Georeferenced Random Sampling

Alba Fuster Alonso

1 Introduction

This paper develops the models applied in R-INLA for the case of random sampling, simulating an ocanographic campaign.

As a reminder, the models will present the response variable in two ways: (1) Catches and (2) CPUE. In addition, the model covariates will be fishing effort (E) and bathymetry (B). Effort will also be included as an offset in one of the models.

Similarly, the following scheme will be followed in the models: (1) models with the response variable transformed to avoid zeros, (1.1) models without spatial term, (1.2) models with spatial term and (2) models with the original response variable.

1.1 Response variable

The response variables CPUE and catches are quantitative, continuous and positive. Therefore, we propose the use of two probability distributions to fit the models: (1) gamma and (2) log normal.

1.2 Explanatory variables

In the case of the explanatory variables in the attaches files in the github repository, it is possible to observe how the response variable and the bathymetry do not follow a linear relationship, but rather, the response variable reaches its maximum at intermediate bathymetries. For this reason, the smoothing of the explanatory variable bathymetry (B) is proposed.

2 Models without zeros

A total of 17 models have been added in the following section, the first 4 of which have no spatial part, while the rest do.

2.1 Models without zeros and spatial term

1. Model 1

$$Catches_i \sim Gamma(\mu_i, \phi)$$

$$log(\mu_i)^{\mathrm{I}} = \beta_0 + \beta_1 E_i + \beta_2 B_i$$

2. Model 2

$$Catches_i \sim lgnormal(\mu_i, \sigma)$$

$${\mu_i}^{\mathrm{II}} = \beta_0 + \beta_1 E_i + \beta_2 B_i$$

[I] Logarithm is the link function of the Gamma distribution

[II] Identity is the link function of the Lognormal distribution

3. Model 3

$$CPUE_i \sim Gamma(\mu_i, \phi)$$

$$log(\mu_i) = \beta_0 + \beta_1 B_i$$

4. Model 4

$$CPUE_i \sim lognormal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 B_i$$

2.2 Models without zeros and with spatial term

Form this section on, the spatial term will be included, in order to understand how this term works in R-INLA, an Appendix is added whit a detailed explanation of the spatial term for one of the models (Appendix I)

1. Model 5

$$Catches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 E_i(s) + \beta_2 B_i(s) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

2. Model 6

$$Catches_i(s) \sim lgnormal(\mu_i(s), \sigma)$$
$$\mu_i(s) = \beta_0 + \beta_1 E_i(s) + \beta_2 B_i(s) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

3. Model 7

$$CPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 B_i(s) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

4. Model 8

$$CPUE_i(s) \sim lgnormal(\mu_i(s), \sigma)$$
$$\mu_i(s) = \beta_0 + \beta_1 B_i(s) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

5. Model 9

$$Catches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 E_i(s) + f_{rw1}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

6. Model 10

$$CPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + f_{rw1}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

7. Model 11

$$Catches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 E_i(s) + f_{rw2}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

8. Model 12

$$CPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + f_{rw2}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

9. Model 13 [III]

$$Catches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$\mu_i(s) = E_i(s)\, x\, \lambda_i(s)$$
$$log(\mu_i(s)) = log(E_i(s)) + log(\lambda_i(s))$$
$$log(\mu_i(s)) = log(E_i(s)) + \beta_0 + f_{rw1}(B_i(s)) + u_i(s)$$
$$\frac{\mu_i(s)}{E_i(s)} = \beta_0 + f_{rw1}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

10. Model 14

$$Catches_i(s) \sim lognormal(\mu_i(s), \sigma)$$
$$l\mu_i(s) = \beta_0 + \beta_1 E_i(s) + f_{rw1}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

11. Model 15

$$CPUE_i(s) \sim lognormal(\mu_i(s), \sigma)$$
$$l\mu_i(s) = \beta_0 + f_{rw1}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

12. Model 16

$$Catches_i(s) \sim lognormal(\mu_i(s), \sigma)$$
$$l\mu_i(s) = \beta_0 + \beta_1 E_i(s) + f_{rw2}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

13. Model 17

$$CPUE_i(s) \sim lognormal(\mu_i(s), \sigma)$$
$$l\mu_i(s) = \beta_0 + f_{rw2}(B_i(s)) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

[III] Effort is apply as a offset in this model

3 Models with zeros: Hurdle models

Hurdle model for continuous data consists of a binary part and a continuous part. The distribution for the binary part will again be the Bernoulli distribution, but for the continuous part we use a gamma distribution. There are two components: a gamma part with the mean μ_i and the Bernoulli part with the mean π_i. We model the mean μ_i as a function of covariates and the log-link is used. Fort the Bernoulli part we use the logistic link. We can use the same or different sets of covariates for the continuous and binary parts, and we can also add spatial correlated random effecto to either part of the model.

1. Hurdle Model 1
$$YCPUE_i(s) \sim Ber(\pi_i(s))$$
$$ZCPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$logit(\pi_i(s)) = \beta_{0YCPUE} + f(B_i(s)) + u_{iYCPUE}(s)$$
$$u_{iYCPUE}(s) \sim GMRF(0, \Sigma)$$
$$log(\mu_i(s)) = \beta_{0ZCPUE} + f(B_i(s)) + u_{iZCPUE}(s)$$
$$u_{iZCPUE}(s) \sim GMRF(0, \Sigma)$$
2. Hurdle Model 2
$$YCPUE_i(s) \sim Ber(\pi_i(s))$$
$$ZCPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$logit(\pi_i(s)) = \beta_{0YCPUE} + f(B_i(s)) + u_{iZYCPUE}(s)$$
$$log(\mu_i(s)) = \beta_{0ZCPUE} + f(B_i(s)) + u_{iZYCPUE}(s)$$
$$u_{ZYCPUE}(s) \sim GMRF(0, \Sigma)$$
3. Hurdle Model 3
$$YCPUE_i(s) \sim Ber(\pi_i(s))$$
$$ZCPUE_i(s) \sim lognormal(\mu_i(s), \sigma)$$
$$logit(\pi_i(s)) = \beta_{0YCPUE} + f(B_i(s)) + u_{iZYCPUE}(s)$$
$$l\mu_i(s) = \beta_{0ZCPUE} + f(B_i(s)) + u_{iZYCPUE}(s)$$
$$u_{ZYCPUE}(s) \sim GMRF(0, \Sigma)$$
4. Hurdle Model 4
$$YCPUE_i(s) \sim Ber(\pi_i(s))$$
$$ZCPUE_i(s) \sim lgnormal(\mu_i(s), \sigma)$$
$$logit(\pi_i(s)) = \beta_{0YCPUE} + f(B_i(s)) + u_{iYCPUE}(s)$$
$$u_{iYCPUE}(s) \sim GMRF(0, \Sigma)$$
$$\mu_i(s) = \beta_{0ZCPUE} + f(B_i(s)) + u_{iZCPUE}(s)$$
$$u_{iZCPUE}(s) \sim GMRF(0, \Sigma)$$
5. Hurdle Model 5
$$YCatches_i(s) \sim Ber(\pi_i(s))$$
$$ZCatches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$logit(\pi_i(s)) = \beta_{0YCatches} + \beta_1 E_i(s) + f(B_i(s)) + u_{iYCatches}(s)$$
$$u_{iYCatches}(s) \sim GMRF(0, \Sigma)$$
$$log(\mu_i(s)) = \beta_{0ZCatches} + \beta_1 E_i(s) + f(B_i(s)) + u_{iZCatches}(s)$$
$$u_{iZCatches}(s) \sim GMRF(0, \Sigma)$$

6. Hurdle Model 6

$$YCPUE_i(s) \sim Ber(\pi_i(s))$$
$$ZCPUE_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$logit(\pi_i(s)) = \beta_{0YCPUE} + B_i(s)^2 + u_{iYCPUE}(s)$$
$$u_{iYCPUE}(s) \sim GMRF(0, \Sigma)$$
$$log(\mu_i(s)) = \beta_{0ZCPUE} + f(B_i(s)) + u_{iZCPUE}(s)$$
$$u_{iZCPUE}(s) \sim GMRF(0, \Sigma)$$

4 Bayesian approximation

It is possible to use Bayes' theorem to obtain a posterior distributions of the fixed and random effects proposed in the linear predictor. To explain how it works the following process is shown:

4.1 Models without zeros

1. Likelihood function:

$$A_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 E_i(s) + \beta_2 B_i(s) + \beta_3 E(s)xB(s)_i + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

In this case, there are four fixed parameters and two hyperparameters to be estimate: the intercept β_0, the coefficients of covariates $\beta_{1,\dots,3}$ and the spatial random effect $u_i(s)$ and the dispersion ϕ:

2. Priors (INLA default, R-INLA works with mean and precision):

   2.1. $\beta_0 \sim N(0, 0)$

   2.2. $\beta_{1,\dots,3} \sim N(0, 0.001)$

   2.3. $\phi \sim lnGamma(1, 0.0005)$

4.2 Hurdle models

1. Likelihood function (the same equation as in section 2.2):

$$Y_i(s) \sim Ber(\pi_i(s))$$
$$Z_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$logit(\pi_i(s))^{\mathrm{IV}} = \beta_{0Y} + \beta_{1Y} E_i(s) + \beta_{2Y} B_i(s) + \beta_{3Y} E(s)xB(s)_i + u_{iY}(s)$$
$$u_{iY}(s) \sim GMRF(0, \Sigma)$$
$$log(\mu_i(s) = \beta_{0Z} + \beta_{1Z} E_i(s) + \beta_{2Z} B_i(s) + \beta_{3Z} E(s)xB(s)_i + u_{iZ}(s)$$
$$u_{iZ}(s) \sim GMRF(0, \Sigma)$$

2. Priors

   2.1. $\beta_{0Y}\, and\, \beta_{0Z} \sim N(0, 0)$

   2.2. $\beta_{1Y,\dots,3Y}\, and\, \beta_{1Z,\dots,3Z} \sim N(0, 0.001)$

   2.3. $\phi \sim lnGamma(1, 0.0005)$

[IV] link function: logit $\frac{\pi}{1-\pi}$

5 Appendix

5.1 Appendix I

The following equation is almost identical to the equation to the equations without a spatial term, except that it contains an extra term u_i at the end. It refers to the spatial correlated random term, which is estimated with the SPDE approach by R-INLA.

$$Catches_i(s) \sim Gamma(\mu_i(s), \phi)$$
$$log(\mu_i(s)) = \beta_0 + \beta_1 E_i(s) + \beta_2 B_i(s) + u_i(s)$$
$$u_i(s) \sim GMRF(0, \Sigma)$$

Where $u = (u(s_1), ..., u(s_N))$ and Σ is the covariance matrix of dimension N x N. Zuur, Ieno and Saveliev (2017) summarises the storyline of spatial terms as follow:

1. N sampling location: s_1 to s_N.
2. Random effect $u(s_i)$.
3. Assume that the GF $u = (u(s_1), ..., u(s_N))$ are normal distributed with mean 0 and covariance Σ_{GF}

4. To get Σ_{GF}: (1) assume a Markovian (=local) behavior, (2) the use of Matern correlation function to quantify the covariance matrix, then (3) to estimate a couple of parameters we use SPDE approach, also (4) using the finite element approach in combination with SPDE we get all the parameters that we need:

4.1. SPDE approach (The components in the next equation are linked to the components of the Matern correlation function):

$$(k^2 - \Delta)^{\alpha/2} \tau U(s) = W(s)$$ [V]

4.2. Finite element approach (elements of the covariance matrix of a GMRF are defined on a regularly or most common irregular spaced two-dimensional, so we use and irregular grid, also called mesh.):

$$u(s_i) = \sum_{k=1}^{G} a_k(s_i)\, x\, w_k$$ [VI]

5.2 Appendix II

The following appendix develops the rw1 (random walk of order 1) models implemented in R-INLA.

The random walk model of order 1 (rw1) for the Gaussian vector $B = (B_1, ..., B_2)$ is constructed assuming independent increments:

$$\Delta B_i = B_i - B_{i-1} \sim N(0, \tau^{-1})$$

The density for B is derived from its $n-1$ increments as

$$\pi(B|\tau) \propto \tau^{n-1/2}\, exp\{-\frac{1}{2}B^T Q B\}$$

Where $Q = \tau R$ and R is the sturcture matrix reflecting the neighbourhood structure of the model.

5.2.1 Hyperparameters

The precision parameter τ is represented as

$$\theta = log\tau$$

and the prior is defined on θ, the default prior is loggama(1, 5e-05)

[V] k is kappa, Δ is the Laplace operator and W is the Gaussian spatial white noise process
[VI] $a_k s(_i)$ is a known value.