:

- Tener en cuenta que estos modelos son un caso particular de **modelos latentes Gaussianos**.

Ventaja

Podemos obtener las **distribuciones a posteriori de los parámetros** y podemos **predecir** utilizando **INLA** (Rue et al., 2009).

Ventaja 2

Podemos utilizar **SPDE** (del inglés Stochastic Partial Differential approach) para aproximar el **Campo Gaussiano continuo** por un **Campo Markoviano Aleatorio Gaussiano** (que es discreto): **computación más sencilla**.

Modelo final con SPDE

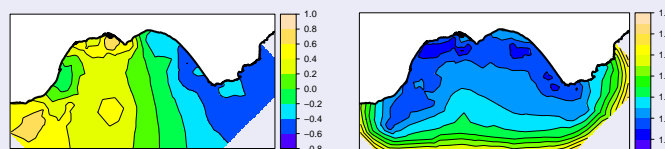
$$\begin{aligned}Z_i &\sim \text{Ber}(\pi_i), i = 1, \dots, n \\ \text{logit}(\pi_i) &= X_i\beta + \theta_i \\ \pi(\beta_j) &\sim N(\mu_{\beta_j}, \rho_{\beta_j}) \\ \theta_i &\sim N(0, \mathbf{Q}(\kappa, \tau)) \\ 2\log\kappa &\sim N(\mu_{\kappa}, \rho_{\kappa}) \\ \log\tau &\sim N(\mu_{\tau}, \rho_{\tau})\end{aligned}$$

- El efecto espacial depende ahora de dos nuevos parámetros: κ y τ , que determinan el rango del efecto y la varianza total, respectivamente.
- Comparación de modelos con dos criterios DIC y CPO (*Conditional Predictive Ordinate*).

Distribución del jurel en el golfo de Almería.

- A pesar de su poco valor comercial, el **jurel** juega un papel importante en la zona de transición entre Mediterráneo y Atlántico.
- Covariables: **profundidad** (efecto neg.) y **clorofil-a** (efecto pos.).

Efecto espacial

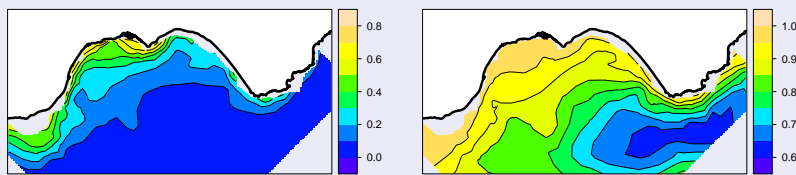
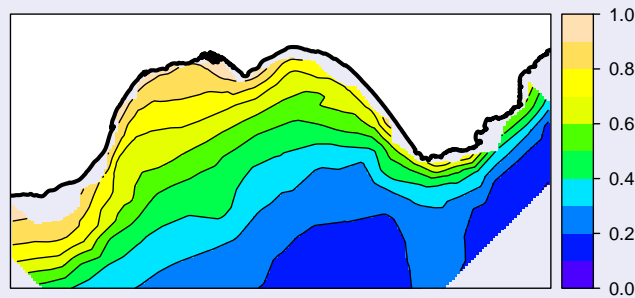


Media a posteriori y desviación estándar del efecto espacial.

- Parece que hay un efecto este-oeste.
- Zona costera al oeste es una zona protegida.

Distribución del jurel en el golfo de Almería (2).

Probabilidad a posteriori de ocurrencia

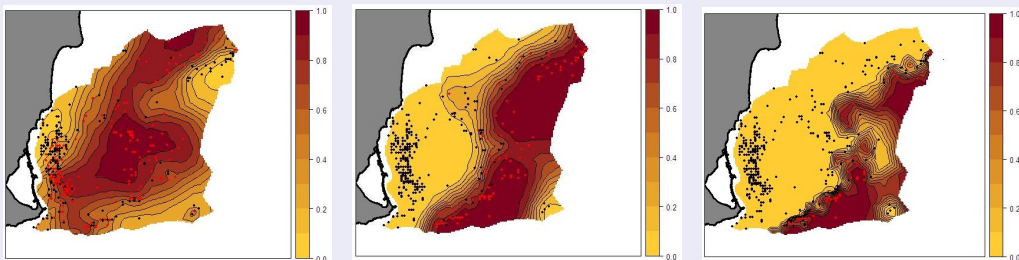


Mediana (arriba) y cuartiles 1 y 3 de π_i |datos

Distribución de 3 especies de elasmobranquios.

- Preocupación creciente en los **elasmobranquios** por su **vulnerabilidad**.
- **Predictores**: profundidad, pendiente del suelo marino y el tipo de sustrato, junto con temperatura y clorofil-a.

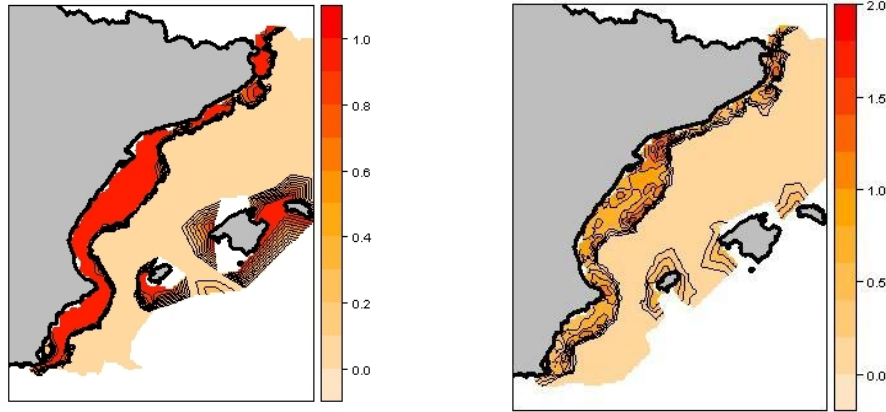
Mediana de la probabilidad a posteriori de la presencia



- Las especies muestran **profundidades óptimas diferentes**.
- Enorme utilidad para identificar **hábitats más sensibles**, y ayudar en la preservación de las especies.

Distribución de los juveniles de merluza en el Mediterráneo.

- Interés en conocer las zonas pesqueras con **más juveniles**.
- Trabajo Fin de Máster de **losu Paradinas** premiado recientemente.



- Probabilidad a posteriori de **presencia de juveniles** (izquierda).
- Media a posteriori de la **cantidad de juveniles** (derecha).

Otras aplicaciones fuera de pesquerías.

- Enorme preocupación en Galicia por la **paramfistomosis en vacas** y la **posibilidad de infectar humanos**.
- Interés en **detectar las zonas de prevalencia mayor**.
- Covariables: **Temperatura** y **log(pendiente)**.
- **Componente espacial del modelo ajustado en vacas de leche** en Galicia: media a posterior (A) y desviación estándar (B).





Modelización Estadística

- **Objetivo:** modelizar la **abundancia** de una especie objetivo.
- Otra vez, datos geo-referenciados → **geo-estadística**.
- Modelos Geostadísticos asumen que las localizaciones y el proceso que se modela son independientes.
- Pero, las localizaciones suelen estar **deliberadamente concentradas** en **áreas donde se espera una alta abundancia de pesca**: → Los pescadores pescan en sitios donde piensan que va a haber pesca.
- Ejemplo claro de **muestreo preferencial**.
- Implementación del Preferential sampling como un proceso puntual marcado.

Preferential sampling como proceso puntual marcado.

- Ambos modelos (localizaciones y marcas) son **modelos Gaussianos latentes Gaussian models**.
- Otra vez **SPDE + INLA**.

El modelo preferencial

Modelo de proceso puntual

$$\begin{aligned} \mathbf{x}|\boldsymbol{\eta} &\sim \mathcal{P}(\exp(\boldsymbol{\eta})) \\ \boldsymbol{\eta} &= \mathbf{1}\beta_{0\eta} + \mathbf{X}_{\eta}\boldsymbol{\beta}_{\eta} + \boldsymbol{\theta} \\ \pi(\beta_0) &\propto 1 \\ \beta_j &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1e-05) \end{aligned}$$

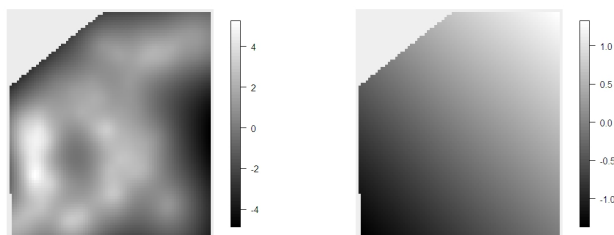
Modelo para las marcas

$$\begin{aligned} y_i|\kappa_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(\kappa_i, \tau^2) \\ \kappa_i &= \beta_{0\kappa} + \mathbf{X}_{i\kappa}\boldsymbol{\beta}_{\kappa} + \theta_i \\ \pi(\beta_0) &\propto 1 \\ \beta_j &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1e-05) \end{aligned}$$

→ donde $\boldsymbol{\theta}$ es un GMRF como en el caso anterior.

Distribución de la merluza en el golfo de Alicante.

- Estudios sobre merluza no tienen presente el **muestreo preferencial**.
- Primera aproximación: **no covariables**, pero si **error del pescador**.
- Media a posteriori del **efecto espacial del modelo con muestreo preferencial y error del pescador** (izda.) y del **modelo sin preferencial** (dcha.).



- Izda: **se aprecian zonas de mayor densidad**.
- Dcha: **demasiado suavizado**, es más difícil encontrar **zonas de interés**.



Conclusiones y trabajo futuro

- Enorme utilidad de la **modelización de la distribución de especies**.
- **INLA** y el **SPDE** nos permite tener una **herramienta rápida y precisa** para realizar inferencia Bayesiana y **predecir** en modelos latentes Gaussianos con componentes espaciales continuas.
- **Preferential sampling** se suele ignorar en los modelos de distribución de especies.
- Líneas futuras:
 - ▶ comparar resultados con hSDM R package de Gelfand y coautores,
 - ▶ incluir en el modelo la **incertidumbre de las covariables**,
 - ▶ unir los modelos **espacio-temporales** con modelos (computacionales) **de dinámica de poblaciones**.

Referencias.

- P.J. Diggle, R. Menezes and T.L. Su, (2010). Geostatistical inference under preferential sampling.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 59 (2): 191–232.
- F. Lindgren, H. Rue and J. Lindstrom, (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion).
Journal of the Royal Statistical Society, Series B, 73: 423-498.
- F. Muñoz, M. G. Pennino, D. Conesa, A. López-Quílez and J.M. Bellido. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models.
Stoch Environ Res Risk Assess, 27: 1171–1180.
- M. González-Warleta, S. Lladosa, J. A. Castro-Hermida, A. M. Martínez-Ibeas, D. Conesa, F. Muñoz, A. López-Quílez, Y. Manga-González and M. Mezo, (2013). Bovine paramphistomosis in Galicia (Spain): Prevalence, intensity, aetiology and geospatial distribution of the infection.
Veterinary Parasitology, 191(3-4): 252-263.
- M.G. Pennino, F. Muñoz, D. Conesa, A. López-Quílez, J.M. Bellido (2013). Modelling sensitive elasmobranch habitats.
Journal of Sea Research, In press.
- H. Rue, S. Martino and N. Chopin, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.
Journal of the Royal Statistical Society, Series B, 71(2): 319-392.