

Supplementary material of the Manuscript: “Spatio-temporal analysis of record-breaking temperature increments across Spain”

Ana C. Cebrián, Jesús Asín, Jorge Castillo-Mateo,

Departamento de Métodos Estadísticos, Universidad de Zaragoza, Zaragoza, Spain.

Alan E. Gelfand

Department of Statistical Science, Duke University, Durham, North Carolina, USA.

This supplementary material contains explicit details for the exploratory analysis, fitting approach, validation measures and plots and additional inference results from the manuscript.

1 Exploratory Analysis

An exploratory descriptive analysis (EDA) is conducted to identify the key features for inclusion in the model. To achieve this, we compute and plot the empirical densities of record increments across different time periods and conditions; the densities are estimated using the Beta boundary kernel for positive variables. In addition to the evolution across decades and the previous day effect, studied in the main manuscript, seasonal behavior and spatial variability are studied here.

Seasonal behavior. The behavior of increments is not homogeneous within the JJA period and the monthly distributions vary across decades, Figure 1 shows the plots of the empirical densities calculated per month in two different 10-year periods, 1992–2001 and 2012–2021,

Spatial variability. To investigate the effect of coastal proximity in the occurrence of records, we compare the empirical density of increments for stations located within 50 km of the coast versus those farther inland. We also explore whether this difference depends on whether a record occurred the previous day or not. Figure 1 (bottom) illustrates this by comparing empirical densities across the four scenarios during the last decade, revealing distinct distributions.

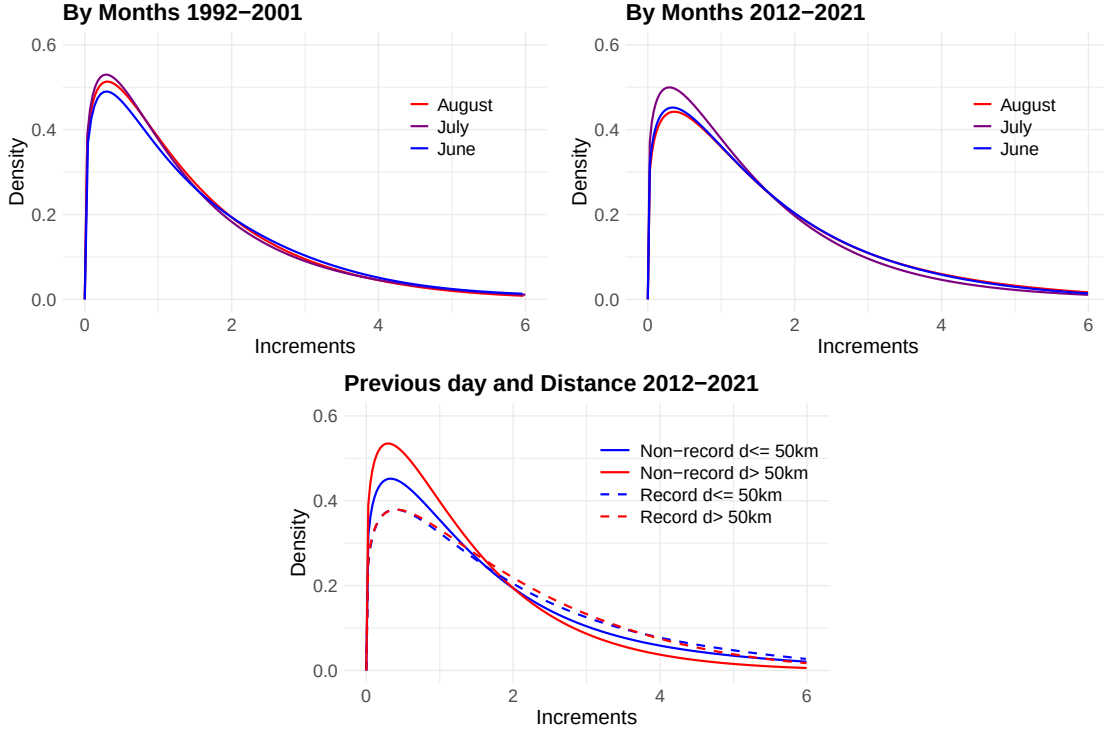


Figure 1: Empirical density of the record increments, in $^{\circ}\text{C}$, per months in periods 1992–2001 and 2012–2021 (top), and conditioned on both the previous day and distance to the coast (bottom).

2 Fitting details

The model is fitted in INLA (version 2024-12-11). We use the *inla* function using the family Gamma and a log lik function. The Gamma distribution is parametrised as follows,

$$f(y) = \frac{b^a}{\Gamma(a)} y^{a-1} \exp(-by), \quad a > 0, \quad b > 0, \quad y > 0,$$

where $E(y) = \mu = \frac{a}{b}$ and $Var(y) = \frac{1}{\tau} = \frac{a}{b^2}$, where τ is the precision and μ is the mean. The parametrisation used by INLA for the precision is:

$$\tau = \frac{\phi}{\mu^2}.$$

The hyperparameter is the shape parameter ϕ , which is represented as $\phi = \exp(\theta)$ and the prior is defined on θ .

For fitting the spatial Gaussian process used in the spatial intercept, INLA uses a Matérn SPDE model (Lindgren et al., 2022). This approach requires a mesh or triangularization covering the observed spatial points to be used. The considered mesh, with 363 points, is shown in Figure 2, and it is obtained using a nonconvex hull of the coordinates of the 40 sites used in the estimation. The Matérn SPDE model in a bidimensional space, with spatial range ρ and standard deviation parameter σ is parametrised as follows

$$(\kappa^2 - \Delta)^{\alpha/2} (\tau x(u)) = \mathcal{W}(u)$$

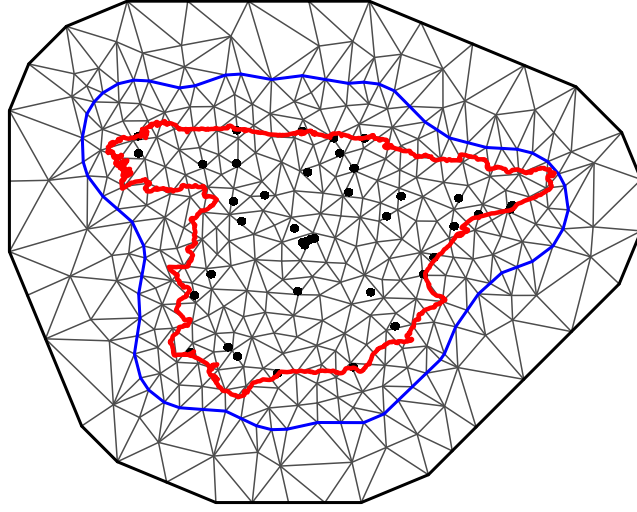


Figure 2: Mesh used to fit the spatial intercept in the spatial models; border of peninsular Spain is plotted in red.

where the spatial scale parameter is given by $\kappa = \frac{8\nu}{\rho}$, where $\nu = \alpha - 1$, and τ is proportional to $1/\sigma$. The hyperparameters are ρ and σ , as α is fixed at 2 by default.

The auto-regressive model of order 1, AR(1), for $\mathbf{x} = (x_1, \dots, x_n)$ is defined in INLA as:

$$x_1 \sim N\left(0, [\tau(1 - \rho^2)]^{-1}\right)$$

$$x_i = \rho x_{i-1} + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^{-1}), \quad i = 2, \dots, n$$

where $|\rho| < 1$.

There are two hyperparameters. The precision parameter κ is represented as: $\theta_1 = \log(\kappa)$, where κ is the marginal precision $\kappa = \tau(1 - \rho^2)$. The parameter ρ is represented as:

$$\theta_2 = \log\left(\frac{1 + \rho}{1 - \rho}\right)$$

and the prior is defined on $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

Since we only consider the three summer months each year, from June to August, this creates time gaps between years. To account for this, the AR(1) structure is implemented using a replicate structure by year. This means that each year is modeled as an independent AR(1) process, while sharing the same within-year day structure.

The prior distributions for the model hyperparameters are the following:

- For the precision parameter θ , the prior is weakly informative, *loggamma* with parameters 1 and 0.01.
- For the fixed intercept and the coefficients of the fixed effects, the prior is weakly informative prior with large variance, $\beta_i \sim N(0, 1000)$.
- For the SPDE structure: a prior that penalises the complexity (PC prior) is used, (Fuglstad et al., 2019). The practical approach for this in INLA is to specify the hyperparameters ρ and σ through

$$P(\rho < \rho_0) = p_\rho$$

and

$$P(\sigma > \sigma_0) = p_\sigma$$

that means that we specify the lower tail quantile and probability for the range (ρ_0 and p_ρ) and the upper tail quantile and probability for the standard deviation (σ_0 and p_σ). In this case, we use a standard option, for the spatial range ρ we consider the 50th quantile equal to the minimum of the range of the latitude and the longitude divided by 2 and for σ the 90th percentile equal to 1.

- For the daily AR1 structure:
 - For θ_1 , the prior is a *loggamma* distribution with parameters 1 and 5e-05.
 - For θ_2 , the prior is a *normal* distribution with parameters 0 and 0.15.

3 Residuals and fitted values

Figure 3 shows some relevant plots of the selected model MS3. The plots in the first row present the residuals (observed minus fitted increments) of the model versus the year, the day within the summer and the site. As expected, the variability of the residuals in the first years is higher. No seasonal or spatial patterns are observed in the plots against the day within the summer and against the locations. The plots in the second row present the fitted values of the model versus the same variables, year, the day within the summer and the site. The plot in the third row presents the observed increments against the posterior mean of the increment.

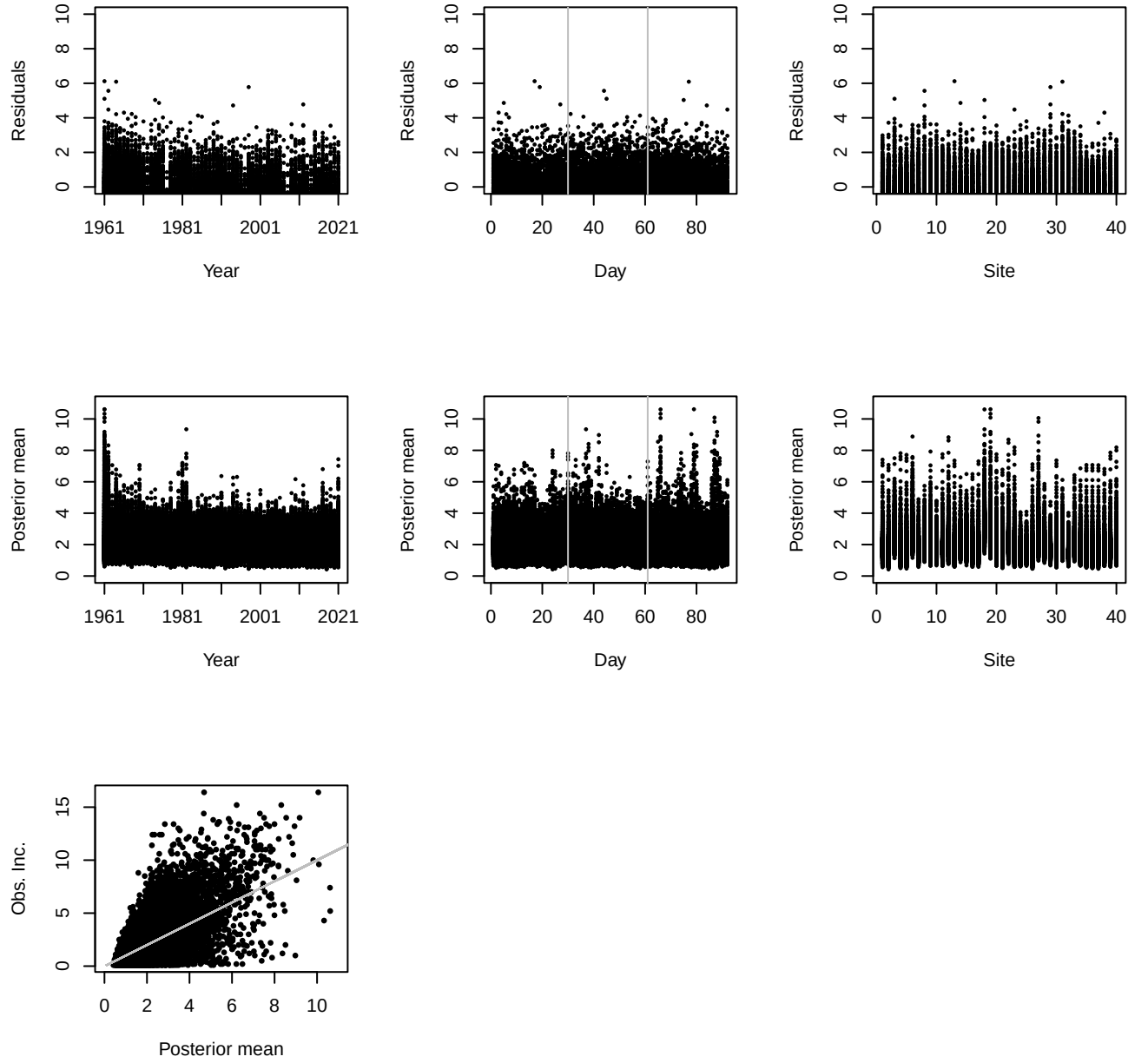


Figure 3: Model MS3. Top: Residuals against year, day within the summer and sites. Middle: Fitted values. Bottom: Observed increments against posterior mean.

4 Validation analysis: coverages

The average coverage, calculated as described in the main manuscript, of the selected model MS3 is 0.93. The plots in Figure 4 summarize the average coverages by year (from 1961 to 2021), by day within the summer (from 1 (first of June) to 92 (the 31th of August) and by site (from 1 to 40); the sites are: Badajoz-Talavera la Real, Madrid-Retiro, Málaga aeropuerto, Navacerrada, Salamanca aeropuerto, San Sebastián-Iguelo, Tortosa-observatorio del Ebro, Valencia, Zaragoza aeropuerto, Barcelona-Fabra, Albacete los Llanos, Burgos-Villafría, Ciudad Real, A Coruña, Murcia, Sevilla-San Pablo, Soria, Bilbao aeropuerto, Santiago de Compostela, Ponferrada, León-Virgen del camino, Logroño-Agoncillo, Zamora, Reus aeropuerto, Barcelona aeropuerto, Madrid Torrejón, Vitoria-Foronda, Almeria aeropuerto, Gijón, Cáceres ciudad, Santander, Castellón, Huelva, Lleida-observatorio, Madrid-Barajas, Madrid-Cuatrovientos, Madrid-Getafe, Morón de la Frontera, Valladolid-Villanubla, Daroca.

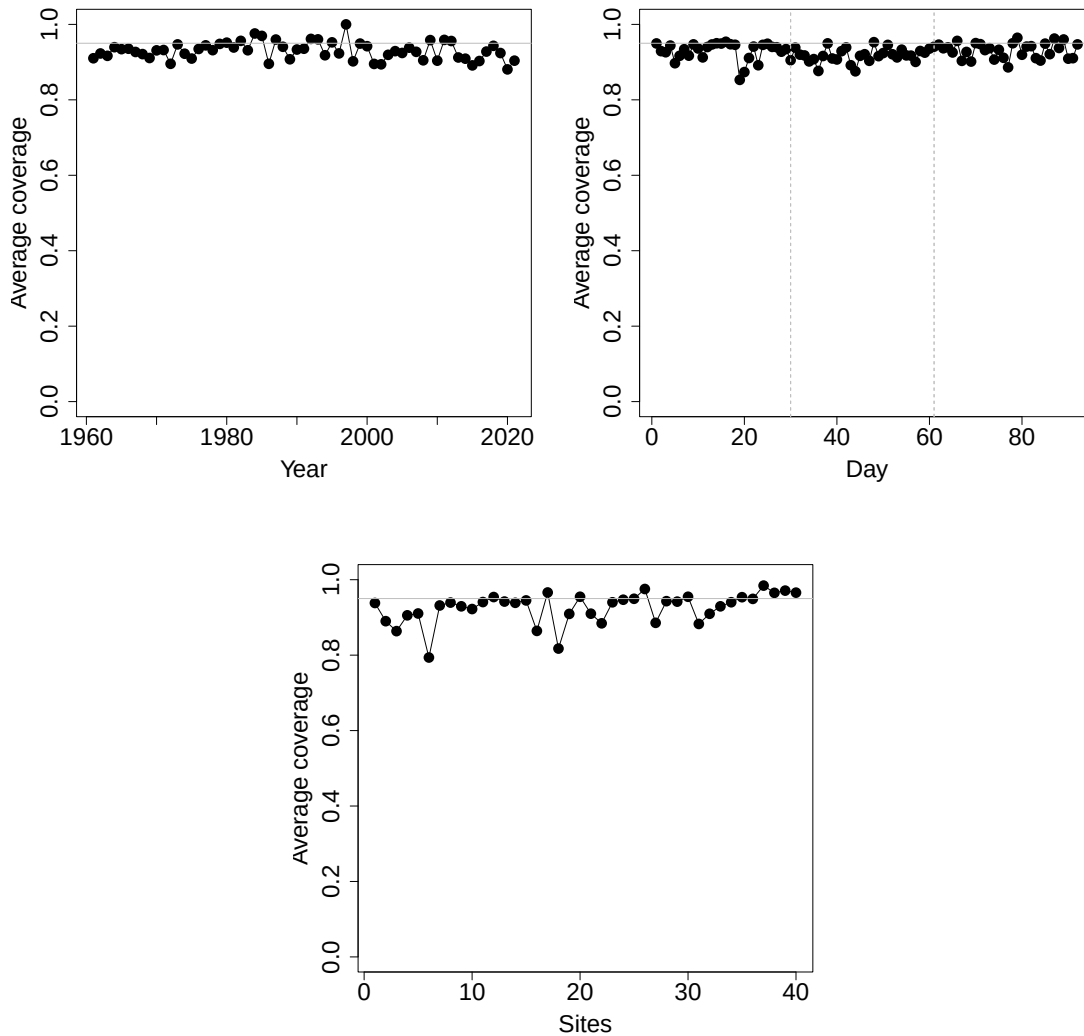


Figure 4: Average coverage of the selected model MS3 against year, day within the summer and by site

5 More post-model inference plots

5.1 Conditional posterior distribution given the previous day indicator

In this section, we show the maps of the posterior mean of the average of the differences between the increments (in $^{\circ}\text{C}$) given that the previous day was a record and given it was not, across days in JJA and one decade as described in Section 6.1 of the main text, for decades $D3$, $D4$, $D5$ and $D6$, that is 1982–1991, 1992–2001, 2002–2011, and 2012–2021 respectively, see Figure 5.

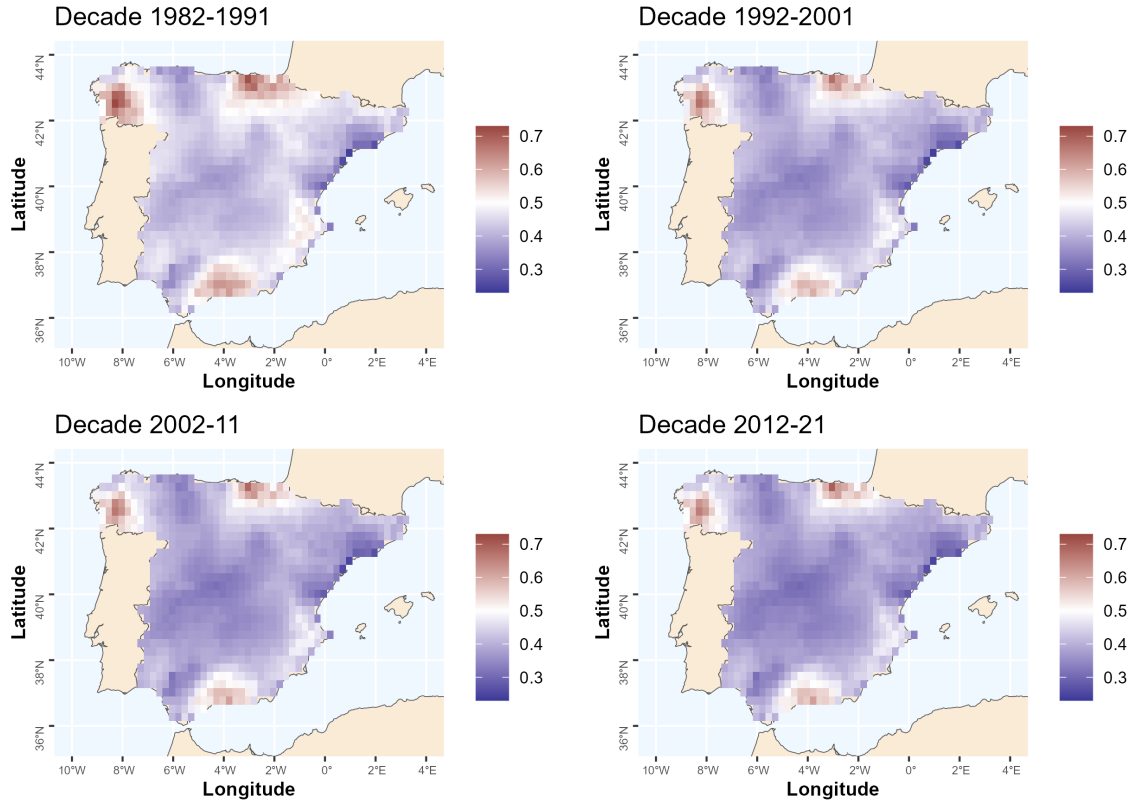


Figure 5: Posterior mean of the average of the differences between the increments (in $^{\circ}\text{C}$) given that the previous day was a record and given it was not, across days in JJA and one decade.

5.2 Marginal posterior distribution

5.2.1 Spatial analysis of increments

In this section, we show some additional maps related to the spatial analysis of the increments. In all the figures in this section, decade $D1$ is not shown due to the high variability in the increments during the early stage.

Figure 6 show maps of the posterior mean and 5th and 95th percentiles of the average of the increments during the three summer months over one decade D , $\bar{J}_{D,JJA}(\mathbf{s})$, for decades $D6$ to $D2$ and also for the entire period; see details on the calculation in Section 6.2.1 in the main manuscript.

Figure 7 shows the maps of the posterior mean and 5th and 95th percentiles of the difference of the averages in the last decade $D6$ and the other decades $D5$, $D4$, $D3$ and $D2$: $\bar{J}_{D6,JJA}(\mathbf{s}) - \bar{J}_{D,JJA}(\mathbf{s})$; see Section 6.2.1 in the main manuscript for more details.

We also show maps of the probabilities of the average increments in July being higher than in June, and in August being higher than in June and July, that is

$$P_D(\bar{J}_{t,M_1}(\mathbf{s}) > \bar{J}_{t,M_2}(\mathbf{s}))$$

where $\bar{J}_{t,M}(\mathbf{s})$ is the average increment in month M at year t and the probability is computed over the 10 years t within the decade D . Figure 8 shows these maps for the entire period and for decades $D2$ to $D6$; see Section 6.2.1 in the main manuscript for more details.

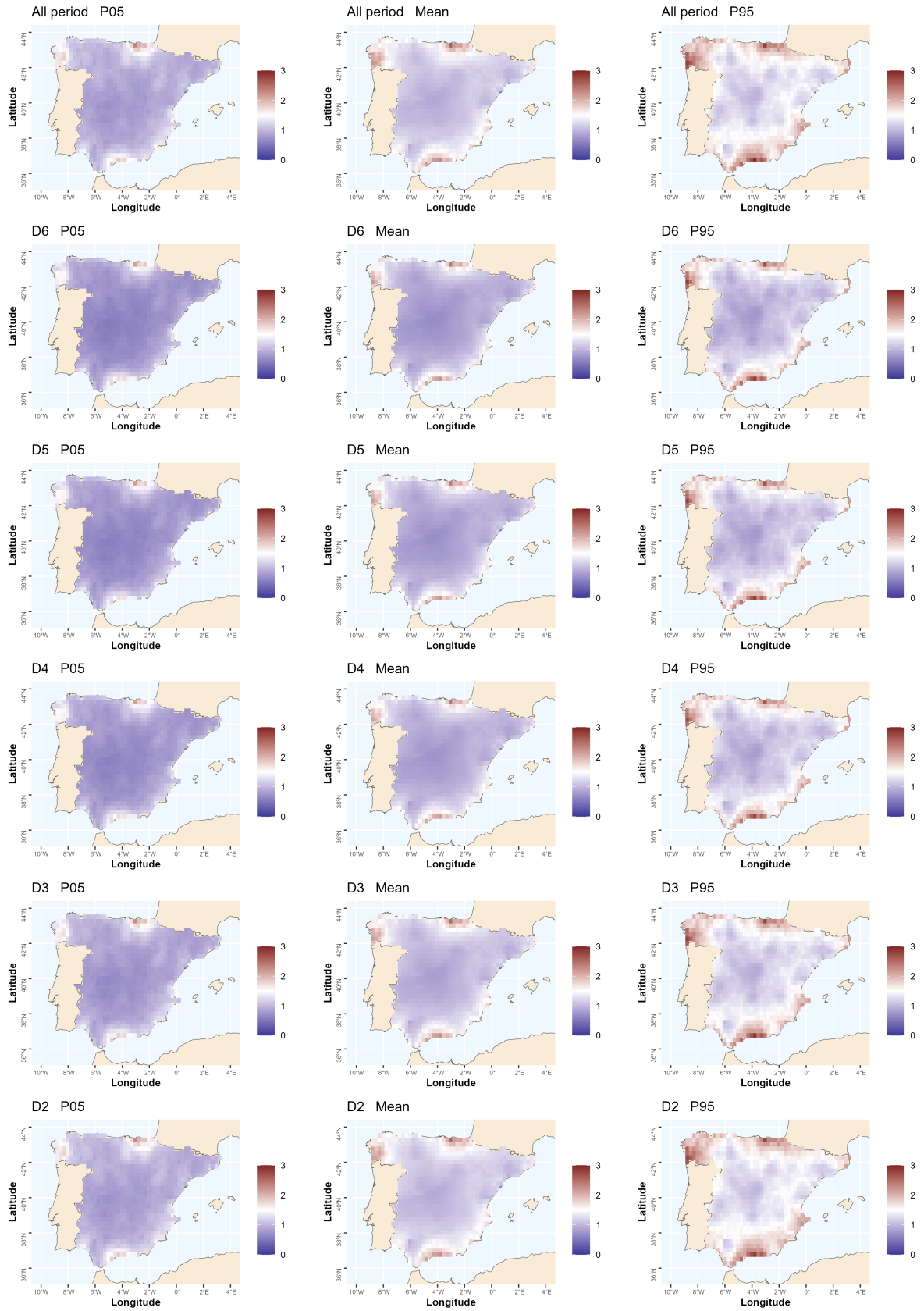


Figure 6: Posterior mean and 5th and 95th percentiles of the average increments, in $^{\circ}\text{C}$, over the entire period and over decades $D6$ to $D2$.

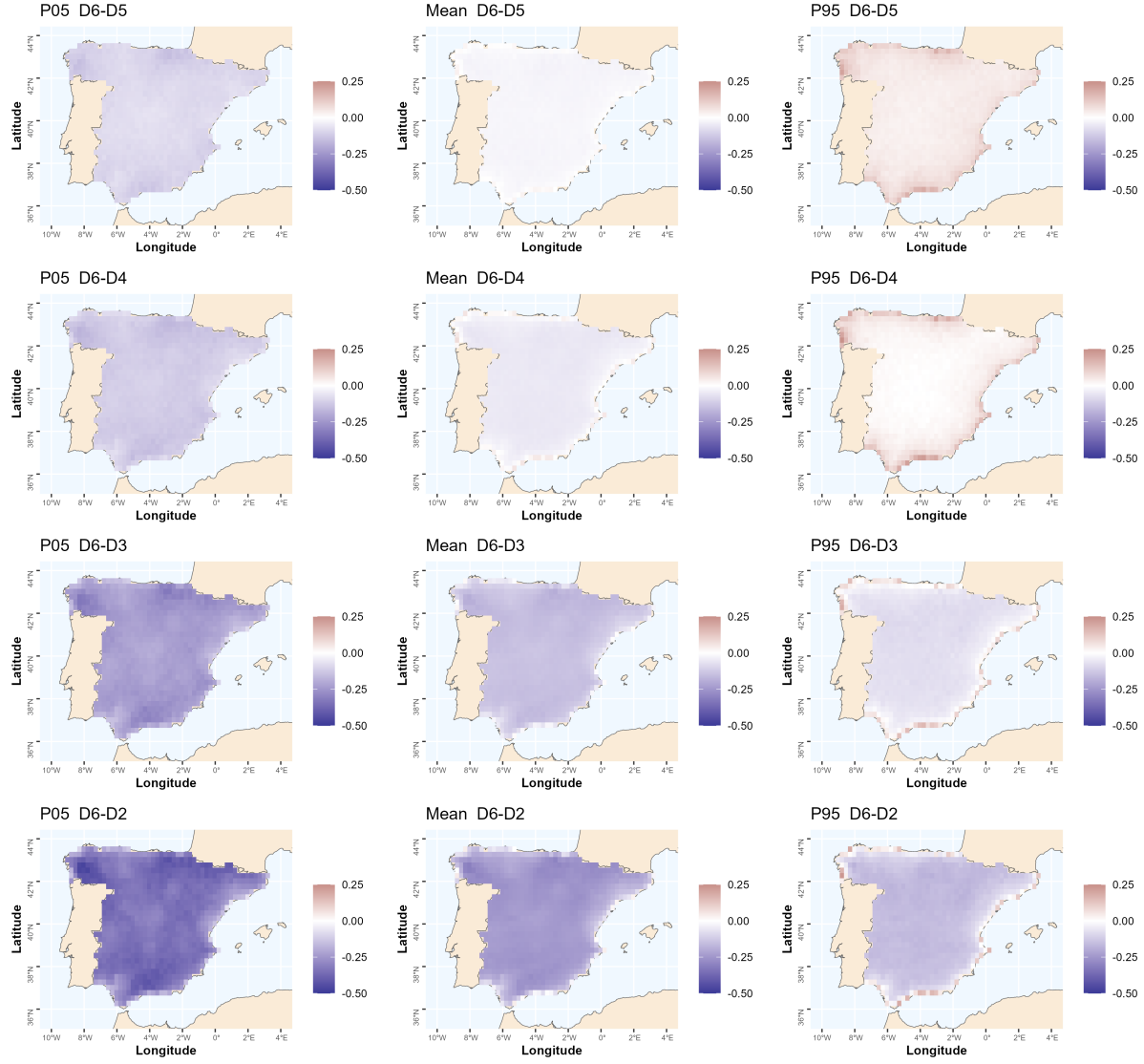


Figure 7: Posterior mean and 5th and 95th percentiles of $\bar{J}_{D_6, JJA}(\mathbf{s}) - \bar{J}_{D, JJA}(\mathbf{s})$, in $^{\circ}\text{C}$, for D , corresponding to decades $D5$, $D4$, $D3$ and $D2$.

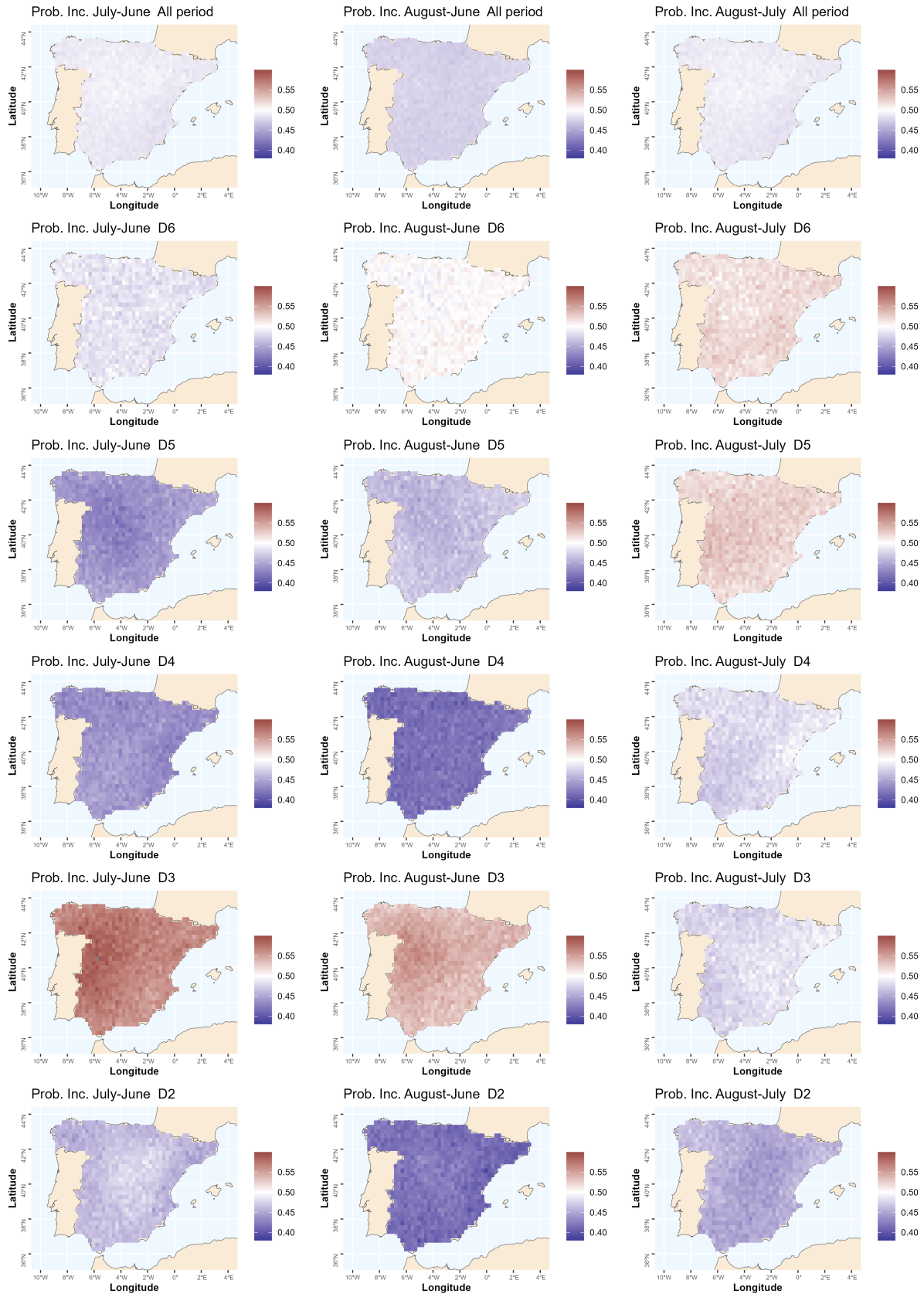


Figure 8: Probabilities of the average increments in July being higher than in June, and in August being higher than in June and July, over the entire period and over decades $D6$ to $D2$.

6 Comparing inland and coastal regions

Given the spatial differences in the behavior of average increments observed in certain decades between inland and some coastal areas, we analyze the annual trends of these increments for inland regions and for three coastal regions, Northern, Eastern, and Southern. The inland area is defined as grid points located more than 50 km from the coastline; the regions considered are shown in Figure 9. The plot showing the annual evolution in the four considered areas, right-hand plot in Figure 9, does not reveal any significant differences in annual trends between inland and coastal regions, nor among the three coastal regions themselves.

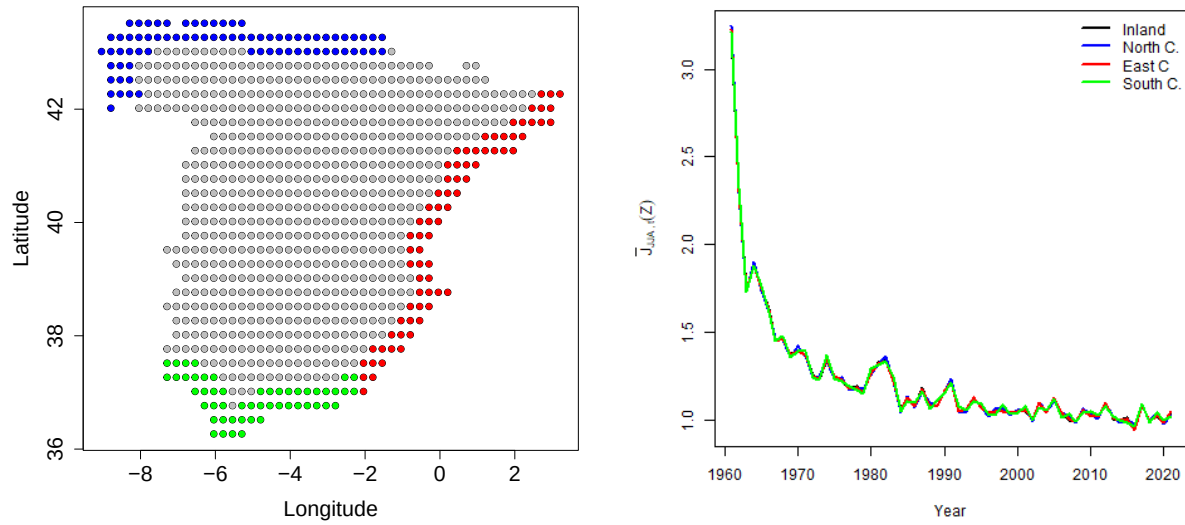


Figure 9: Left: Map of the considered areas: inland (gray), North coast (blue), East coast (red) and South coast (green). Right: Posterior mean of the annual average of the increments across four different areas.

References

- Fuglstad GA, Simpson D, Lindgren F, Rue H (2019) Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association* 114(525):445–452
- Lindgren F, Bolin D, Rue H (2022) The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spatial Statistics* 50:100599