

# **Métodos Hamiltonian Monte Carlo para la estimación de modelos de series climáticas**



**Alejandro Camón Fernández**  
Trabajo de fin de grado en Matemáticas  
Universidad de Zaragoza

Director del trabajo:  
Jesús Asín Lafuente  
9 de septiembre de 2022

# Summary

This work presents a detailed explanation of the HMC algorithm used for bayesian inference and an application to estimate, in the bayesian framework, a new proposed autoregressive model for the maximum daily temperatures.

The proposed model is based on the previous work of Castillo-Mateo et al. (2022) ([1]), where the time structure of mean was modeled. A more flexible structure is considered modeling also the variance and including interaction terms to reflect a seasonal variability of trend and persistence. The model has easily interpretable terms, it is able to represent the short and long-term dynamics of the temperatures, specially in relation to the effect of a possible climate change. Also a procedure of selection of covariates is designed and all the estimation process is implemented using RStan library, in the R workspace.

Models are fitted to series in a database built with data obtained by 18 stations placed in Aragón and its surroundings during a period of over 60 years. The estimation with the HMC-NUTS algorithm is possible but computationally slow. The results show progress towards a more complete model, because both, a non-constant variance and the addition of seasonal-trend and seasonal-persistence interactions, are necessary in Aragón series.

# Índice general

<b>Summary</b>	<b>II</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. La estimación bayesiana utilizando el algoritmo HMC</b>	<b>3</b>
2.1. Inferencia bayesiana y algoritmo MCMC . . . . .	3
2.1.1. El algoritmo MCMC . . . . .	4
2.1.2. Muestreo de Gibbs . . . . .	4
2.1.3. Control sobre la convergencia de la estimación MCMC . . . . .	5
2.2. El algoritmo HMC . . . . .	5
2.3. Software para la estimación bayesiana . . . . .	6
<b>3. Modelos estadísticos para la temperatura máxima diaria</b>	<b>8</b>
3.1. Análisis exploratorio de los datos . . . . .	8
3.2. Modelización estadística de las series diarias de temperatura máxima . . . . .	10
3.2.1. Modelo propuesto para la serie diaria de temperatura máxima en una localidad	11
3.3. Procedimiento para la selección de términos en los submodelos de media y de varianza	13
3.3.1. Resultados del procedimiento de selección del modelo . . . . .	14
3.4. Resultados del ajuste de los modelos bayesianos . . . . .	15
3.4.1. Grado de ajuste y crítica de los modelos . . . . .	15
3.4.2. Estructura de los predictores lineales para los submodelos de media y de varianza	15
3.4.3. Estimación del cambio observado en el valor medio de la distribución marginal de la temperatura máxima diaria . . . . .	19
3.4.4. Resultados de otros autores . . . . .	19
<b>4. Conclusiones y trabajo futuro</b>	<b>21</b>
4.1. Líneas de trabajo abiertas . . . . .	22
<b>Anexo A: Análisis exploratorio</b>	<b>24</b>
<b>Anexo B. Resultados de la modelización bayesiana</b>	<b>29</b>



# Capítulo 1

## Introducción

El fuerte incremento en los últimos años de la investigación en modelización bayesiana, Banerjee et al. ([2]), está asociado a la necesidad, en numerosos campos científicos, de desarrollar modelos estadísticos temporales y espacio-temporales. Se trata de modelos que deben considerar aspectos de la dinámica temporal de la variable respuesta y la posible relación con otras variables que se consideran explicativas; estas pueden ser medidas asociadas a causas, o variables relacionadas que se conocen antes en el tiempo o que son más fáciles de registrar que la propia respuesta.

También se puede representar la dependencia espacial. Habitualmente la correlación espacial entre posiciones geográficas se expresa como función de la distancia geográfica, pero también puede asociarse a otras distancias relativas a la naturaleza de la variable, basadas en altitud, fenotipos, genética,... Estos modelos son complejos en cuanto al número de parámetros y su estimación es factible dentro del marco bayesiano y no es abordable en el marco frecuentista.

Entre las líneas donde se avanza en la metodología bayesiana, se encuentra la definición de nuevos algoritmos que permitan reducir el esfuerzo en computación asociado al trabajo con grandes bases de datos. Estos avances se requieren para mejorar las prestaciones del algoritmo básico Markov Chain Monte Carlo (MCMC). Específicamente, los algoritmos objeto de este Trabajo Fin de Grado (TFG) son el denominado Hamiltonian Monte Carlo (HMC), Gelman et al. ([3]), y la modificación de este que plantea el algoritmo No U-Turns Sampler (NUTS), Hoffman y Gelman ([4]), utilizado en la librería *RStan*, ampliamente usada en modelización bayesiana.

En este TFG aplicaremos estos algoritmos para encontrar un modelo estadístico capaz de caracterizar la distribución de las temperaturas máximas diarias ( $T_{max}$ ) en varios lugares de Aragón y su entorno. Un objetivo de particular interés es la estimación del cambio observado en la distribución. Dado que se requieren medidas de variabilidad de ese cambio, no es suficiente utilizar herramientas descriptivas, sino que será necesario disponer de una modelización estadística en la escala más próxima a la realidad, en este caso la escala diaria. Para ello nos basaremos en el trabajo previo de Castillo-Mateo et al. ([1]), que utiliza modelos bayesianos espacio-temporales para estimar la media de la distribución de probabilidad de la temperatura máxima diaria.

El modelo que se va a plantear tendrá en cuenta la estacionalidad de la variable a lo largo del año, la autocorrelación debida a la persistencia de las condiciones atmosféricas de días previos y, finalmente, la evolución a largo plazo, que va a expresar el posible efecto del calentamiento global. A diferencia del trabajo de Castillo-Mateo et al, se va a plantear la modelización de la serie completa, lo que supone considerar épocas del año con distinta variabilidad de  $T_{max}$ , por ello se requiere modelizar tanto la media como la varianza de la distribución a lo largo del año.

El objetivo fundamental en esta aplicación del algoritmo es estimar el cambio observado en esta zona de la Cuenca del Ebro a lo largo de los últimos 60 años, y estudiar si el cambio es homogéneo tanto espacial como temporalmente, es decir, la modelización debe incluir términos para valorar si ha habido un cambio similar en las distintas estaciones del año y en las distintas zonas de la región de estudio. Para esto debe considerarse un modelo que permita analizar un posible mayor calentamiento en ciertas épocas del año. Esto implica la utilización de interacciones capaces de estimar una tendencia distinta

para invierno y para verano. Además, nos interesa comparar estadísticamente el cambio estimado en las zonas más áridas del centro de la cuenca frente al obtenido para otras zonas climáticas.

La memoria incluye un capítulo para revisar el procedimiento usual de la estimación bayesiana MCMC, donde también se presentarán con detenimiento los pasos del algoritmo HMC y el fundamento del algoritmo NUTS, una alternativa que supone mejoras en la velocidad de convergencia. El capítulo 3 se dedica a la aplicación de esta metodología de estimación para modelar la serie de temperatura máxima diaria en observatorios del entorno de Aragón. En particular, se presentará (i) una revisión crítica del modelo de Castillo-Mateo et al. ([1]), (ii) el planteamiento de un modelo para ser aplicado a la serie diaria en todo el año, (iii) la propuesta de un procedimiento de preselección de los componentes del modelo para cada serie, (iv) la aplicación de la estimación bayesiana usando el algoritmo NUTS, (v) el estudio en profundidad de algunos de los términos ajustados, (vi) una estimación del cambio neto observado en 50 años en la región y (vii) una comparación de los resultados con trabajos previos. También se identifican patrones espaciales gracias a la representación de los parámetros estimados en mapas obtenidos mediante un suavizado (LOESS). El último capítulo resumirá las conclusiones y planteará las líneas de trabajo abiertas. Se incluyen con la memoria anexos que recogen con más detalle el análisis exploratorio de la distribución observada en  $Tmax$ , los resultados del algoritmo propuesto para la selección del modelo en cada una de las series, así como resúmenes de los modelos bayesianos estimados, su crítica y su explotación, incluyendo el estudio en profundidad de algunos términos del modelo.

## Capítulo 2

# La estimación bayesiana utilizando el algoritmo HMC

Este capítulo incluye una revisión de conceptos básicos de inferencia bayesiana. Posteriormente se revisa el algoritmo MCMC y una sección se dedica a presentar el HMC, Gelman et al. ([3]). Finalmente, se consideran algunas opciones de software para realizar la estimación.

### 2.1. Inferencia bayesiana y algoritmo MCMC

La inferencia bayesiana plantea una perspectiva distinta a la *frecuentista* en cuanto a la estimación de los parámetros de un modelo, afirmando que cada parámetro es una variable aleatoria con una distribución de probabilidad, en lugar de un valor concreto desconocido. Por lo tanto, el objetivo de la inferencia bayesiana es estimar la distribución de los parámetros.

Así, sea  $\theta$  el vector de parámetros que queremos estimar y sean  $y$  los datos observados. Se pueden utilizar la definición de probabilidad condicional y la Regla de Bayes (de ahí el nombre de *inferencia bayesiana*) para concluir la siguiente proporcionalidad:

$$P(\theta|y) = \frac{P(\theta, y)}{P(y)} = \frac{P(y|\theta)P(\theta)}{P(y)} \propto P(y|\theta)P(\theta).$$

Puesto que la distribución marginal de la respuesta  $P(y)$  será igual para todos los valores de  $\theta$ , podemos considerar que la distribución de  $\theta$  condicionada a  $y$  es una expresión reescalada de  $P(y|\theta)P(\theta)$ , donde el valor de  $P(y)$  es una constante. Las que sí deben establecerse explícitamente son la distribución de probabilidad de  $y$  condicionado a  $\theta$ , que se corresponde con la *verosimilitud* frecuentista y debe estar de acuerdo con la naturaleza aleatoria de la respuesta, y la distribución de probabilidad de  $\theta$ , llamada *distribución a priori* de los parámetros.

La distribución de probabilidad de  $\theta$  condicionada a  $y$  es la distribución que queremos estimar, denominada *distribución a posteriori* de los parámetros. El procedimiento de estimación busca hacer máxima la distribución a posteriori o, equivalentemente,  $P(y|\theta)P(\theta)$ .

Una vez obtenida la distribución a posteriori de un parámetro  $\theta_i$ , se puede estimar el valor medio a posteriori o la moda a posteriori del parámetro. También se puede obtener un percentil de esa distribución o un intervalo de credibilidad para el parámetro, por ejemplo, el intervalo de credibilidad para  $\theta_i$  al 95 % (IC95) se construye a partir de los percentiles 0,025 y 0,975 de su distribución a posteriori.

Además, el modelo bayesiano permitirá obtener muestras de la distribución a posteriori de características de la variable respuesta  $y$ , como su valor medio o su varianza en condiciones dadas por posibles covariables. En consecuencia, se pueden obtener muestras de la respuesta en situaciones no observadas  $\tilde{y}$  a partir de su distribución predictiva a posteriori.

### 2.1.1. El algoritmo MCMC

La inferencia requiere considerar distribuciones a posteriori donde se evalúan integrales multidimensionales cuya solución analítica es muy difícil de obtener. Como alternativa, en la estimación bayesiana se usan métodos de Monte Carlo basados en cadenas de Markov (MCMC).

Los métodos MCMC son un tipo de algoritmo utilizado para el muestreo de una distribución de probabilidad y se basan en la construcción de cadenas de Markov que tienen la distribución buscada como distribución estacionaria. Dado un valor inicial para los parámetros, se simulan sucesivamente valores de una distribución de probabilidad ‘sencilla’, donde cada valor generado depende solo del anterior valor simulado, de ahí que la simulación corresponda a una cadena de Markov. Con una adecuada distribución del sorteo en cada paso, la cadena de Markov tendrá como distribución estacionaria  $p(\theta|y)$ . Por lo tanto, a partir de un paso en una trayectoria suficientemente larga, los valores simulados del parámetro  $\theta$  corresponden a esta distribución estacionaria, es decir, a la distribución a posteriori  $p(\theta|y)$ . De ahí, que la estimación bayesiana MCMC se base en la parte final de la trayectoria simulada de la cadena de Markov.

El algoritmo de Metropolis supone una primera propuesta para obtener simulaciones de cadenas de Markov cuya distribución estacionaria coincide con la distribución a posteriori que se pretende estimar.

Este algoritmo consiste en una variación de un camino aleatorio en el que, a cada paso  $t$  de la trayectoria, se le añade una cierta probabilidad de rechazar el movimiento aleatorio propuesto para el parámetro  $\theta$  y, en su lugar, quedarse en el mismo sitio. Los pasos seguidos en cada iteración de la trayectoria serán los siguientes:

1. Tomamos un punto de partida  $\theta^0$ , con  $p(\theta^0|y) > 0$ , de una distribución inicial  $p_0(\theta)$ .
2. Para cada iteración  $t = 1, 2, \dots$ :
  - a) Tomamos un candidato a nueva posición  $\theta^*$  de una distribución de salto del tiempo  $t - 1$  al tiempo  $t$ ,  $J_t(\theta^*|\theta^{t-1})$ . Esta distribución de salto debe ser simétrica, cumpliendo que  $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$ , para todo  $\theta_a, \theta_b$  y  $t$ .
  - b) Calculamos el ratio de probabilidades:

$$r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)}.$$

- c) Decidimos si aceptar o rechazar el candidato a  $\theta^t$  de la siguiente manera:

$$\theta^t = \begin{cases} \theta^* & , \text{ con probabilidad } \min(r, 1) \\ \theta^{t-1} & , \text{ en otro caso} \end{cases}.$$

El algoritmo de Metropolis-Hastings propone una mejora para reducir la frecuencia de rechazo, lo que puede acelerar la convergencia respecto al algoritmo Metropolis. La distribución de los saltos no tiene que ser simétrica, a diferencia de lo que ocurre en el Metropolis, así que se introduce una variación en el ratio de densidades para compensar esta asimetría:

$$r = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)}.$$

### 2.1.2. Muestreo de Gibbs

El muestreo de Gibbs es un método MCMC en el que se divide el vector de parámetros  $\theta$  en  $d$  subvectores y, con cada paso de tiempo o iteración  $t$ , se actualizan dichos subvectores en cierto orden condicionándolos a la última versión del resto de parámetros. De esta manera, de los subvectores ya actualizados, se utiliza el valor del paso  $t$  ( $\theta_i^t$ ) y, de los parámetros pendientes de actualizar, se usa el valor del paso  $t - 1$  ( $\theta_j^{t-1}$ ).

Las densidades propuestas en el muestreo de Gibbs coinciden con las distribuciones a posteriori condicionadas de los parámetros, de modo que la probabilidad de aceptar la innovación en cada paso es siempre 1. Esto implica que en ningún paso hay rechazo.



### 2.1.3. Control sobre la convergencia de la estimación MCMC

Es importante asegurar la convergencia del algoritmo MCMC en el ajuste del modelo propuesto. Para ello, es usual simular al menos dos cadenas con distintos puntos iniciales y utilizar herramientas que permitan establecer si las cadenas convergen a la misma distribución límite. Una herramienta gráfica sencilla corresponde a representar en una misma figura la trayectoria de valores simulados por cada cadena, los trazos deberían solapar.

El *factor de reducción potencial de escala*  $\hat{R}$  mide la convergencia con un cálculo que compara la varianza entre las cadenas y la varianza dentro de cada cadena. Para cada una de las  $M$  cadenas se obtiene  $s_m^2$ , la cuasivarianza muestral del parámetro  $\theta$  calculada a partir de  $N$  elementos de la trayectoria, para la cadena  $m$  se tiene  $\theta_m^{(1)}, \theta_m^{(2)}, \dots, \theta_m^{(N)}$  cuya media es  $\bar{\theta}_m$ . Con ellas se obtiene  $W$ , el valor medio de la cuasivarianza dentro de las cadenas.

$$s_m^2 = \frac{1}{N-1} \sum_{n=1}^N (\theta_m^{(n)} - \bar{\theta}_m)^2, \quad W = \frac{1}{M} \sum_{m=1}^M s_m^2.$$

Así, se consigue un estimador de la variación con una mixtura de  $W$  y la varianza 'cross-chain',  $B$ :

$$B = \frac{N}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2, \quad \hat{v}ar^+(\theta|y) = \frac{N-1}{N} W + \frac{1}{N} B.$$

Finalmente, la expresión de  $\hat{R}$  es:

$$\hat{R} = \sqrt{\frac{\hat{v}ar^+(\theta|y)}{W}}.$$

En el caso ideal, el valor de  $\hat{R}$  para cada parámetro debe estar cerca de 1 y Gelman aconseja que sea menor que 1.1. En caso de no converger, se debería aumentar el número de iteraciones.

Por otra parte, se obtiene para cada parámetro el *tamaño de muestra efectivo*, que refleja el número efectivo de sorteos en la simulación que aportan información para el parámetro.

El tamaño de muestra efectivo de un estimador se obtiene a partir de las  $N$  simulaciones de  $M$  cadenas con la expresión  $\hat{N}_{eff} = MN/\hat{\tau}$ , donde  $\hat{\tau} = 1 + 2 \sum_{r=1}^{2m+1} \hat{\rho}_r$ . A su vez, la estimación de la autocorrelación de retardo  $r$ , denotada  $\hat{\rho}_r$ , combina en una las estimaciones de esta misma autocorrelación de cada una de las  $M$  cadenas,  $\hat{\rho}_{r,m}$ :

$$\hat{\rho}_r = 1 - \frac{W - \frac{1}{M} \sum_{m=1}^M s_m^2 \hat{\rho}_{r,m}}{\hat{v}ar^+(\theta|y)}.$$

## 2.2. El algoritmo HMC

El algoritmo de Metropolis-Hastings converge a través de una caminata aleatoria, pero este proceso se puede mejorar incorporando, junto al vector de parámetros  $\theta$  de la trayectoria, un vector de momentos  $\phi$  que ayude a guiar la simulación hacia la dirección óptima. Para este proceso definiremos una matriz diagonal de masas  $M$ , un valor para el número de saltos por iteración  $L$ , y un valor  $\varepsilon$  que influirá en la amplitud de los saltos tanto de  $\phi$  como de  $\theta$ . La Figura 2.1 representa el esquema de avance en una iteración, que se describe con los siguientes pasos:

1. Al comienzo de la iteración tomamos un valor inicial para el vector de momentos, obtenido aleatoriamente de su distribución  $\phi \sim N(0, M)$ , es decir,  $\phi_j \sim N(0, M_{jj})$ ,  $j = 1, \dots, \dim(\phi)$ .
2. El paso principal del algoritmo es realizar  $L$  saltos que consisten, cada uno, en las siguientes actualizaciones:
  - a) Añadimos a  $\phi$  medio salto  $\varepsilon$  en la dirección del gradiente del logaritmo de la distribución a posteriori:

$$\phi \leftarrow \phi + \frac{1}{2} \varepsilon \frac{d \log p(\theta|y)}{d\theta}$$

b) Usamos  $\phi$  para actualizar  $\theta$ :

$$\theta \leftarrow \theta + \varepsilon M^{-1} \phi$$

c) Una vez actualizado  $\theta$ , hacemos que  $\phi$  dé otro medio salto en la dirección del nuevo gradiente:

$$\phi \leftarrow \phi + \frac{1}{2} \varepsilon \frac{d \log p(\theta|y)}{d\theta}$$

3. Al final de este proceso de  $L$  saltos obtenemos los candidatos a nueva posición de  $\theta$  y  $\phi$ , que denominaremos  $\theta^*$  y  $\phi^*$ , pero aún existe la posibilidad de rechazar este avance, por lo que definiremos un ratio de aceptación utilizando los candidatos de los pasos anteriores  $\theta^{t-1}$  y  $\phi^{t-1}$ :

$$r = \frac{p(\theta^*|y)p(\phi^*)}{p(\theta^{t-1}|y)p(\phi^{t-1})}$$

4. Para terminar la iteración decidiremos aleatoriamente si aceptamos los parámetros candidatos o no de la siguiente manera:

$$\theta^t = \begin{cases} \theta^* & , \text{ con probabilidad } \min(r, 1) \\ \theta^{t-1} & , \text{ en otro caso} \end{cases}$$

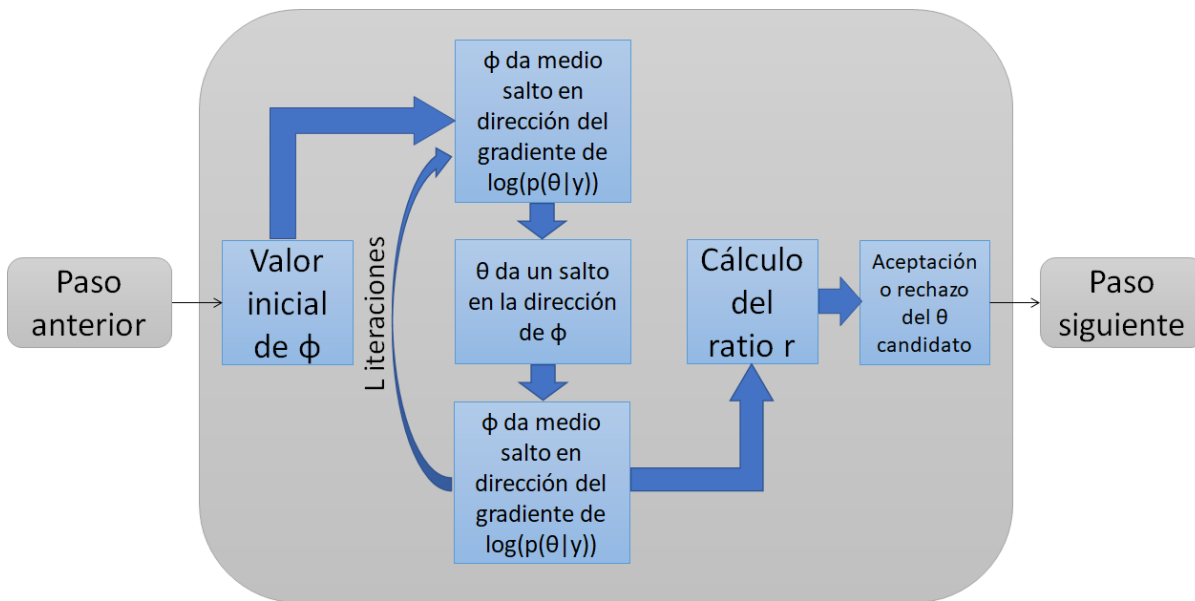


Figura 2.1: Esquema de un paso intermedio del algoritmo HMC.

El software que utilizaremos para la estimación de los modelos es Stan, a través de su versión en R: *RStan*. Este software usa una versión del algoritmo HMC llamada No-U-Turns Sampler (NUTS), Hoffman y Gelman ([4]), que favorece la convergencia del proceso cambiando la dirección de la trayectoria de manera aleatoria cuando detecta que el nuevo paso está dando un giro brusco que puede llevar a que la trayectoria quede atrapada en un bucle o una órbita. Estos autores evalúan que NUTS genera un muestreo más eficiente puesto que dará lugar a mayor valor de  $\hat{N}_{eff}$  con menos computación que el HMC básico. El nombre procede de que a estos giros bruscos se los denomina en inglés 'U-turns'.

### 2.3. Software para la estimación bayesiana

Existe una amplia variedad de software para realizar la inferencia bayesiana en R, en Python, en Matlab y en BUGS, entre otros.

En el entorno R destaca *jagsUI* que permite trabajar en el entorno *JAGS*, llamado así por las siglas de Just Another Gibbs Sampler, Plummer ([5]). Es un programa para análisis de modelos bayesianos jerárquicos usando simulaciones MCMC con muestreo de Gibbs. *JAGS* fue creado con el objetivo de permitir a los usuarios escribir sus propias funciones, distribuciones y 'samplers'. Es una plataforma de experimentación para nuevas ideas en modelización bayesiana. Además, es muy útil para plantear una introducción al modelado estadístico, favorece la comprensión de la estructura de los modelos bayesianos porque en su programación se requiere explicitar todos los elementos que definen el modelo: distribución de la respuesta para construir la verosimilitud, parámetros y su distribución a priori.

En las primeras fases de este proyecto se trabajó con esta librería, para construir algunos modelos sencillos. Sin embargo, se descartó su uso en la aplicación de este TFG, debido al enorme tiempo de computación que requiere la estimación de modelos para la varianza.

*BAMLSS* es un paquete en el entorno R que proporciona un marco de trabajo computacional modular para modelos de regresión bayesiana flexibles. Utiliza su propia versión de algoritmo MCMC, cuyos pasos intercala con un algoritmo de Newton-Raphson, Umlauf et al. ([6]). Una de sus principales ventajas es su uso intuitivo debido a que la expresión del modelo se realiza mediante la introducción de una fórmula del tipo que utiliza la función 'lm' de R, que incluye la respuesta, las variables predictoras, la base de datos y el tipo de distribución de la respuesta.

Finalmente, Stan es una plataforma de última generación para el modelado estadístico y el cálculo estadístico de alto rendimiento, Stan Development Team ([7]). Permite realizar inferencia bayesiana con algoritmos HMC y NUTS. En particular, existe *RStan* para el entorno R, cuya ventaja es una reducción considerable en el tiempo de computación respecto a *JAGS*. Esta librería es utilizada en los ejemplos desarrollados en el libro de Gelman et al. ([3]). *RStan* contiene funciones para la explotación de la distribución a posteriori, permitiendo la elaboración de gráficas de los resultados obtenidos, de distribuciones marginales de los términos de interés o la predicción de nuevos valores. Otras librerías complementarias para la crítica y explotación de los modelos obtenidos con *RStan* son *coda*, *bayesplot* o *rstantools*.

Una desventaja de *Rstan* es que requiere la construcción de 'scripts' con la definición explícita de la verosimilitud y las distribuciones a priori. Para simplificar este inconveniente, el equipo de Stan ha generado la librería *Rstanarm*, Muth et al. ([8]), de modo que los usuarios pueden utilizar una definición de modelos estándar similar a la habitual en la función 'lm' de R. Los modelos propuestos en este TFG, sin embargo, no se incluyen entre los disponibles en *Rstanarm*, de modo que no ha sido posible utilizar esta herramienta facilitadora.

## Capítulo 3

# Modelos estadísticos para la temperatura máxima diaria

Para ejemplificar el uso de los algoritmos explicados, vamos a diseñar un modelo que represente la evolución de las temperaturas máximas diarias de 18 estaciones meteorológicas de Aragón y su entorno. Los modelos ajustados serán útiles para estimar el calentamiento observado desde mitad del siglo pasado (de 1953 a 2015) en esta región.

Existen numerosas referencias en la literatura para abordar la modelización de series de temperatura diaria, donde los componentes considerados en el modelo dependen de los objetivos que se pretenden alcanzar, por ejemplo, caracterizar la distribución o disponer de un simulador de tipos de tiempo, Kleiber et al. ([9]), Verdin et al. ([10]).

La línea seguida aquí se basa en el trabajo de Castillo-Mateo et al. ([1]), pero en lugar de aplicar la misma versión del algoritmo MCMC, nuestra estimación se obtendrá utilizando un algoritmo HMC, concretamente el NUTS.

El capítulo presenta en primer lugar la base de datos, junto con un análisis exploratorio de las características de la distribución de temperatura máxima diaria ( $T_{max}$ ), que mostrarán la estacionalidad de media y varianza y la existencia de tendencias temporales distintas en cada época del año y en cada lugar. Una posterior sección se dedica a revisar los principales elementos de la modelización de Castillo-Mateo et al. y, tras ello, se plantea el modelo que se ajusta en este trabajo, indicando las diferencias entre ambas propuestas. El proceso de estimación incluye el planteamiento de un procedimiento de selección de términos de los submodelos de media y varianza, que será descrito en la Sección 3.3. La sección final está dedicada al análisis de los resultados obtenidos y del cambio climático observado.

### 3.1. Análisis exploratorio de los datos

La región de Aragón presenta una considerable diversidad climática debido a su variado relieve y a su posición entre dos mares y dos cordilleras. Esto hace que la zona sea difícil de modelar, lo que ha motivado varios trabajos que tratan de afrontar este problema tanto a nivel regional, Castillo-Mateo et al. ([1]), como extendiendo el planteamiento a nivel nacional, Peña-Angulo et al. ([11]), ya que el resto de la Península Ibérica también presenta retos al modelado.

La Figura 3.1 representa un mapa con el relieve de la región considerada, expresando la altitud con curvas de nivel, y la posición de los 18 observatorios cuyas series diarias 1953-2015 constituyen la base de datos, construida a partir de los registros de AEMET, a los cuales agradecemos su colaboración.

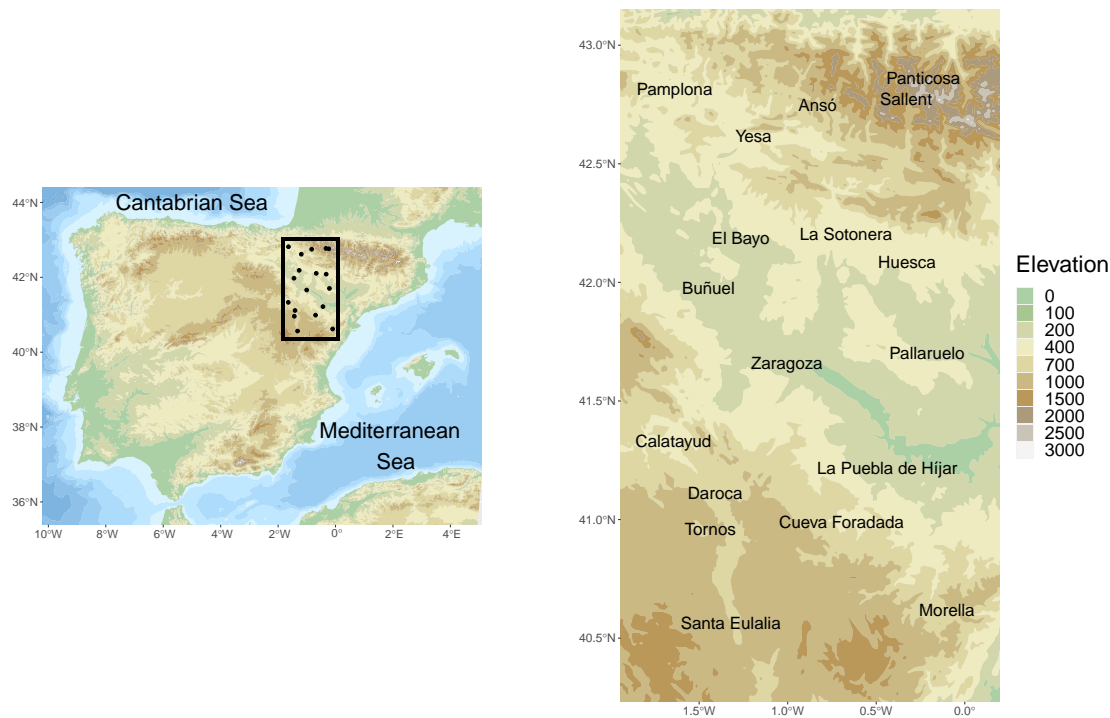


Figura 3.1: Mapas del relieve de la Península Ibérica y del entorno de Aragón donde se indica la localización de los 18 observatorios considerados.

A continuación, se realizará una exploración de la base de datos con el objetivo de detectar algunos de las características que deberían incluirse en la propuesta del modelo estadístico.

La Figura 3.2a presenta el diagrama de cajas de las temperaturas máximas diarias por meses. Podemos observar en ella el obvio comportamiento estacional de las temperaturas a lo largo del año. La gráfica de la Figura 3.2b muestra el diagrama de cajas equivalente de las anomalías, definidas como  $T_{max}$  menos la media de las temperaturas observadas en cada día del año. Aquí se puede ver una estacionalidad en la variabilidad de las temperaturas, esto último se evidencia, por ejemplo, al comparar los rangos intercuartílicos, que van desde  $5,1^{\circ}\text{C}$  en enero hasta  $7,275^{\circ}\text{C}$  en junio.

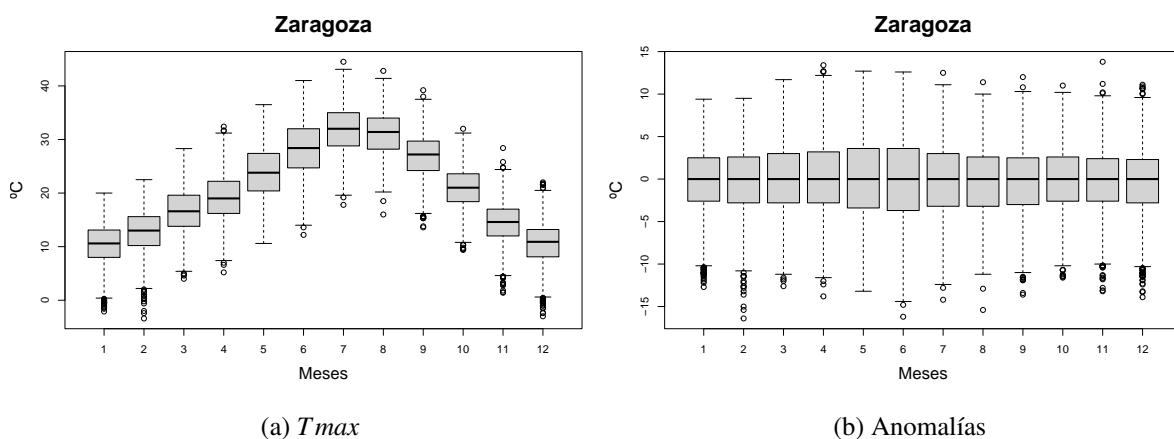


Figura 3.2: Diagramas de cajas por meses de  $T_{max}$  en Zaragoza (izd.) y de sus anomalías (dch.).

La tendencia lineal ajustada para la serie de valor medio anual en 4 observatorios se representa en la Figura 3.3a. Zaragoza y Ansó presentan distintos ritmos de cambio de las temperaturas a lo largo de los años, una clara prueba de la variabilidad espacial en el calentamiento observado. Las pendientes asociadas a cada mes en Zaragoza, en la Figura 3.3b, evidencian que existe variabilidad en el calenta-

miento observado en las distintas épocas del año, es decir existe un comportamiento estacional en esta tendencia.

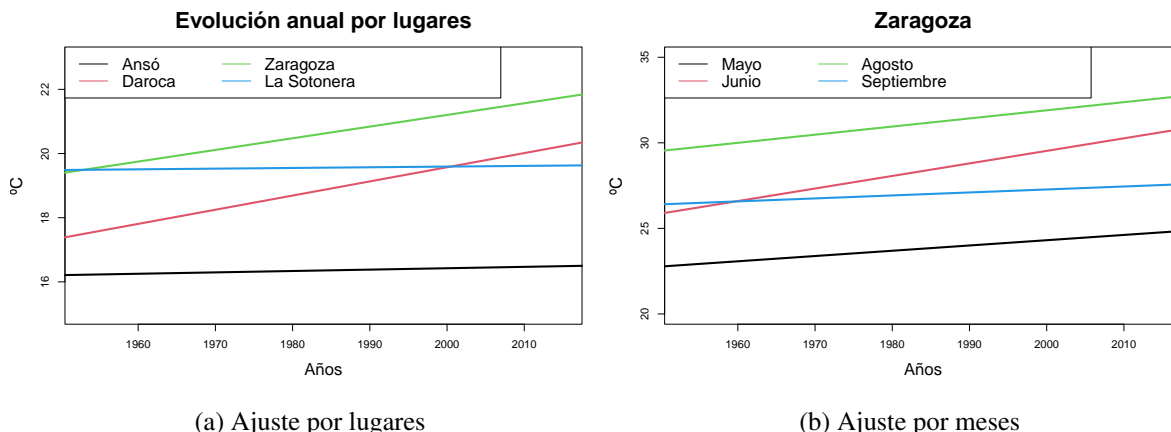


Figura 3.3: Ajustes lineales respecto a los años de las temperaturas medias anuales para cuatro observatorios (izd.) y de las temperaturas medias mensuales para cuatro meses en Zaragoza (dch.).

Una vez identificadas estas características de los datos observados, podemos concluir que nuestro modelo, para poder representar el comportamiento de las temperaturas con términos fácilmente interpretables, deberá reflejar los siguientes efectos:

- Tendencia al aumento de temperaturas a lo largo de los años, que trata de explicar el efecto del cambio climático.
- Persistencia de la temperatura del día anterior debida a la inercia de las temperaturas en la atmósfera.
- Estacionalidad de la temperatura media, del efecto de la tendencia y, posiblemente, de la persistencia.
- Variabilidad de las temperaturas no constante a lo largo del año, lo que requerirá una modelización de la varianza.

### 3.2. Modelización estadística de las series diarias de temperatura máxima

En el trabajo de Castillo-Mateo et al. ([1]), se tienen en cuenta los efectos antes mencionados (tendencia, persistencia y estacionalidad), junto con la elevación del terreno en cada lugar, para tratar de modelar la media de la distribución diaria de  $Tmax$  en un modelo bayesiano espacial. Este modelo se aplicó únicamente al periodo del año entre Mayo y Septiembre, ambos incluidos.

Llamaremos  $Y_{t\ell}(s)$  a la temperatura máxima diaria para el día  $\ell$ ,  $\ell = 2, \dots, L$  del año  $t$ ,  $t = 1, \dots, T$  en la posición  $s \in D$ , donde  $D$  es la región de estudio. De esta manera, para tratar de separar efectos fijos y aleatorios, así como distinguir claramente posición, año y día en la notación, se presenta la siguiente estructura para el modelo:

$$Y_{t\ell}(s) = \mu_{t\ell}(s; \theta_f) + \gamma_t(s) + \rho_Y(s) (Y_{t\ell-1}(s) - (\mu_{t\ell-1}(s; \theta) + \gamma_t(s))) + \epsilon_{t\ell}^Y(s), \quad (3.1)$$

donde  $\mu_{t\ell}(s; \theta_f)$  representa la componente de los efectos fijos,  $\gamma_t(s)$  la de los efectos aleatorios,  $\rho_Y(s)$  representa la persistencia respecto del día anterior mediante un término autorregresivo y  $\epsilon_{t\ell}^Y(s)$  es un error diario y local. Se especifica la estructura de efectos fijos en  $\mu_{t\ell}(s; \theta_f)$ :

$$\mu_{t\ell}(s; \theta_f) = \beta_0 + \alpha t + \beta_1 \sin(2\pi\ell/365) + \beta_2 \cos(2\pi\ell/365) + \beta_3 \text{elev}(s), \quad (3.2)$$

siendo  $\beta_0$  un término independiente global,  $\alpha$  un coeficiente de tendencia lineal global, los términos de seno y coseno con periodo de un año forman una componente estacional, y  $elev(s)$  es la elevación en la posición  $s$ . Se denotan estos parámetros por  $\theta_f = (\beta_0, \alpha, \beta_1, \beta_2, \beta_3)$ .

Para los efectos aleatorios se especifica:

$$\gamma_t(s) = \beta_0(s) + \alpha(s)t + \psi_t + \eta_t(s), \tag{3.3}$$

donde  $\beta_0(s)$  es un campo gaussiano (CG) espacial que corrige el efecto de la altitud,  $\alpha(s)$  es el CG asociado a la variabilidad espacial de la tendencia,  $\psi_t = \rho_\psi \psi_{t-1} + \lambda_t$  es un término de autoregresión anual que puede capturar la correlación entre años, y  $\eta_t(s)$  es un efecto aleatorio de términos independientes con dependencia espacial. Los parámetros de los efectos aleatorios se reúnen en  $\theta_r = (\rho_\psi, \sigma_\lambda^2, \sigma_{\beta_0}^2, \phi_{\beta_0}, \sigma_\alpha^2, \phi_\alpha)$ .

Los tres términos de error introducidos son gaussianos:  $\lambda_t \sim N(0, \sigma_\lambda^2)$ ,  $\eta_t(s) \sim N(0, \sigma_\eta^2)$  y  $\epsilon_{t\ell}^Y(s) \sim N(0, \sigma_\epsilon^2(s))$ . Sin embargo, como en el trabajo los términos asociados a  $\lambda_t$  y  $\eta_t(s)$  resultaron no ser significativos, solo nos interesa el efecto de  $\epsilon_{t\ell}^Y(s)$ , que añade variabilidad al modelo mediante una varianza  $\sigma_\epsilon^2(s)$  constante en el tiempo. Esta especificación no concuerda con el exploratorio, que mostraban que la varianza debería presentar, al menos, estacionalidad a lo largo del año, por lo que será una de las propuestas en nuestro modelo, añadiendo un submodelo para la varianza.

### 3.2.1. Modelo propuesto para la serie diaria de temperatura máxima en una localidad

En este trabajo se obtendrá un modelo local, es decir, un modelo con una estructura de predictores distinta para cada uno de los 18 observatorios. Se modelizará la varianza de las  $Tmax$  diarias, para poder estudiar la posible variabilidad espacial y temporal de esta. Además, nuestra propuesta será aplicada a todo el año.

Seguiremos asumiendo que la respuesta  $Y_{t\ell}$  ( $Tmax$ ) tiene, cada día, una distribución normal condicional a un vector de covariables que representan persistencia, tendencia y estacionalidad. En la Figura 3.4 se muestra un esquema con las dependencias entre variables que estarán representadas en el nuevo modelo propuesto. Se utilizan las notaciones  $\mu_{t\ell}$  y  $\sigma_{t\ell}$  para referirse, respectivamente, al submodelo propuesto para la media de  $Tmax$  el día  $\ell$  del año  $t$  y al submodelo que representa su desviación típica.

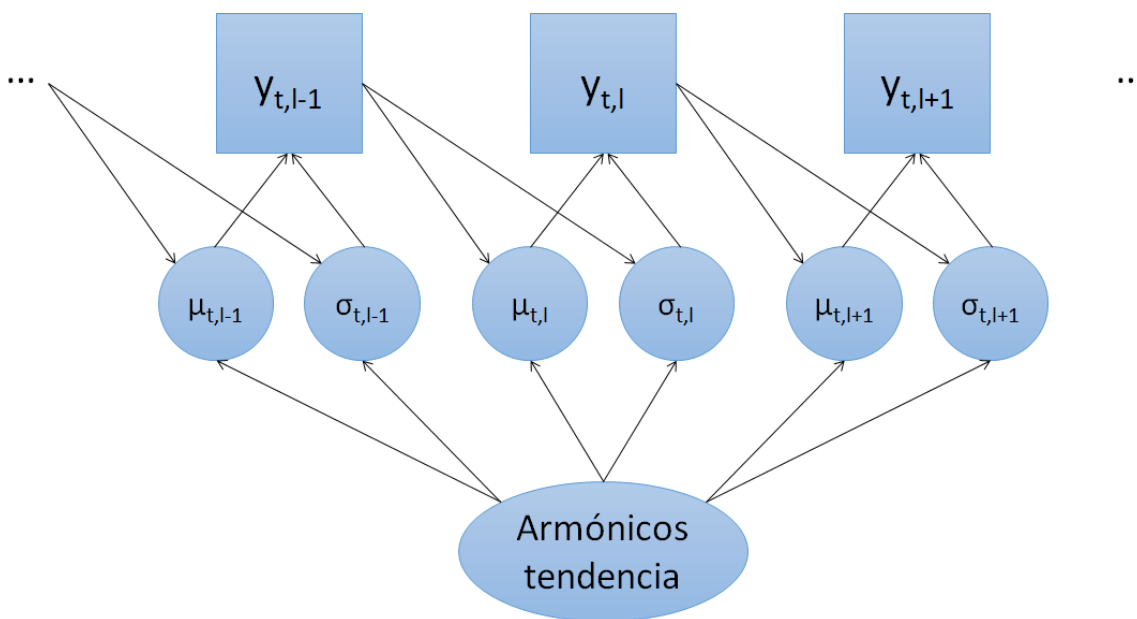


Figura 3.4: Esquema de las dependencias del modelo propuesto para cada elemento de la serie.

Los dos submodelos de media y de varianza tendrán la misma estructura base: un término independiente, un término de tendencia anual y un término autorregresivo que expresa una persistencia proporcional a la  $Tmax$  del día anterior. Cada uno de estos tres términos tiene a su vez un coeficiente independiente y una parte estacional representada por uno o varios armónicos.

$$\begin{aligned}\tilde{\mu}_{t\ell} := E[Y_{t\ell}|\mathbf{X}_{t\ell}] &= \beta_0 + \sum_{i=1}^N (\beta_{i,s}Si_\ell + \beta_{i,c}Ci_\ell) + \left( \alpha_0 + \sum_{i=1}^{N_\alpha} (\alpha_{i,s}Si_\ell + \alpha_{i,c}Ci_\ell) \right) t + \left( \rho_0 + \sum_{i=1}^{N_\rho} (\rho_{i,s}Si_\ell + \rho_{i,c}Ci_\ell) \right) Y_{t\ell-1} \\ \tilde{\sigma}_{t\ell}^2 := Var[Y_{t\ell}|\mathbf{X}_{t\ell}] &= \exp \left\{ \beta_0^\sigma + \sum_{i=1}^{N^\sigma} (\beta_{i,s}^\sigma Si_\ell + \beta_{i,c}^\sigma Ci_\ell) + \left( \alpha_0^\sigma + \sum_{i=1}^{N_\alpha^\sigma} (\alpha_{i,s}^\sigma Si_\ell + \alpha_{i,c}^\sigma Ci_\ell) \right) t \right. \\ &\quad \left. + \left( \rho_0^\sigma + \sum_{i=1}^{N_\rho^\sigma} (\rho_{i,s}^\sigma Si_\ell + \rho_{i,c}^\sigma Ci_\ell) \right) Y_{t\ell-1} \right\}.\end{aligned}$$

Se señalan a continuación algunas especificaciones del modelo propuesto:

- Como la varianza debe ser positiva, al predictor lineal del submodelo de la varianza se le aplica una transformación exponencial, de manera que el predictor será el logaritmo de la varianza estimada.
- $\ell = 1, \dots, L$ , con  $L = 366$  en años bisiestos y  $L = 365$  en el resto.
- La temperatura del día anterior será  $Y_{t\ell-1}$  todos los días del año salvo el 1 de enero, que tendrá como día anterior  $Y_{-1,366}$  si el anterior era un año bisiesto o  $Y_{-1,365}$  en caso contrario.
- $Ci_\ell = \cos(2\pi i(\ell - 1)/366)$  y  $Si_\ell = \sin(2\pi i(\ell - 1)/366)$ .
- El vector de covariables es  $\mathbf{X}_{t\ell} = (t, Y_{t\ell-1}, S1_\ell, C1_\ell, S2_\ell, C2_\ell, \dots, S_{\mathcal{N}_\ell}, C_{\mathcal{N}_\ell})$ , con  $\mathcal{N} = \max(N, N^\sigma)$ .
- $N_\alpha, N_\rho \leq N$  y  $N_\alpha^\sigma, N_\rho^\sigma \leq N^\sigma$ .

■ Llamaremos a los vectores de parámetros

$$\begin{aligned}\theta &= (\beta_0, \beta_{1,s}, \beta_{1,c}, \dots, \beta_{N,s}, \beta_{N,c}, \alpha_0, \alpha_{1,s}, \alpha_{1,c}, \dots, \alpha_{N_\alpha,s}, \alpha_{N_\alpha,c}, \rho_0, \rho_{1,s}, \rho_{1,c}, \dots, \rho_{N_\rho,s}, \rho_{N_\rho,c}) \text{ y} \\ \theta^\sigma &= (\beta_0^\sigma, \beta_{1,s}^\sigma, \beta_{1,c}^\sigma, \dots, \beta_{N,s}^\sigma, \beta_{N,c}^\sigma, \alpha_0^\sigma, \alpha_{1,s}^\sigma, \alpha_{1,c}^\sigma, \dots, \alpha_{N_\alpha^\sigma,s}^\sigma, \alpha_{N_\alpha^\sigma,c}^\sigma, \rho_0^\sigma, \rho_{1,s}^\sigma, \rho_{1,c}^\sigma, \dots, \rho_{N_\rho^\sigma,s}^\sigma, \rho_{N_\rho^\sigma,c}^\sigma)\end{aligned}$$

Dada la complejidad de los modelos bayesianos que se proponen, es necesario ajustarlos mediante MCMC. A continuación, se indican algunas consideraciones sobre el procedimiento de estimación:

- Las distribuciones a priori serán gaussianas y no informativas,  $\mathbf{N}(0, 1000)$ , para todos los parámetros.
- La estimación HMC se realizará utilizando dos cadenas, con un periodo de calentamiento de 500 simulaciones, y se consideran 2000 simulaciones (1000 por cadena), lo que da cadenas de 1500 simulaciones de longitud. En los casos en que esta longitud no es suficiente para la convergencia se recurrirá a 9000 simulaciones de calentamiento y guardar 2000 simulaciones totales, lo que da cadenas de 10000 simulaciones.
- Tras el proceso de ajuste se establecerá a partir de intervalos de credibilidad al 95% (IC95) si las covariables correspondientes tienen un efecto significativo en los submodelos de media y de varianza. Un parámetro se considerará significativo si y solo si su intervalo de credibilidad no contiene al 0. Sin embargo, en el caso de las funciones armónicas, hará falta que tanto el seno como el coseno de una frecuencia concreta sean no significativos para que el armónico se considere no significativo.



### 3.3. Procedimiento para la selección de términos en los submodelos de media y de varianza

El procedimiento de estimación bayesiano resulta computacionalmente muy costoso, por lo que se ha tenido que recurrir a una fase previa de preselección de los términos que pueden tener capacidad explicativa. En esta fase se ha diseñado un procedimiento donde se ajustan sucesivos modelos lineales no bayesianos a los datos de cada localización. Se prueban en primer lugar modelos ensayados que incorporan la información que describe la evolución atmosférica asociada a la inercia atmosférica a corto plazo que expresan el término autorregresivo y el ciclo estacional. Sobre esos modelos se analiza si existe evidencia de efectos asociados a la tendencia (cambio climático). Para ello utilizamos el siguiente procedimiento constructivo:

1. Modelo 0.0: Ajuste lineal de los datos respecto al valor de la temperatura máxima en el día previo (término autorregresivo) para modelar la persistencia de las temperaturas.
2. Modelo 1.0: Añadimos al Modelo 0.0 ajustado las funciones armónicas para modelar la periodicidad. Se incluye el armónico de periodo 1 año y se añade cada vez un nuevo armónico de periodo más corto ( $1/2, 1/3, \dots$ ). Comparamos cada paso con el anterior con la ayuda de un test ANOVA, hasta que este indique que el nuevo modelo ya no mejora al anterior y, entonces, nos quedamos con el número de armónicos de dicho modelo anterior ( $N$  o  $N^\sigma$  armónicos).
3. Modelo 1.1: Añadimos al ajuste 1.0 las interacciones entre la persistencia y los armónicos (productos de armónico por persistencia), que pueden indicar una persistencia distinta según la época del año. Estas interacciones se añaden una a una en el mismo orden que los armónicos y comparando cada modelo con el anterior hasta que el ajuste no mejore o hasta que hayamos añadido interacciones con todos los armónicos del modelo 1.0 ( $N_\rho$  o  $N_\rho^\sigma$  armónicos con interacciones).
4. Modelo 2.0: Añadimos al Modelo 1.1 ajustado la covariable que expresa el año, para modelizar la evolución anual de las temperaturas, lo que nos permitirá percibir los efectos directos del calentamiento global.
5. Modelo 2.1: Añadimos al ajuste 2.0 las interacciones entre el año y los armónicos de la misma manera que las interacciones con la persistencia en el modelo 1.1 ( $N_\alpha$  o  $N_\alpha^\sigma$  interacciones). Estos coeficientes representarían un efecto distinto del calentamiento según la época del año.

Una vez contruidos los 5 modelos lineales para un lugar concreto se compararán empleando tests ANOVA para elegir el que mejor represente los datos. Estas comparaciones se realizarán en orden de construcción de los modelos, empezando por comparar el 0.0 con el 1.0 y quedándonos cada vez con la mejor de las propuestas para después compararla con el siguiente ajuste. La estructura del predictor lineal de este "mejor modelo" será la que se utilice para modelizar la media de la distribución de probabilidad de temperaturas en la inferencia bayesiana.

Para explorar los componentes que se incluirán en el submodelo de varianza, se van a ajustar modelos lineales a los residuos del modelo lineal seleccionado para la media. Tomamos el logaritmo del cuadrado de los residuos de los datos con respecto del denominado "mejor modelo" los utilizamos como nuevos datos para modelizar el logaritmo de la varianza de la distribución de temperaturas. Se emplea el mismo proceso que se ha utilizado para modelizar la media. De ahí, obtendremos análogamente un "mejor modelo del logaritmo de la varianza", de modo que la estructura de su predictor lineal se utilizará en la inferencia bayesiana para modelizar la varianza de la distribución de probabilidad de temperaturas.

De esta manera, cada uno de los 18 lugares estudiados tendrá un modelo con una composición de predictores lineales que puede ser completamente distinta.

### 3.3.1. Resultados del procedimiento de selección del modelo

En esta sección se resumen las comunalidades identificadas en la selección de los modelos de medias y de varianzas. La información extensa de los resultados en cada observatorio se muestra en el Anexo A. La Figura 3.5 presenta la posición geográfica de los 18 observatorios utilizados, señalando, en el gráfico de la izquierda, el submodelo de media que ha sido seleccionado en cada posición y, en el gráfico de la derecha, el correspondiente submodelo de varianza tomado en cada lugar.

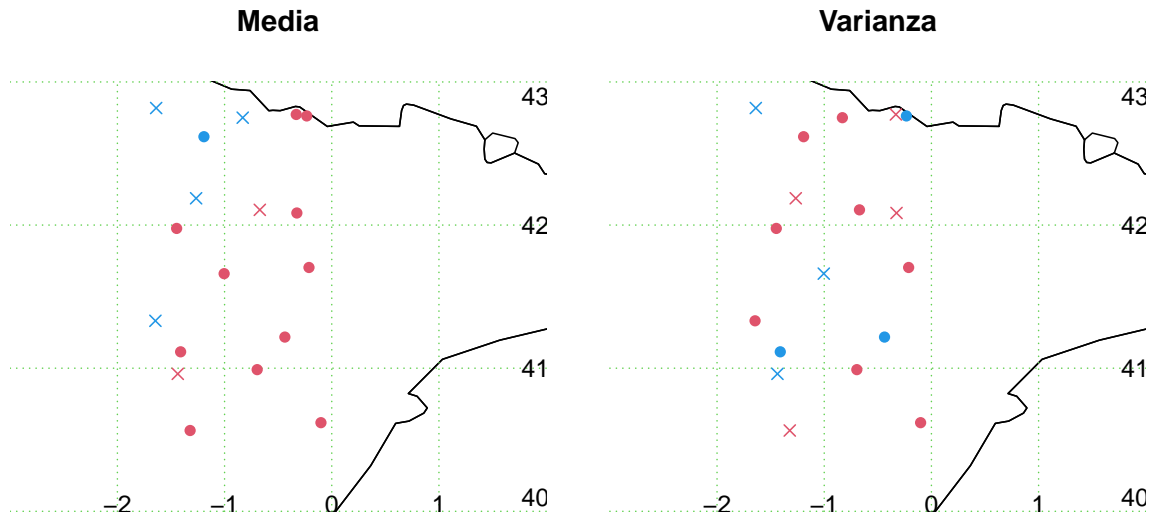


Figura 3.5: Mapas indicando los submodelos seleccionados en cada lugar. Se utilizan símbolos y colores para representarlos; 1.0 (punto azul), 1.1 (cruz azul), 2.0 (punto rojo), 2.1 (cruz roja).

#### Submodelo de medias

El grado de ajuste de estos submodelos es elevado, puesto que su  $R^2$  se sitúa por encima de 0,83 en todos los lugares. En el Anexo A, la Tabla A3 muestra las medidas de resumen de los modelos seleccionados. Se encuentra que en varios observatorios se selecciona el modelo 1.0, por ejemplo en Yesa. Por lo tanto, en este submodelo son comunes a todos los observatorios los términos que expresan estacionalidad (armónicos) y la dependencia del día previo ( $Y_{t,\ell-1}$ ). El número de armónicos seleccionados es al menos 2 en todos los casos.

Observando el panel izquierdo de la Figura 3.5, vemos que los modelos de la media seleccionados indican un patrón espacial respecto a la evidencia de tendencia, con dos zonas bien diferenciadas. Las localidades de la zona noroeste no presentan tendencia, efecto que sí tienen los lugares situados al Este y Sur. Se identifican dos observatorios (Tornos y La Sotonera, marcados con cruces rojas) que representan una transición entre ambas zonas, siendo los únicos lugares con tendencia no homogénea a lo largo del año. Estas tendencias son claramente positivas durante primavera y verano.

#### Submodelo de varianzas

Los únicos términos comunes a todos los ajustes seleccionados para el submodelo de varianzas son el primer armónico y el término autorregresivo. En este caso no hay zonas bien diferenciadas en el mapa del modelo de la varianza, Figura 3.5 derecha. A lo sumo se puede ver una reducida área que abarca Zaragoza y algunas poblaciones al sur que se distinguen de las demás por no presentar términos de tendencia anual.

El grado de ajuste es reducido porque los valores de  $R^2$  de los submodelos de varianza no superan 0,05 e indican que deberá analizarse con precaución su significación en el posterior ajuste del modelo bayesiano completo. Por otra parte, este es solo un primer intento de modelizar la varianza junto a la media, en trabajos futuros el ajuste de la varianza necesitará un modelo con mayor información. A pesar de ello, los modelos seleccionados contienen predictores lineales que corresponden al modelo 1.0 en 3 lugares, al 1.1 en 3 lugares, al 2.0 en 8 lugares y al 2.1 en 4 lugares, es decir, se encuentran indicios de que la varianza no es homogénea y cambia estacionalmente o dependiendo de la situación previa.

La presencia de patrones espaciales y comunales en ambos submodelos sugiere que sería útil utilizar un modelo espacial, al igual que hicieron Castillo-Mateo y colaboradores. Esta idea se propondrá para futuros trabajos.

### 3.4. Resultados del ajuste de los modelos bayesianos

En esta sección se resumen los ajustes obtenidos mediante *RStan* en los 18 observatorios, en concreto se sintetizan los resultados relativos al grado de ajuste obtenido, así como la composición de los predictores lineales, junto con herramientas que facilitan su interpretación. Los resultados completos de todos los ajustes se incluyen en el Anexo B.

#### 3.4.1. Grado de ajuste y crítica de los modelos

En primer lugar, se observa que en todos los parámetros de todos los modelos el valor de  $\hat{R}$  se encuentra entre 0,999 y 1,003. Según Gelman debería ser lo más cercano posible a 1 y son aceptables los valores de  $\hat{R}$  por debajo de 1,1.

En cuanto al tamaño efectivo de muestra, se requeriría que estuviese en todos casos por encima de 100 por lo menos. Los resultados muestran que, para casi todos los parámetros, este valor está por encima de 800, e incluso entre los 7 casos donde no se superan los 800, el mínimo tamaño efectivo de muestra es 320,3.

Se ha realizado un análisis de residuos del modelo bayesiano. Para ello, dado que se modelan la media y la varianza, las herramientas usuales de *RStan* no son útiles. Por tanto, se ha definido el residuo  $e_{t\ell}$  considerando una estandarización de la respuesta a partir de las medias a posteriori del valor medio  $\tilde{\mu}_{t\ell}$  y de la desviación típica  $\tilde{\sigma}_{t\ell}$  condicionados por el vector de covariables  $\mathbf{X}_{t\ell}$  en cada día. En el Anexo B se muestran figuras para mostrar la distribución de estos residuos para algunas localidades.

$$e_{t\ell} = \frac{y_{t\ell} - \tilde{\mu}_{t\ell}}{\tilde{\sigma}_{t\ell}}.$$

#### 3.4.2. Estructura de los predictores lineales para los submodelos de media y de varianza

Debe señalarse que el procedimiento de preselección de los términos de los modelos ha identificado correctamente términos que se requieren en el ajuste. Se consideran no significativos aquellos términos cuyo intervalo de credibilidad al 95% (IC95) contenga al 0. En concreto, entre los 18 modelos ajustados, cada uno con un número de coeficientes entre 13 (Yesa) y 26 (Pallaruelo), encontramos solo 11 términos no significativos. Cabe destacar que, a pesar del escaso grado de explicación obtenido en la fase de preselección, los efectos incluidos en el predictor lineal de la varianza han sido significativos. Los coeficientes no significativos no corresponden al efecto principal de la tendencia o la persistencia, sino con una componente armónica, el término del seno o el del coseno resulta no significativo pero la otra función sí que es significativa.

La Tabla 3.1 muestra un resumen de las distribuciones a posteriori obtenidas para todos los parámetros del modelo bayesiano ajustado en Zaragoza, como ejemplo de los modelos estimados. Los únicos coeficientes no significativos, indicados en rojo, están asociados a un término armónico que tiene asociado un coeficiente en el otro término del armónico que sí es significativo. El predictor de la varianza requiere estacionalidad y un término autorregresivo.

Submodelo	Parametro	Media	IC inferior	IC superior
	$\beta_0$	-160	-200	-120
Media	$\beta_{1,s}$	-15	-17	-12
	$\beta_{1,c}$	-34	-37	-31
Modelo 2.0	$\beta_{2,s}$	6.8	5.9	7.6
	$\beta_{2,c}$	-1.2	-2	-0.32
$N = 3$	$\beta_{3,s}$	0.001	-0.53	0.54
	$\beta_{3,c}$	-1.2	-1.7	-0.67
$N_\alpha = 0$	$\alpha_0$	0.11	0.094	0.13
	$\rho_0$	0.68	0.67	0.69
$N_\rho = 1$	$\rho_{1,s}$	0.03	0.018	0.043
	$\rho_{1,c}$	0.0027	-0.011	0.016
	$\beta_0^\sigma$	6.1	6	6.2
Varianza	$\beta_{1,s}^\sigma$	0.13	-0.0039	0.28
	$\beta_{1,c}^\sigma$	0.81	0.67	0.94
Modelo 1.1	$\beta_{2,s}^\sigma$	-0.1	-0.15	-0.062
	$\beta_{2,c}^\sigma$	-0.05	-0.093	-0.0085
$N = 3$	$\beta_{3,s}^\sigma$	-0.015	-0.042	0.012
	$\beta_{3,c}^\sigma$	0.092	0.065	0.12
$N_\alpha = 0$	$\rho_0^\sigma$	0.0019	0.0015	0.0023
	$\rho_{1,s}^\sigma$	0.00025	-0.00043	0.00088
$N_\rho = 1$	$\rho_{1,c}^\sigma$	-0.0037	-0.0043	-0.0031

Cuadro 3.1: Tabla con el resumen de las distribuciones a posteriori de cada parámetro en el modelo de Zaragoza. Los coeficientes no significativos están indicados en rojo.

### Efecto de la tendencia

Los modelos ajustados permiten valorar la variabilidad espacial y temporal de los efectos de tendencia. La tendencia estimada en la media de  $Tmax$  se expresa en  $^{\circ}C/década$  y corresponde para una fecha  $\ell$  del año con la distribución a posteriori de  $\alpha_0 + \sum_{i=1}^{N_\alpha} (\alpha_{i,s}Si_\ell + \alpha_{i,c}Ci_\ell)$ . Se muestran, a modo de ejemplo, en la Figura 3.6, resultados en 4 fechas, una en cada una de las estaciones del año, para comparar la distribución a posteriori de la tendencia en el valor medio en Tornos, Zaragoza y Panticosa, situadas en un eje sur-norte y con muy distintas altitudes. Se puede ver que Tornos presenta estacionalidad en el efecto de la tendencia, teniendo valores positivos en primavera y verano y negativos en invierno. Por otro lado, Zaragoza y Panticosa mantienen la misma tendencia durante todo el año, ya que sus submodelos de la media carecen de estacionalidad en este efecto.

Siguiendo esta misma idea de mostrar la variabilidad espacio-temporal de la tendencia, los mapas de la Figura 3.7 representan las medias a posteriori de la tendencia en la media de  $Tmax$  ajustados en los 18 lugares para los mismos 4 días del año; los valores se obtienen mediante un procedimiento *LOESS*. El mapa representa con una escala de colores rojizos las zonas con un mayor tendencia positiva y amarillos las zonas en las que se obtiene la tendencia cercana a 0 o negativa; se representa con una línea negra la curva de nivel de tendencia nula. Al fijarnos en los patrones de los mapas, aparecen similitudes en las tendencias en primavera y verano, cuando solo la zona noroeste tiene tendencias reducidas. El patrón espacial de la tendencia en otoño e invierno es distinto, siendo más extensa el área con tendencias reducidas o nulas y situándose el mayor calentamiento en el sur y el este.

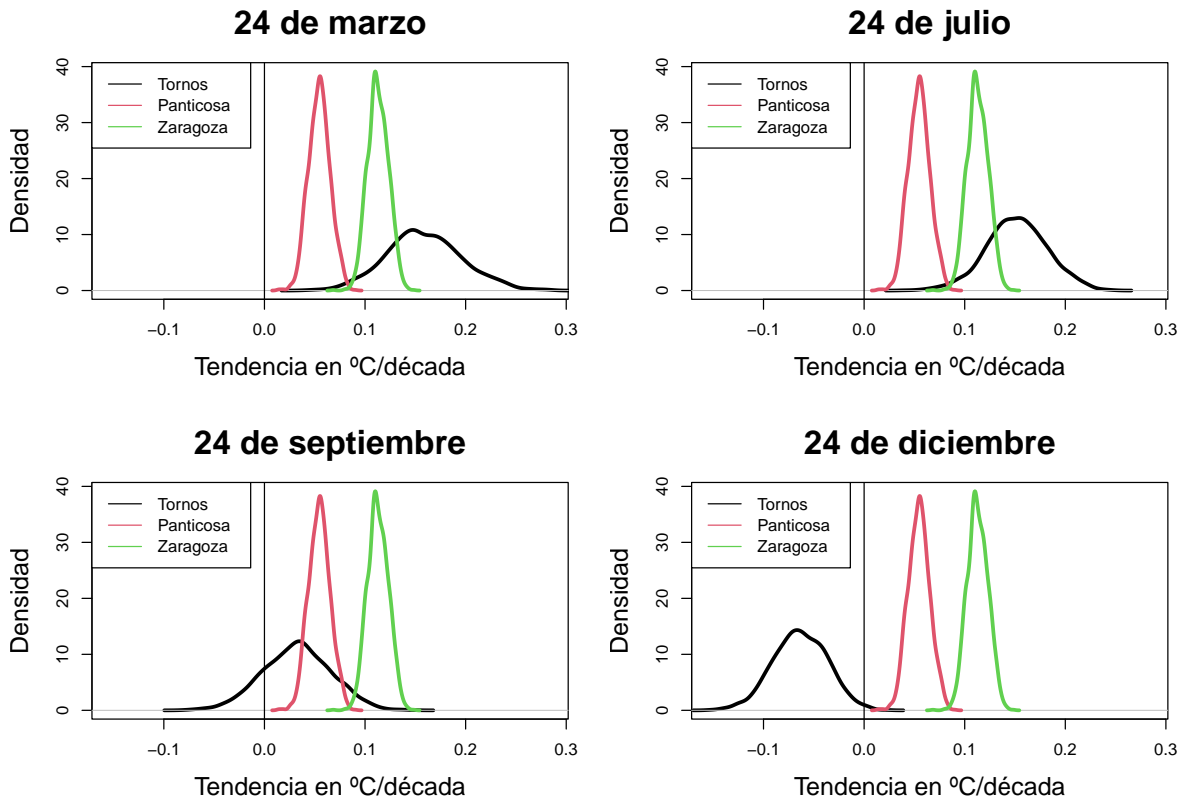


Figura 3.6: Densidades a posteriori obtenidas para la tendencia temporal en la media de la  $T_{max}$ , en Tornos, Zaragoza y Panticosa, en 4 días distintos del año.

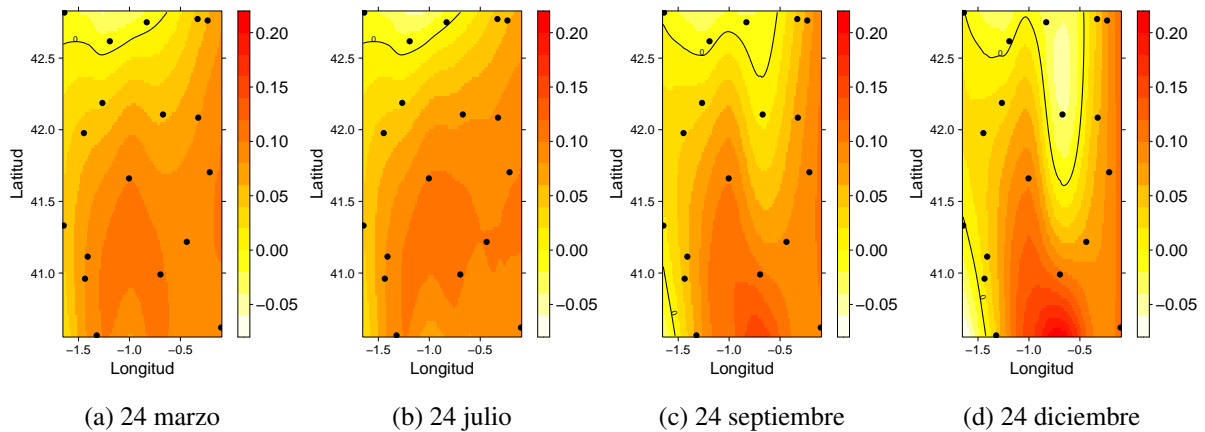


Figura 3.7: Mapa de la media a posteriori de las tendencias del valor medio, en  $^{\circ}\text{C}/\text{década}$ , para 4 días del año.

En cuanto a los efectos de tendencia de los submodelos de varianza, no tienen patrones tan claros como en los submodelos de media. Se ha encontrado que en algunos lugares, como por ejemplo en Santa Eulalia, este término proporcional al año puede variar con las estaciones. Esto se evidencia en la Figura 3.8, que muestra las medias a posteriori de los términos de tendencia en Pamplona, Ansó y Santa Eulalia en función del día del año; se representan los términos de los predictores de la varianza, divididos entre 2 para representar la tendencia del logaritmo de la desviación típica estimada.

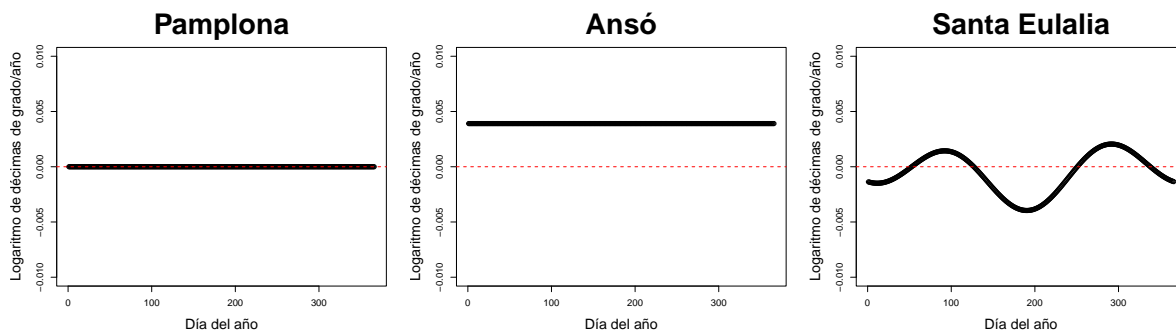


Figura 3.8: Valor medio a posteriori de la tendencia de la desviación típica, en escala logarítmica, frente al día del año.

**Efecto de la persistencia**

Los términos de la persistencia también presentan algunos resultados de interés. La Figura 3.9 presenta las medias a posteriori de los términos autorregresivos de los dos submodelos de Zaragoza. En la gráfica de la derecha se representa el predictor lineal entre 2 para expresar el efecto sobre el logaritmo de la desviación típica.

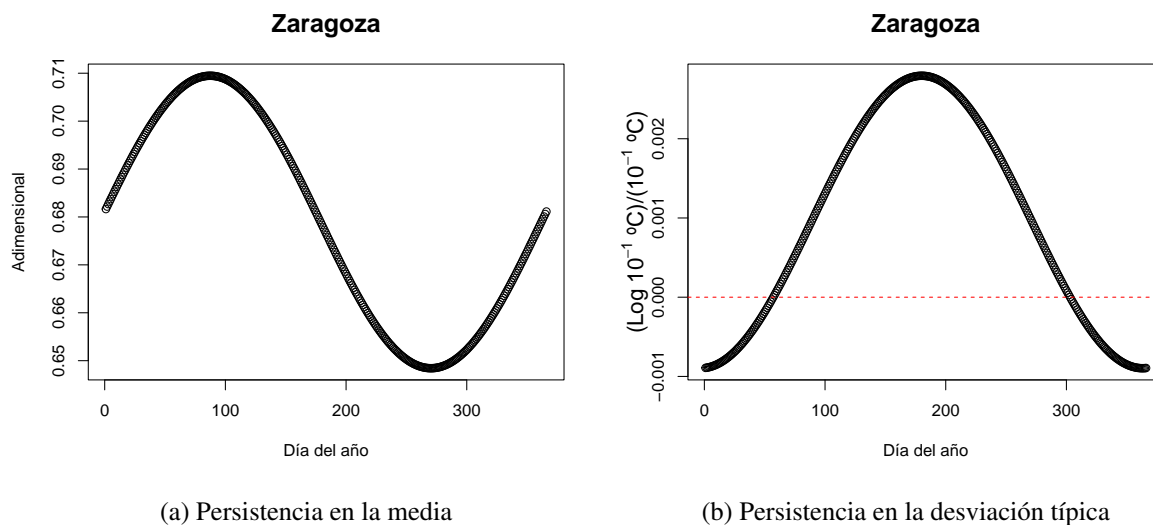


Figura 3.9: Valor medio a posteriori del término autorregresivo del submodelo de la media (izd.) y del submodelo de la desviación típica (dch.), en escala logarítmica, frente al día del año.

Ambos submodelos tienen un efecto autorregresivo estacional: en la media se presentan los valores mayores en primavera y los menores durante el otoño, indicando que en esta estación hay menos persistencia en la temperatura diaria. En la desviación típica aparece un máximo a mitad de año y valores incluso negativos en invierno. Estos valores negativos del término de la persistencia implicarán que a un día con temperaturas altas le seguirá uno con una menor varianza. Por el contrario, el resto del año, el término autorregresivo es positivo, es decir, hay más varianza cuanto mayor haya sido la temperatura del día anterior.

### 3.4.3. Estimación del cambio observado en el valor medio de la distribución marginal de la temperatura máxima diaria

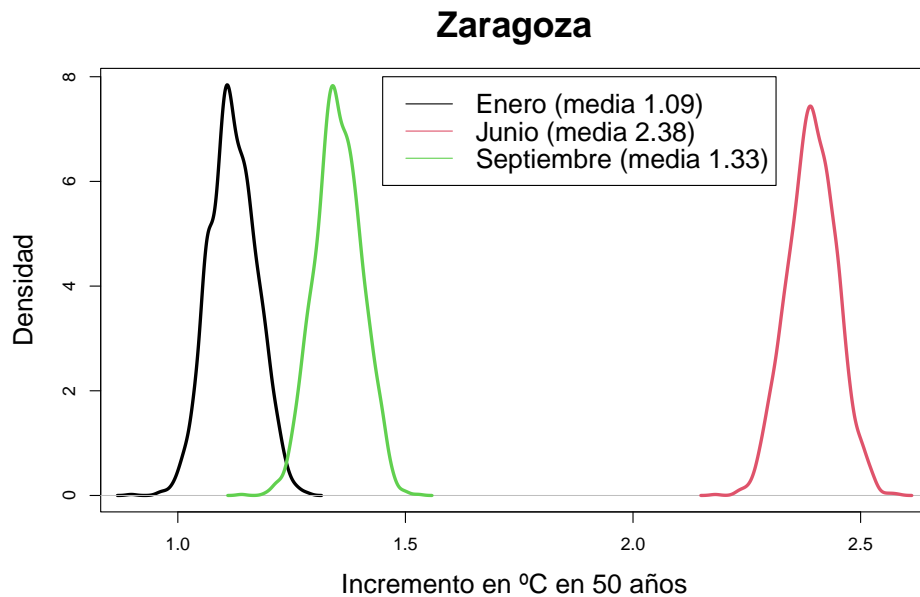


Figura 3.10: Densidad a posteriori de la diferencia del valor medio de  $T_{max}$  entre las décadas de 1955-1964 y 2005-2014, para los meses de enero, junio y septiembre, en Zaragoza.

El modelo autorregresivo que se ha estimado permite también estimar características de la distribución marginal de la temperatura máxima diaria o de características agregadas, como el valor medio o un percentil de una media mensual. Se va a aprovechar esta utilidad del marco bayesiano para obtener las densidades a posteriori de la diferencia entre las temperaturas medias mensuales estimadas en dos décadas separadas 50 años. La Figura 3.10 muestra esto para los meses de enero, junio y septiembre en Zaragoza. Se encuentra un incremento efectivo con dos conclusiones principales: i) Junio y septiembre han sufrido un calentamiento más rápido que enero, de donde se deduce que el periodo más cálido del año se está ampliando; ii) se observan distintos calentamiento efectivos para cada mes del año, pero debe advertirse que el término de la tendencia de Zaragoza no presenta estacionalidad, así que estos diferentes ritmos de calentamiento provienen de un desplazamiento de la distribución del término autorregresivo, a largo plazo se desplazan progresivamente las temperaturas.

### 3.4.4. Resultados de otros autores

Es conocido que el efecto del calentamiento global es distinto para cada clima y región geográfica, Fisher et al. ([12]). La heterogeneidad espacial que se encuentra en el entorno de Aragón también ha sido identificada por otros autores. Peña-Angulo et al. ([11]) estudiaron la evolución del valor medio estacional de las temperaturas máximas y mínimas en observatorios de la Península Ibérica entre 1916 y 2015. Encontraron una evolución espacialmente heterogénea y con diferencias para las distintas épocas del año, con una tendencia estimada desde 1956 de  $0,31^{\circ}\text{C}/\text{década}$  en primavera,  $0,27^{\circ}\text{C}/\text{década}$  en verano,  $0,13^{\circ}\text{C}/\text{década}$  en otoño y  $0,17^{\circ}\text{C}/\text{década}$  en invierno.

#### Comparación con el trabajo de Castillo-Mateo et al. (2022)

La aportación más relevante respecto al trabajo de Castillo-Mateo et al. (2022) es la modelización de la varianza. Este submodelo permite representar el comportamiento estacional de la variabilidad de

las temperaturas y, más relevante, considerar que la varianza del modelo condicional también incluye la dependencia respecto al día previo.

Como consecuencia de trabajar con todo el año aparecen términos significativos relativos a un comportamiento estacional tanto en la persistencia como en la tendencia. Estos términos no requieren ser considerados en Castillo-Mateo et al. porque se trabaja con los días del periodo cálido de mayo a septiembre.

Nuestro modelo se ajusta localmente, esto tiene como ventaja que la particularización del modelo seleccionado en cada lugar ayuda a evitar sobreparametrizaciones. Por otro lado, la consideración de un modelo espacial tiene dos ventajas fundamentales respecto al propuesto en este trabajo: permite modelar efectos ligados a variables geográficas como la altitud y permite considerar un efecto estacional común. Además, nuestro modelo no contiene términos aleatorios espaciales, que ayudarían a representar la variabilidad espacial de las tendencias (no asociada a la altitud).

Los mapas para identificar patrones espaciales se presentan en este trabajo mediante interpolación de los resultados locales. Sin embargo, no se puede hacer inferencia sobre posiciones geográficas donde no existe registro, lo que sí se puede obtener a partir del modelo espacial de Castillo-Mateo et al.

Los resultados obtenidos por Castillo-Mateo et al. muestran a coincidencias entre el patrón espacial que aparece en el término  $\rho_0$  y los valores obtenidos por nuestro modelo para el valor medio a posteriori del efecto de la persistencia durante el verano, que se sitúa alrededor del valor 0,7, con una reducción en la zona noroeste, sometida a mayor variabilidad atmosférica. Esto se puede comprobar en la Figura B7 del Anexo B.

La similitud es también evidente en el patrón espacial del calentamiento en verano: la media a posteriori de la tendencia está situada en mayores valores, próximos a  $0,35^\circ\text{C}/\text{década}$ , en el centro del Valle del Ebro, mientras que en el noroeste se encuentran, principalmente, distribuciones a posteriori centradas en 0.



## Capítulo 4

# Conclusiones y trabajo futuro

En este trabajo se ha revisado el algoritmo HMC para la estimación de modelos bayesianos. Este algoritmo de tipo MCMC y algunos métodos que lo mejoran son la base del software Stan, ampliamente utilizado por la comunidad científica.

Este marco de estimación se ha aplicado para obtener un modelo estadístico útil para analizar la tendencia en la distribución de la temperatura diaria. En primer lugar, se ha propuesto un modelo estadístico para una serie de temperatura máxima diaria, en la línea del propuesto por Castillo-Mateo et al. ([1]) pero con mayor versatilidad para que tanto el valor medio como la varianza representen las características observadas del fenómeno: estacionalidad, persistencia y posible tendencia al calentamiento. Esto proporciona una flexibilidad al modelo que permitiría incluir fácilmente los componentes que resultasen necesarios, por ejemplo, términos cuadráticos en la tendencia o en el efecto de la temperatura en el día previo. En segundo lugar, la estimación se plantea en el ámbito bayesiano usando el algoritmo HMC. Para esto se ha desarrollado una librería propia de funciones de R.

La aplicación se lleva a cabo en una base de datos con observatorios en el entorno de Aragón y medidas obtenidas desde mediados del siglo XX. Todos los modelos estimados presentan efecto de la persistencia tanto en el valor medio como en la varianza. También se ha obtenido una caracterización de las subregiones donde existe evidencia de tendencia, así como una estimación del cambio observado a escala diaria.

La presencia, en los submodelos de varianza, de términos significativos para representar la estacionalidad y la dependencia de la temperatura en el día previo, así como, en muchos casos, de términos significativos de tendencia, muestra que la modelización de la varianza en esta región ha sido, efectivamente, necesaria.

La estructura de la modelización propuesta presenta varias ventajas. Para empezar, los modelos obtenidos están en la escala diaria, la misma escala en la que se observan los registros, y por lo tanto pueden obtenerse estimaciones de estadísticos relativos a la distribución de la respuesta diaria.

El modelo es simulable y esto será una ventaja para valorar el grado de representación de los datos extremos. La modelización de las series diarias propone una distribución condicional de tipo gaussiano cuyos parámetros esperanza y varianza vienen determinados por los parámetros del modelo, para los que se dispone de su distribución a posteriori.

Por otro lado, se ha comprobado que el algoritmo HMC es costoso computacionalmente cuando se utiliza RStan. En la aplicación mostrada aquí esto es problemático debido a la complejidad de los modelos, porque hay una convergencia lenta en modelos que incluyen términos ligados a la tendencia en la varianza y porque se aplica sobre series de longitud considerable (alrededor de 20000 datos por cada lugar). Esto implica que implementar un proceso de selección del modelo bajo este paradigma de estimación bayesiana tendría un tiempo inasumible en un ordenador de prestaciones habituales.

En su lugar, se ha diseñado un procedimiento para la selección de los términos del modelo. La estrategia incluye una fase de análisis de modelos lineales no bayesianos aplicados dos veces: una sobre los datos, para la selección de variables que muestren evidencia de ser útiles para el submodelo de la media; y otra sobre sus residuos, para realizar la selección análoga en el submodelo de varianza.

En lo que respecta a los modelos ajustados, se han encontrado varios patrones espaciales en los términos de tendencia y persistencia que, en el periodo cálido del año, coinciden con los resultados obtenidos por Castillo-Mateo et al. También se ha observado un aumento neto de las temperaturas distinto para cada mes del año que parece indicar que el periodo cálido del año se está volviendo cada vez más largo. Por último, este resultado también parece insinuar que el efecto a largo plazo del calentamiento no se refleja únicamente en los términos de la tendencia, sino que también los términos de persistencia podrían estar representando una parte de este comportamiento.

#### 4.1. Líneas de trabajo abiertas

A partir de este TFG surgen numerosas líneas de trabajo. A corto plazo, se desarrollará un modelo espacio-temporal. Los patrones espaciales en algunos componentes identificados en este trabajo han indicado que podría ser útil pasar de los modelos locales empleados a un único modelo espacial, añadiendo las covariables espaciales que se consideren necesarias y un campo gaussiano que exprese, en cada día  $(t, l)$ , la dependencia entre los términos  $\varepsilon_{tl}(s)$ . Para ello se estudiarán las posibilidades de modelos espacio-temporales planteados por Mínguez et al. ([13]), implementados en la librería *pspatreg*, y por el profesor Sahu y colaboradores ([14]), que han desarrollado la librería *bmstdr*.

En paralelo, será necesario mejorar la estimación bayesiana para reducir el tiempo de computación y, por lo tanto, poder abordar modelos más complejos o poder realizar la fase de preselección en el ámbito bayesiano. Una opción es realizar operaciones empleando C++ en algunas o todas las fases del algoritmo HMC. Otra opción que se ensayará es completar el algoritmo HMC con una fase de muestreo Gibbs, lo que requiere obtener analíticamente las distribuciones condicionales de los parámetros de la varianza, puesto que las del submodelo de medias fueron desarrolladas por Castillo-Mateo et al. [1].

Otro objetivo a corto plazo es la definición de metodología para generar herramientas, basadas en unos residuos adecuados, para permitir la crítica de aspectos concretos de estos modelos con predictores relativamente complejos. En particular, deberían ser útiles para valorar si se requieren algunos términos que podrían ayudar al ajuste tanto del submodelo de medias como del de varianzas, por ejemplo, una componente cuadrática para la persistencia o para la tendencia. En especial, el submodelo de la varianza necesita pulirse, ya que esta ha sido una primera aproximación a modelizar una varianza no constante a lo largo del tiempo. Las herramientas de crítica permitirán un estudio en mayor profundidad de la variabilidad observada en  $T_{max}$ , donde se pueden encontrar otros términos útiles.

A más largo plazo, también se propone incluir en los modelos covariables atmosféricas, bien representadas en los modelos climáticos, como, por ejemplo, las alturas geopotenciales.

Un objetivo ambicioso será aplicar una modelización conjunta a las temperatura máximas y mínimas diarias para obtener una representación más completa de la evolución climática.

Finalmente, una vez tengamos un modelo que represente correctamente la evolución de las temperaturas, podremos probarlo en otras situaciones, como, por ejemplo, escenarios globales definidos por el Panel Intergubernamental del Cambio Climático (IPCC), y también aplicarlo para intentar realizar proyecciones sobre las temperaturas futuras.

# Bibliografía

- [1] Jorge Castillo-Mateo, Miguel Lafuente, Jesús Asín, Ana C Cebrián, Alan E Gelfand, and Jesús Abaurrea. Spatial modeling of day-within-year temperature time series: an examination of daily maximum temperatures in Aragón, Spain. *Journal of Agricultural, Biological and Environmental Statistics*, 27:487–505, 2022.
- [2] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.
- [3] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [4] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [5] Martyn Plummer et al. Jags: A program for analysis of bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–10. Vienna, Austria., 2003.
- [6] Nikolaus Umlauf, Nadja Klein, and Achim Zeileis. Bamlss: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, 27(3):612–627, 2018.
- [7] Andrew Gelman, Daniel Lee, and Jiqiang Guo. Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543, 2015.
- [8] Chelsea Muth, Zita Oravecz, and Jonah Gabry. User-friendly bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quantitative Methods for Psychology*, 14(2):99–119, 2018.
- [9] William Kleiber, Richard W Katz, and Balaji Rajagopalan. Daily minimum and maximum temperature simulation over complex terrain. *The Annals of Applied Statistics*, 7(1):588–612, 2013.
- [10] Andrew Verdin, Balaji Rajagopalan, William Kleiber, Guillermo Podestá, and Federico Bert. BayGEN: A Bayesian space-time stochastic weather generator. *Water Resources Research*, 55(4):2900–2915, 2019.
- [11] Dhais Peña-Angulo, José Carlos Gonzalez-Hidalgo, Leire Sandonís, Santiago Beguería, Miquel Tomas-Burguera, Joan Albert López-Bustins, Marc Lemus-Canovas, and Javier Martin-Vide. Seasonal temperature trends on the spanish mainland: A secular study (1916–2015). *International Journal of Climatology*, 41(5):3071–3084, 2021.
- [12] EM Fischer, Sebastian Sippel, and Reto Knutti. Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11(8):689–695, 2021.
- [13] Román Mínguez, Roberto Basile, and María Durbán. An alternative semiparametric model for spatial panel data. *Statistical Methods & Applications*, 29(4):669–708, 2020.
- [14] Sujit K Sahu. *Bayesian modeling of spatio-temporal data with R*. Chapman and Hall/CRC, 2022.

# Anexo A: Análisis exploratorio

En este Anexo se incluyen figuras del análisis exploratorio en los 18 observatorios.

Se muestran los resultados de la fase de preselección de los modelos. En particular, se muestran resúmenes de los modelos lineales que se han ajustado y los que se han seleccionado para la posterior estimación bayesiana.

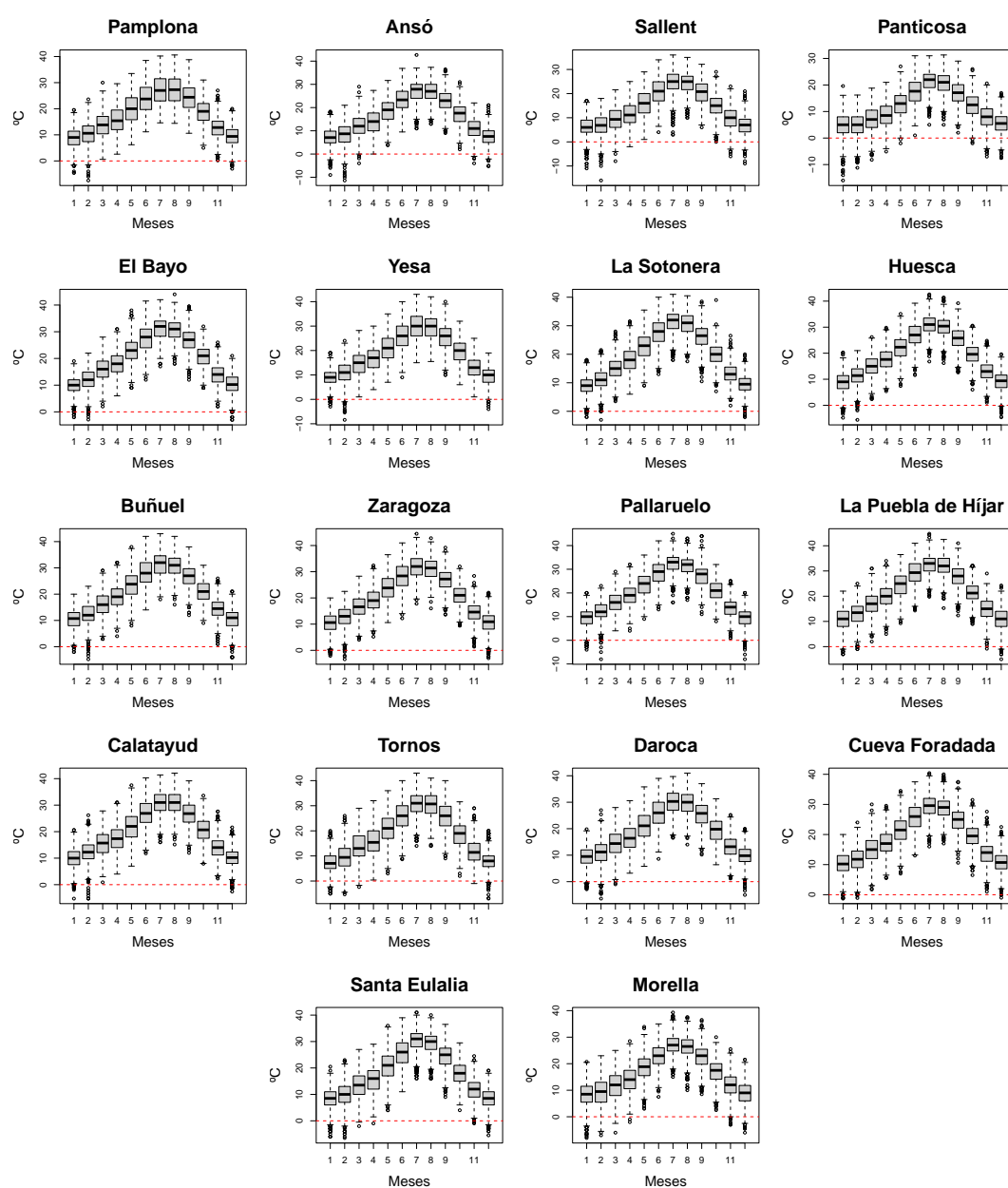


Figura A1: Diagrama de cajas por meses de  $T_{max}$  en cada lugar.

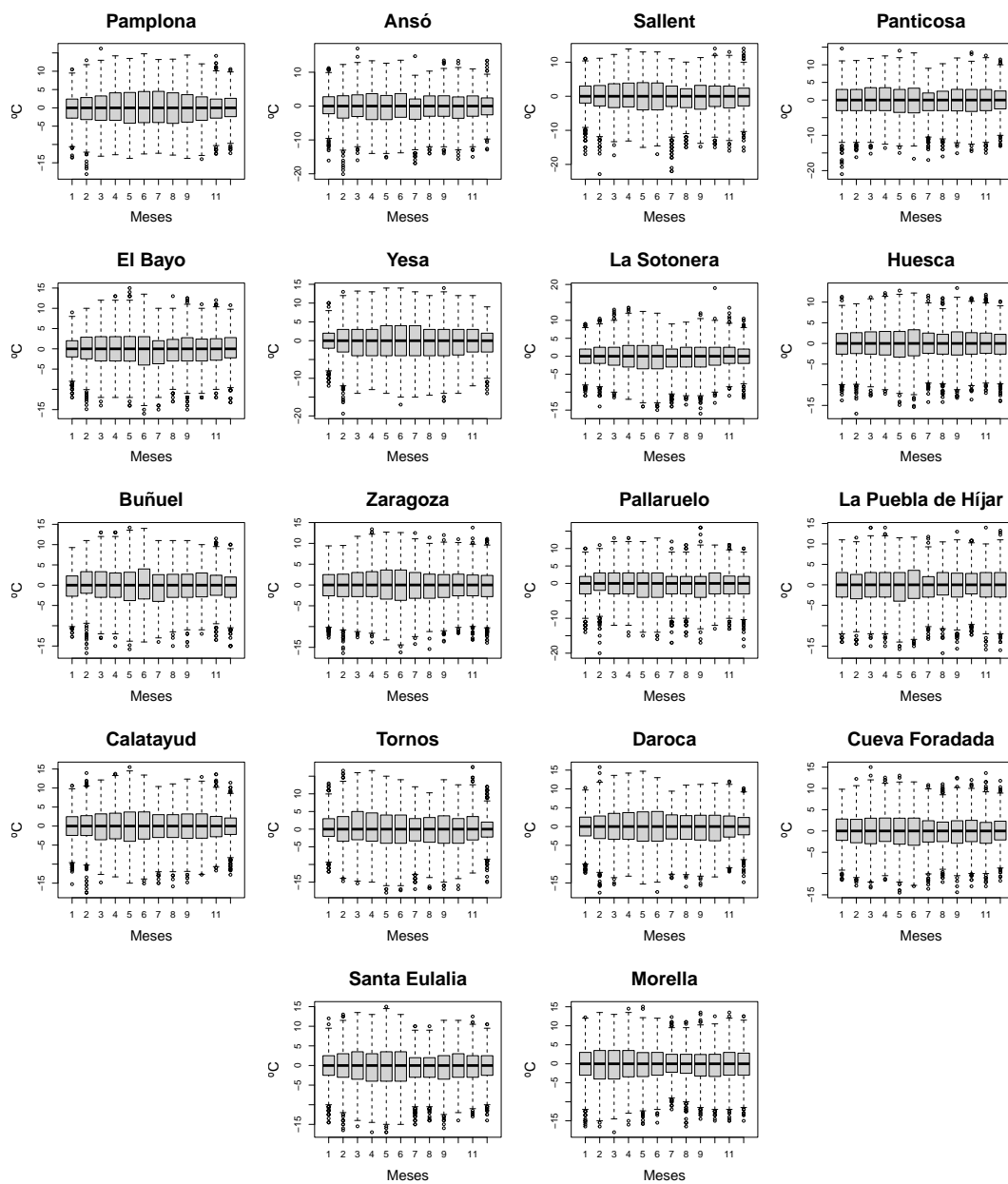


Figura A2: Diagrama de cajas por meses de la anomalía en cada lugar.

<i>MEDIA</i>		0.0	1.0	1.1	2.0	2.1
Pamplona	$R^2$	0.80	0.83	0.83	0.83	0.83
	acf de residuos	-0.05	0.043	0.044	0.044	0.044
Buñuel	$R^2$	0.86	0.88	0.88	0.88	0.88
	acf de residuos	-0.081	0.027	0.027	0.028	0.028
El Bayo	$R^2$	0.88	0.90	0.90	0.90	0.90
	acf de residuos	-0.082	0.025	0.025	0.025	0.025
Morella	$R^2$	0.85	0.86	0.86	0.86	0.86
	acf de residuos	-0.076	0.011	0.011	0.014	0.014
Huesca	$R^2$	0.89	0.91	0.91	0.91	0.91
	acf de residuos	-0.098	0.011	0.013	0.017	0.017
Tornos	$R^2$	0.86	0.87	0.87	0.87	0.87
	acf de residuos	-0.057	0.036	0.036	0.037	0.037
Santa Eulalia	$R^2$	0.88	0.89	0.89	0.89	0.89
	acf de residuos	-0.064	0.031	0.031	0.032	0.032
Calatayud	$R^2$	0.87	0.88	0.88	0.88	0.88
	acf de residuos	-0.074	0.022	0.023	0.023	0.023
Panticosa	$R^2$	0.83	0.84	0.84	0.84	0.84
	acf de residuos	-0.036	0.039	0.039	0.04	0.04
La Puebla de Híjar	$R^2$	0.87	0.89	0.89	0.89	0.89
	acf de residuos	-0.103	0.012	0.012	0.014	0.014
Ansó	$R^2$	0.87	0.88	0.88	0.88	0.88
	acf de residuos	-0.006	0.07	0.07	0.07	0.07
Daroca	$R^2$	0.86	0.87	0.87	0.87	0.87
	acf de residuos	-0.033	0.056	0.056	0.06	0.06
Zaragoza	$R^2$	0.87	0.89	0.89	0.89	0.89
	acf de residuos	-0.081	0.03	0.03	0.033	0.034
La Sotonera	$R^2$	0.9	0.91	0.91	0.91	0.91
	acf de residuos	-0.093	0.017	0.017	0.017	0.018
Pallaruelo	$R^2$	0.89	0.90	0.90	0.90	0.90
	acf de residuos	-0.106	0.008	0.009	0.01	0.01
Embalse de Cueva Foradada	$R^2$	0.87	0.89	0.89	0.89	0.89
	acf de residuos	-0.071	0.026	0.027	0.03	0.03
Sallent de Gállego	$R^2$	0.85	0.86	0.86	0.86	0.86
	acf de residuos	-0.055	0.026	0.026	0.026	0.027
Yesa	$R^2$	0.84	0.86	0.86	0.86	0.86
	acf de residuos	-0.082	0.024	0.024	0.024	0.024

Cuadro A1: Valores de  $R^2$  y del primer término de la función de autocorrelación de los residuos para todos los modelos lineales de la preselección de modelos de la media.

<b>VARIANZA</b>		0.0	1.0	1.1	2.0	2.1
Pamplona	$R^2$	0.04	0.05	0.05	0.05	0.05
	acf de residuos	0.03	0.018	0.017	0.017	0.017
Buñuel	$R^2$	0.006	0.01	0.01	0.02	0.02
	acf de residuos	0.046	0.04	0.039	0.032	0.032
El Bayo	$R^2$	0.005	0.01	0.01	0.02	0.02
	acf de residuos	0.066	0.061	0.06	0.053	0.053
Morella	$R^2$	0.0008	0.006	0.007	0.008	0.008
	acf de residuos	0.051	0.048	0.048	0.047	0.047
Huesca	$R^2$	0.0001	0.002	0.003	0.003	0.004
	acf de residuos	0.041	0.04	0.039	0.039	0.039
Tornos	$R^2$	0.005	0.02	0.02	0.02	0.02
	acf de residuos	0.045	0.035	0.034	0.033	0.033
Santa Eulalia	$R^2$	0.0004	0.008	0.009	0.009	0.01
	acf de residuos	0.064	0.057	0.056	0.056	0.054
Calatayud	$R^2$	0.006	0.01	0.01	0.01	0.01
	acf de residuos	0.047	0.043	0.042	0.041	0.041
Panticosa	$R^2$	0.0002	0.001	0.001	0.002	0.002
	acf de residuos	0.063	0.062	0.062	0.061	0.061
La Puebla de Híjar	$R^2$	0.00001	0.007	0.008	0.008	0.008
	acf de residuos	0.059	0.054	0.055	0.054	0.054
Ansó	$R^2$	0.002	0.01	0.01	0.02	0.02
	acf de residuos	0.058	0.048	0.047	0.044	0.044
Daroca	$R^2$	0.008	0.01	0.01	0.01	0.01
	acf de residuos	0.028	0.023	0.023	0.023	0.023
Zaragoza	$R^2$	0.005	0.01	0.01	0.01	0.01
	acf de residuos	0.055	0.048	0.049	0.049	0.049
La Sotonera	$R^2$	0.009	0.02	0.02	0.02	0.02
	acf de residuos	0.053	0.043	0.043	0.042	0.042
Pallaruelo	$R^2$	0.000004	0.006	0.007	0.007	0.008
	acf de residuos	0.059	0.054	0.052	0.052	0.052
Embalse de Cueva Foradada	$R^2$	0.003	0.01	0.01	0.01	0.01
	acf de residuos	0.076	0.071	0.072	0.069	0.068
Sallent de Gállego	$R^2$	0.001	0.007	0.008	0.01	0.01
	acf de residuos	0.065	0.059	0.057	0.054	0.053
Yesa	$R^2$	0.01	0.02	0.02	0.02	0.02
	acf de residuos	0.051	0.045	0.045	0.044	0.044

Cuadro A2: Valores de  $R^2$  y del primer término de la función de autocorrelación de los residuos para todos los modelos lineales de la preselección de modelos de la varianza.

<b>MEDIA</b>	Nº de armónico ( $N$ )	Nº de interacciones de persistencia ( $N_{\rho}$ )	Nº de interacciones de tendencia ( $N_{\alpha}$ )	Mejor modelo
Pamplona	2	1	1	1.1
Buñuel	3	1	1	2.0
El Bayo	3	1	1	1.1
Morella	2	1	1	2.0
Huesca	3	1	1	2.0
Tornos	2	1	1	2.1
Santa Eulalia	3	1	1	2.0
Calatayud	3	1	1	1.1
Panticosa	2	1	1	2.0
La Puebla de Hajar	3	1	1	2.0
Ansó	2	1	1	1.1
Daroca	2	1	1	2.0
Zaragoza	3	1	1	2.0
La Sotonera	3	1	1	2.1
Pallaruelo	3	1	1	2.0
Embalse de Cueva Foradada	2	1	1	2.0
Sallent de Gállego	2	1	1	2.0
Yesa	2	1	1	1.0

Cuadro A3: Número de armónicos utilizados para cada modelo lineal de la preselección de submodelos de la media y los submodelos elegidos para cada lugar.

<b>VARIANZA</b>	Nº de armónico ( $N^{\sigma}$ )	Nº de interacciones de persistencia ( $N_{\rho}^{\sigma}$ )	Nº de interacciones de tendencia ( $N_{\alpha}^{\sigma}$ )	Mejor modelo
Pamplona	3	1	1	1.1
Buñuel	2	2	1	2.0
El Bayo	1	1	1	2.1
Morella	2	1	1	2.0
Huesca	1	1	1	2.1
Tornos	2	1	1	1.1
Santa Eulalia	2	1	2	2.1
Calatayud	2	2	1	2.0
Panticosa	2	1	1	1.0
La Puebla de Hajar	3	1	1	1.0
Ansó	2	1	1	2.0
Daroca	2	1	1	1.0
Zaragoza	3	1	1	1.1
La Sotonera	1	1	1	2.0
Pallaruelo	4	2	1	2.0
Embalse de Cueva Foradada	2	1	1	2.0
Sallent de Gállego	2	1	1	2.1
Yesa	1	1	1	2.0

Cuadro A4: Número de armónicos utilizados para cada modelo lineal de la preselección de submodelos de la varianza y los submodelos elegidos para cada lugar.



# Anexo B. Resultados de la modelización bayesiana

En este Anexo se incluyen al comienzo algunos ejemplos de histogramas de residuos de los modelos bayesianos ajustados.

Después, se presentan las tablas que muestran el modelo bayesiano ajustado en cada observatorio. Para cada parámetro se resume su distribución a posteriori con la media, desviación típica e intervalo de credibilidad al 95 %, también se incorporan el tamaño efectivo de muestra y  $\hat{R}$ .

Finalmente, se presentan figuras que resumen el efecto en el predictor lineal de la persistencia y de la tendencia, tanto para el submodelo de medias como para el de varianza. Los gráficos aparecen en disposición pseudogeográfica para ayudar a la interpretación espacial de la estructura de los predictores lineales.

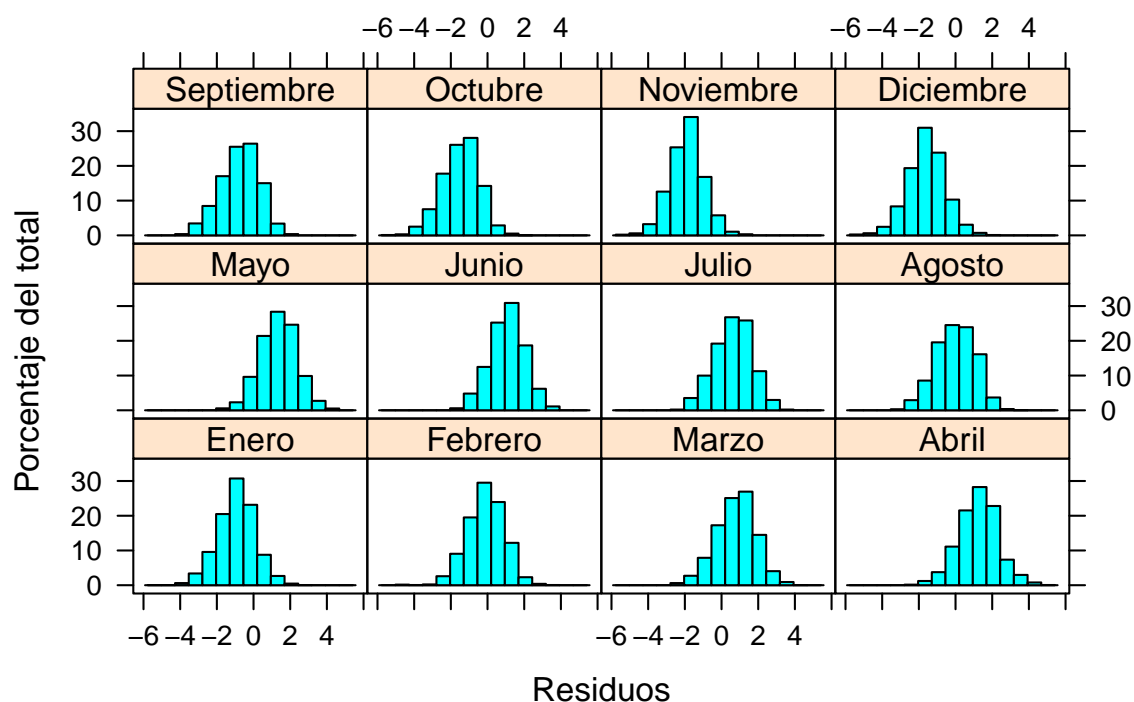


Figura B3: Histogramas de los residuos del modelo Pamplona agrupados por meses.

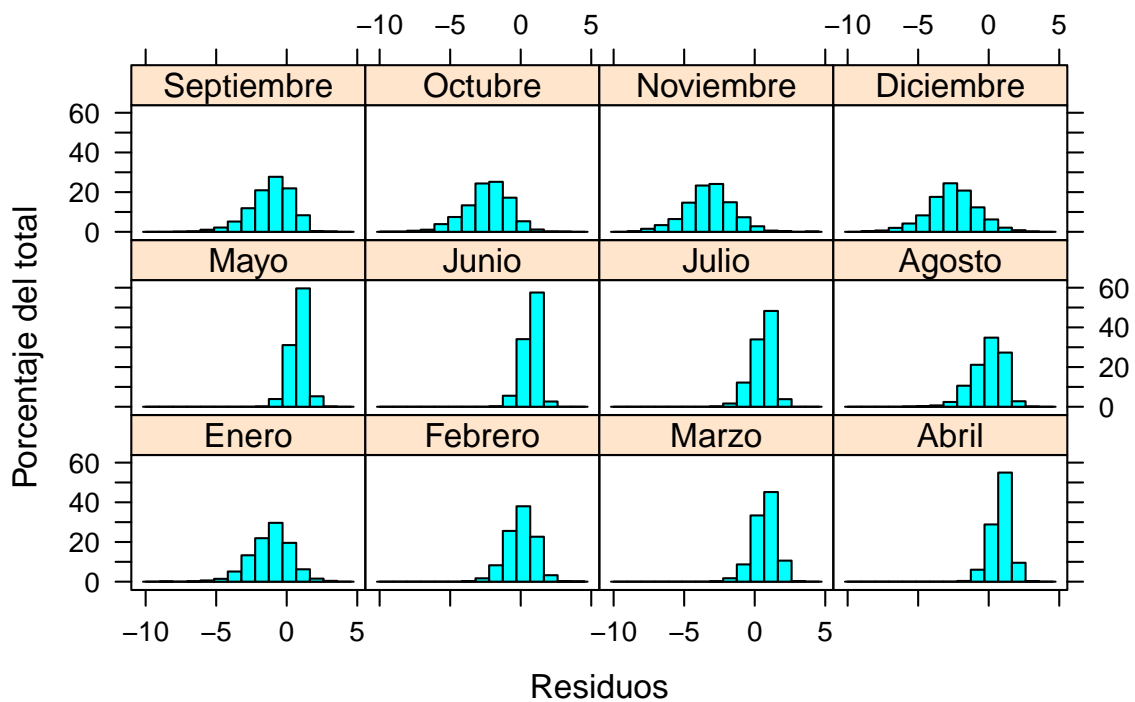


Figura B4: Histogramas de los residuos del modelo Zaragoza agrupados por meses.

<b><i>Pamplona</i></b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	55.93	0.89	[54.16,57.75]	1000	1
media y.lag	0.6994	0.00467	[0.69,0.7086]	1500	1
media sin1	-14.27	1.2	[-16.61,-11.95]	880	1
media cos1	-35.65	1.26	[-38.16,-33.22]	1100	1
media sin2	6.465	0.443	[5.635,7.374]	1000	1
media cos2	0.829	0.427	[0.008114,1.667]	1300	1
media y.lag:sin1	0.02876	0.00654	[0.01625,0.04125]	960	1
media y.lag:cos1	0.04969	0.00671	[0.03674,0.06328]	1200	1
varianza (Intercept)	6.493	0.0376	[6.422,6.57]	1700	1
varianza y.lag	0.00227	0.000202	[0.001859,0.002658]	2000	1
varianza sin1	-0.03773	0.0548	[-0.15,0.07211]	1400	1
varianza cos1	-0.008251	0.0537	[-0.1143,0.09558]	1600	1
varianza sin2	0.07105	0.0188	[0.03336,0.1079]	1500	1
varianza cos2	-0.04443	0.0191	[-0.08176,-0.006604]	1800	1
varianza sin3	-0.04866	0.0133	[-0.07551,-0.02327]	2400	1
varianza cos3	0.0545	0.0132	[0.02787,0.08122]	2700	1
varianza y.lag:sin1	0.0009305	0.000289	[0.0003549,0.00151]	1500	1
varianza y.lag:cos1	-0.001771	0.000303	[-0.002358,-0.001188]	1600	1

Cuadro B5: Tabla con el resumen de las distribuciones a posteriori de Pamplona y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Buñuel</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-7.198	21.1	[-49.6,34.06]	1300	1
media y.lag	0.6975	0.00464	[0.6884,0.7066]	8400	1
media sin1	-14.32	1.36	[-16.94,-11.65]	3500	1
media cos1	-34.47	1.45	[-37.22,-31.66]	3200	1
media sin2	6.341	0.422	[5.505,7.155]	3500	1
media cos2	-0.3798	0.446	[-1.256,0.5046]	3800	1
media sin3	0.05243	0.267	[-0.4621,0.5793]	9000	1
media cos3	-0.9107	0.269	[-1.438,-0.3848]	9300	1
media year	0.0353	0.0107	[0.01445,0.05669]	1300	1
media y.lag:sin1	0.03194	0.00652	[0.01885,0.04466]	3600	1
media y.lag:cos1	0.01716	0.00673	[0.004022,0.03032]	3300	1
varianza (Intercept)	-15.19	1.06	[-17.25,-13.13]	8200	1
varianza y.lag	0.001298	0.000228	[0.0008567,0.001747]	10000	1
varianza sin1	0.04963	0.0686	[-0.08312,0.1853]	6700	1
varianza cos1	0.3083	0.0671	[0.1762,0.4388]	8200	1
varianza sin2	-0.1099	0.0452	[-0.2001,-0.02444]	7300	1
varianza cos2	0.1789	0.0444	[0.09195,0.2678]	8900	1
varianza year	0.01086	0.000537	[0.009815,0.01191]	8200	1
varianza y.lag:sin1	0.0007434	0.000319	[0.0001147,0.001368]	7600	1
varianza y.lag:cos1	-0.002066	0.000315	[-0.002687,-0.001451]	9000	1
varianza y.lag:sin2	0.0002342	0.000198	[-0.0001562,0.0006246]	10000	1
varianza y.lag:cos2	-0.0009624	0.000194	[-0.001344,-0.0005824]	12000	1

Cuadro B6: Tabla con el resumen de las distribuciones a posteriori de Buñuel y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>El Bayo</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	58.48	1.01	[56.49,60.4]	2900	1
media y.lag	0.7139	0.00462	[0.7048,0.7228]	3800	1
media sin1	-12.43	1.3	[-14.99,-9.938]	2800	1
media cos1	-33.29	1.42	[-36.11,-30.51]	2500	1
media sin2	5.957	0.415	[5.16,6.779]	3200	1
media cos2	0.06393	0.44	[-0.7863,0.9286]	3300	1
media sin3	0.01615	0.254	[-0.487,0.5101]	7500	1
media cos3	-0.893	0.256	[-1.396,-0.3875]	7300	1
media y.lag:sin1	0.02098	0.00633	[0.008904,0.03337]	2800	1
media y.lag:cos1	0.0187	0.00663	[0.005701,0.03177]	2800	1
varianza (Intercept)	-14.72	1.12	[-16.91,-12.55]	7400	1
varianza y.lag	0.0006428	0.000222	[0.0002078,0.001074]	10000	1
varianza sin1	-4.78	1.54	[-7.854,-1.782]	5800	1
varianza cos1	-5.856	1.59	[-9.022,-2.727]	5900	1
varianza year	0.01067	0.000563	[0.009571,0.01178]	7400	1
varianza y.lag:sin1	0.001068	0.000202	[0.0006769,0.001467]	11000	1
varianza y.lag:cos1	-0.0008062	0.000193	[-0.001178,-0.0004268]	10000	1
varianza sin1:year	0.002361	0.000777	[0.0008508,0.003907]	5800	1
varianza cos1:year	0.002931	0.000801	[0.001352,0.004537]	6000	1

Cuadro B7: Tabla con el resumen de las distribuciones a posteriori de EL Bayo y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Morella</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-140.1	20.5	[-180.6,-99.71]	3300	1
media y.lag	0.7272	0.00447	[0.7186,0.7361]	7900	1
media sin1	-14.19	1.08	[-16.27,-12.1]	3300	1
media cos1	-27.86	1.13	[-30.06,-25.63]	3400	1
media sin2	5.983	0.415	[5.17,6.786]	3600	1
media cos2	1.614	0.402	[0.8367,2.397]	4300	1
media year	0.0942	0.0104	[0.07369,0.1145]	3300	1
media y.lag:sin1	0.03493	0.00633	[0.02254,0.04717]	3000	1
media y.lag:cos1	0.022	0.00618	[0.01004,0.03411]	3500	1
varianza (Intercept)	3.079	1.04	[1.065,5.151]	7700	1
varianza y.lag	0.001897	0.000211	[0.001482,0.002308]	8600	1
varianza sin1	0.1145	0.0511	[0.01465,0.214]	5400	1
varianza cos1	0.1824	0.0552	[0.07304,0.2888]	5400	1
varianza sin2	0.02457	0.0192	[-0.013,0.06222]	5400	1
varianza cos2	-0.0674	0.0191	[-0.1052,-0.03025]	6100	1
varianza year	0.001697	0.000527	[0.0006527,0.002715]	7600	1
varianza y.lag:sin1	0.000367	0.000288	[-0.0001914,0.0009258]	5600	1
varianza y.lag:cos1	0.0006021	0.000291	[3.055e-05,0.001183]	5700	1

Cuadro B8: Tabla con el resumen de las distribuciones a posteriori de Morella y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Huesca</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-160.3	19	[-197,-122.7]	3400	1
media y.lag	0.6976	0.00474	[0.6883,0.7068]	7200	1
media sin1	-10.6	1.3	[-13.14,-8.066]	3100	1
media cos1	-27.28	1.38	[-29.99,-24.58]	2800	1
media sin2	5.598	0.429	[4.771,6.437]	3300	1
media cos2	-1.975	0.446	[-2.849,-1.102]	3700	1
media sin3	-0.0224	0.251	[-0.5173,0.4653]	7600	1
media cos3	-1.427	0.251	[-1.92,-0.9298]	8200	1
media year	0.1096	0.00963	[0.09045,0.1284]	3300	1
media y.lag:sin1	0.01409	0.00652	[0.001473,0.02681]	3100	1
media y.lag:cos1	-0.02385	0.00672	[-0.03686,-0.01067]	3000	1
varianza (Intercept)	3.985	1.02	[2.015,5.966]	8100	1
varianza y.lag	-0.0001945	0.000238	[-0.0006604,0.0002703]	10000	1
varianza sin1	0.1968	1.43	[-2.619,2.989]	6400	1
varianza cos1	-3.711	1.45	[-6.524,-0.8652]	7300	1
varianza year	0.001286	0.000515	[0.0002891,0.00228]	8100	1
varianza y.lag:sin1	0.000691	0.000209	[0.0002831,0.001101]	10000	1
varianza y.lag:cos1	-0.001007	0.000193	[-0.001387,-0.0006252]	11000	1
varianza sin1:year	-0.0001145	0.000722	[-0.001529,0.001306]	6400	1
varianza cos1:year	0.001941	0.000731	[0.000499,0.003369]	7300	1

Cuadro B9: Tabla con el resumen de las distribuciones a posteriori de Huesca y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Tornos</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-55.34	23.9	[-102.7,-9.311]	1000	1
media y.lag	0.7145	0.00479	[0.7054,0.7237]	2500	1
media sin1	-102.4	34.3	[-169.2,-35.61]	940	1
media cos1	104.2	31.3	[40.34,166.2]	320	1
media sin2	7.062	0.482	[6.088,7.99]	2400	1
media cos2	1.526	0.51	[0.5243,2.517]	2500	1
media year	0.0551	0.0121	[0.03185,0.0792]	1000	1
media y.lag:sin1	0.01214	0.0061	[0.0001331,0.0235]	2400	1
media y.lag:cos1	0.03303	0.00639	[0.02051,0.04537]	2200	1
media sin1:year	0.04557	0.0173	[0.012,0.07895]	940	1
media cos1:year	-0.07133	0.0158	[-0.1025,-0.0388]	320	1
varianza (Intercept)	6.876	0.038	[6.801,6.953]	1800	1
varianza y.lag	0.001367	0.00019	[0.000992,0.001732]	2100	1
varianza sin1	-0.08221	0.0509	[-0.1849,0.0179]	1900	1
varianza cos1	-0.384	0.0522	[-0.4907,-0.2845]	1900	1
varianza sin2	0.08433	0.02	[0.04607,0.1227]	1900	1
varianza cos2	-0.1099	0.0205	[-0.1502,-0.07018]	2500	1
varianza y.lag:sin1	0.00146	0.000261	[0.0009438,0.001972]	1900	1
varianza y.lag:cos1	0.001438	0.000268	[0.0009301,0.001982]	2100	1

Cuadro B10: Tabla con el resumen de las distribuciones a posteriori de Tornos y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Santa Eulalia</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-43.28	20.8	[-83.92,-2.618]	3200	1
media y.lag	0.7261	0.00453	[0.7172,0.7351]	6300	1
media sin1	-11.51	1.18	[-13.83,-9.196]	2800	1
media cos1	-33.45	1.28	[-35.94,-30.96]	2400	1
media sin2	6.496	0.45	[5.599,7.397]	3100	1
media cos2	2.273	0.441	[1.398,3.15]	2800	1
media sin3	-0.998	0.278	[-1.544,-0.4517]	6400	1
media cos3	-0.8937	0.273	[-1.429,-0.3625]	6700	1
media year	0.04757	0.0105	[0.02724,0.06804]	3200	1
media y.lag:sin1	0.01961	0.00636	[0.007242,0.03194]	2800	1
media y.lag:cos1	0.02462	0.00635	[0.01209,0.03706]	2400	1
varianza (Intercept)	8.581	1.03	[6.568,10.59]	6700	1
varianza y.lag	0.001354	0.000216	[0.0009327,0.001775]	10000	1
varianza sin1	0.5957	1.47	[-2.293,3.473]	6100	1
varianza cos1	-5.017	1.47	[-7.872,-2.129]	5900	1
varianza sin2	2.411	1.47	[-0.4215,5.269]	6500	1
varianza cos2	8.214	1.48	[5.357,11.15]	6300	1
varianza year	-0.001062	0.00052	[-0.002081,-4.957e-05]	6700	1
varianza y.lag:sin1	0.0009729	0.000296	[0.0004095,0.00155]	10000	1
varianza y.lag:cos1	-0.0003379	0.000303	[-0.0009334,0.0002692]	10000	1
varianza sin1:year	-0.0002707	0.000744	[-0.001716,0.001185]	6100	1
varianza cos1:year	0.002532	0.000741	[0.001056,0.003972]	5900	1
varianza sin2:year	-0.001204	0.000742	[-0.002647,0.0002204]	6500	1
varianza cos2:year	-0.00421	0.000744	[-0.005692,-0.002772]	6300	1

Cuadro B11: Tabla con el resumen de las distribuciones a posteriori de Santa Eulalia y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Calatayud</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	56.16	0.952	[54.31,58.02]	4000	1
media y.lag	0.7174	0.00454	[0.7084,0.7262]	5500	1
media sin1	-14.4	1.28	[-16.91,-11.9]	3900	1
media cos1	-27.79	1.34	[-30.4,-25.18]	3500	1
media sin2	7.026	0.424	[6.21,7.867]	4200	1
media cos2	-1.131	0.441	[-2.023,-0.267]	4700	1
media sin3	0.002665	0.28	[-0.5452,0.5559]	11000	1
media cos3	-1.105	0.273	[-1.643,-0.5616]	11000	1
media y.lag:sin1	0.02963	0.0063	[0.01728,0.04195]	4100	1
media y.lag:cos1	-0.00293	0.00655	[-0.01563,0.009811]	4000	1
varianza (Intercept)	1.355	1.05	[-0.6986,3.386]	9200	1
varianza y.lag	0.001424	0.000214	[0.0009969,0.001837]	11000	1
varianza sin1	-0.04389	0.0612	[-0.1661,0.07374]	8700	1
varianza cos1	0.2004	0.0609	[0.08095,0.3194]	8300	1
varianza sin2	0.05333	0.042	[-0.02892,0.1372]	9300	1
varianza cos2	-0.02105	0.0426	[-0.1052,0.06258]	8800	1
varianza year	0.002543	0.000528	[0.001524,0.003581]	9200	1
varianza y.lag:sin1	0.0009391	0.000292	[0.0003667,0.001515]	9300	1
varianza y.lag:cos1	-0.001709	0.000301	[-0.002302,-0.00112]	9500	1
varianza y.lag:sin2	-0.000158	0.000195	[-0.0005363,0.0002239]	12000	1
varianza y.lag:cos2	-0.0007752	0.000195	[-0.001158,-0.0003889]	12000	1

Cuadro B12: Tabla con el resumen de las distribuciones a posteriori de Calatayud y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Panticosa</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-77.16	21.2	[-119.7,-36.98]	1100	1
media y.lag	0.7372	0.00435	[0.7285,0.7452]	2500	1
media sin1	-8.276	0.78	[-9.793,-6.763]	1100	1
media cos1	-20.88	0.889	[-22.65,-19.13]	1100	1
media sin2	3.481	0.38	[2.738,4.206]	1200	1
media cos2	2.264	0.372	[1.553,3.001]	1400	1
media year	0.05457	0.0107	[0.03443,0.07608]	1100	1
media y.lag:sin1	0.0004923	0.00611	[-0.01173,0.01225]	1100	1
media y.lag:cos1	0.006126	0.00629	[-0.006289,0.0183]	1300	1
varianza (Intercept)	6.821	0.0247	[6.772,6.869]	1200	1
varianza y.lag	-0.0006431	0.000196	[-0.001019,-0.0002553]	1300	1
varianza sin1	0.06901	0.0149	[0.03958,0.09862]	1200	1
varianza cos1	-0.02274	0.0198	[-0.0597,0.01672]	1300	1

Cuadro B13: Tabla con el resumen de las distribuciones a posteriori de Panticosa y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>La Puebla de Híjar</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-125.6	21.2	[-167,-84.14]	530	1
media y.lag	0.6772	0.00482	[0.6679,0.6868]	3400	1
media sin1	-12.83	1.37	[-15.45,-10.06]	850	1
media cos1	-36.49	1.55	[-39.4,-33.57]	1100	1
media sin2	6.592	0.427	[5.741,7.401]	950	1
media cos2	-0.9861	0.491	[-1.967,-0.03543]	1100	1
media sin3	-0.1602	0.283	[-0.7289,0.3985]	2000	1
media cos3	-1.235	0.285	[-1.779,-0.6656]	2100	1
media year	0.09821	0.0107	[0.07738,0.1192]	530	1
media y.lag:sin1	0.02239	0.00632	[0.009916,0.03445]	850	1
media y.lag:cos1	0.009092	0.00704	[-0.004372,0.02239]	1000	1
varianza (Intercept)	6.483	0.047	[6.391,6.571]	1000	1
varianza y.lag	0.001354	0.000218	[0.0009404,0.001785]	1000	1
varianza sin1	0.1767	0.0148	[0.1465,0.2054]	1900	1
varianza cos1	0.1526	0.0269	[0.1019,0.2058]	1100	1
varianza sin2	-0.08944	0.0141	[-0.1163,-0.06086]	2400	1
varianza cos2	0.124	0.0133	[0.0974,0.1495]	2300	1
varianza sin3	-0.02258	0.0134	[-0.04913,0.003333]	2700	1
varianza cos3	0.1144	0.0133	[0.08846,0.1399]	3100	1

Cuadro B14: Tabla con el resumen de las distribuciones a posteriori de La Puebla de Híjar y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Anso</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	40.52	0.807	[38.92,42.08]	3100	1
media y.lag	0.7591	0.00436	[0.7508,0.7676]	4200	1
media sin1	-9.681	1.02	[-11.68,-7.637]	3300	1
media cos1	-26.46	1.12	[-28.66,-24.26]	2600	1
media sin2	4.856	0.398	[4.054,5.639]	3500	1
media cos2	0.7356	0.407	[-0.06343,1.534]	3700	1
media y.lag:sin1	0.01349	0.00584	[0.001942,0.02501]	3500	1
media y.lag:cos1	0.02014	0.00618	[0.008123,0.03231]	3000	1
varianza (Intercept)	-8.76	1.07	[-10.85,-6.665]	6600	1
varianza y.lag	-0.0005188	0.000206	[-0.0009161,-0.0001153]	11000	1
varianza sin1	-0.03973	0.0498	[-0.1356,0.05719]	4900	1
varianza cos1	-0.1774	0.0527	[-0.2809,-0.07437]	4900	1
varianza sin2	0.08323	0.0199	[0.04435,0.122]	4700	1
varianza cos2	-0.1778	0.02	[-0.2166,-0.1388]	5600	1
varianza year	0.007816	0.000537	[0.006763,0.008867]	6600	1
varianza y.lag:sin1	0.0009043	0.000283	[0.0003499,0.001458]	5200	1
varianza y.lag:cos1	-0.0005825	0.000296	[-0.001162,-1.55e-05]	4900	1

Cuadro B15: Tabla con el resumen de las distribuciones a posteriori de Anso y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.



<b>Daroca</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-188.6	21.9	[-230.8,-144.6]	930	1
media y.lag	0.7082	0.00446	[0.6998,0.7172]	2500	1
media sin1	-14.63	1.19	[-16.96,-12.27]	1000	1
media cos1	-32.44	1.26	[-34.9,-29.95]	800	1
media sin2	6.916	0.424	[6.073,7.729]	1000	1
media cos2	0.7035	0.425	[-0.09905,1.554]	880	1
media year	0.1235	0.0111	[0.1009,0.145]	930	1
media y.lag:sin1	0.02926	0.00618	[0.01726,0.04158]	1000	1
media y.lag:cos1	0.01896	0.00624	[0.006534,0.03129]	800	1
varianza (Intercept)	6.461	0.0413	[6.381,6.543]	1100	1
varianza y.lag	0.001893	0.000213	[0.001467,0.002307]	1100	1
varianza sin1	0.1687	0.0149	[0.1395,0.1989]	1700	1
varianza cos1	-0.0578	0.0245	[-0.1062,-0.009806]	1100	1
varianza sin2	0.00385	0.0137	[-0.02429,0.02958]	2000	1
varianza cos2	-0.0781	0.0136	[-0.1049,-0.05163]	1800	1

Cuadro B16: Tabla con el resumen de las distribuciones a posteriori de Daroca y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Zaragoza</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-157.2	20.4	[-196.1,-119.3]	400	1
media y.lag	0.6789	0.0046	[0.6703,0.6879]	2600	1
media sin1	-14.59	1.37	[-17.34,-11.9]	1100	1
media cos1	-33.93	1.44	[-36.76,-31.13]	1300	1
media sin2	6.758	0.42	[5.949,7.571]	1200	1
media cos2	-1.162	0.442	[-2.017,-0.3158]	1300	1
media sin3	0.001023	0.276	[-0.5291,0.5381]	2100	1
media cos3	-1.217	0.27	[-1.73,-0.6666]	2400	1
media year	0.1129	0.0103	[0.09375,0.1325]	400	1
media y.lag:sin1	0.0304	0.0065	[0.01806,0.0432]	1100	1
media y.lag:cos1	0.002702	0.00679	[-0.01052,0.01598]	1200	1
varianza (Intercept)	6.123	0.051	[6.023,6.22]	1600	1
varianza y.lag	0.001899	0.000226	[0.00147,0.002345]	2100	1
varianza sin1	0.133	0.0741	[-0.003901,0.2806]	1500	1
varianza cos1	0.8059	0.0698	[0.6707,0.9423]	1300	1
varianza sin2	-0.1049	0.0228	[-0.1494,-0.06178]	1400	1
varianza cos2	-0.04966	0.0216	[-0.09266,-0.008539]	1600	1
varianza sin3	-0.01484	0.0136	[-0.04215,0.01171]	2800	1
varianza cos3	0.09223	0.0138	[0.06498,0.1193]	3100	1
varianza y.lag:sin1	0.000249	0.000341	[-0.0004337,0.0008777]	1500	1
varianza y.lag:cos1	-0.003682	0.000317	[-0.004288,-0.003071]	1400	1

Cuadro B17: Tabla con el resumen de las distribuciones a posteriori de Zaragoza y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>La Sotonera</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	36.1	18.6	[-0.005442,72.46]	2900	1
media y.lag	0.7102	0.00462	[0.7012,0.7192]	11000	1
media sin1	-53.97	25.4	[-104.4,-4.094]	2100	1
media cos1	94.8	26	[44.48,146.2]	3300	1
media sin2	5.356	0.432	[4.514,6.208]	5800	1
media cos2	0.3127	0.449	[-0.5623,1.2]	5400	1
media sin3	-0.2172	0.245	[-0.705,0.2631]	9400	1
media cos3	-1.011	0.24	[-1.477,-0.5413]	9900	1
media year	0.01075	0.00939	[-0.007848,0.0291]	2900	1
media y.lag:sin1	0.01064	0.00638	[-0.001686,0.02313]	5900	1
media y.lag:cos1	0.01191	0.0066	[-0.001222,0.02482]	5400	1
media sin1:year	0.02205	0.0128	[-0.003227,0.04741]	2100	1
media cos1:year	-0.06471	0.0131	[-0.09063,-0.03932]	3300	1
varianza (Intercept)	15.91	0.975	[14,17.83]	8800	1
varianza y.lag	0.0009654	0.000241	[0.0004872,0.00145]	11000	1
varianza sin1	-0.003891	0.0415	[-0.08499,0.07847]	11000	1
varianza cos1	-0.3402	0.0436	[-0.4254,-0.2546]	10000	1
varianza year	-0.004848	0.000492	[-0.005816,-0.003882]	8800	1
varianza y.lag:sin1	0.0007515	0.000201	[0.0003561,0.001141]	11000	1
varianza y.lag:cos1	0.0001971	0.00019	[-0.0001662,0.0005717]	9700	1

Cuadro B18: Tabla con el resumen de las distribuciones a posteriori de La Sotonera y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Pallaruelo</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-86.37	20.8	[-128.4,-46.57]	3800	1
media y.lag	0.6894	0.00478	[0.68,0.6987]	8700	1
media sin1	-13.42	1.42	[-16.16,-10.61]	3400	1
media cos1	-32.22	1.48	[-35.13,-29.3]	3400	1
media sin2	6.683	0.453	[5.802,7.567]	3600	1
media cos2	-2.292	0.482	[-3.244,-1.378]	4100	1
media sin3	0.1838	0.27	[-0.3443,0.7119]	8300	1
media cos3	-1.625	0.27	[-2.157,-1.092]	9200	1
media year	0.07539	0.0105	[0.05532,0.09669]	3800	1
media y.lag:sin1	0.0255	0.0067	[0.01216,0.03854]	3600	1
media y.lag:cos1	-0.009662	0.00685	[-0.02322,0.00385]	3500	1
varianza (Intercept)	13.95	1.04	[11.92,16.01]	8700	1
varianza y.lag	0.0004277	0.000228	[-1.923e-05,0.0008715]	10000	1
varianza sin1	0.05555	0.0705	[-0.08442,0.1911]	6400	1
varianza cos1	0.04804	0.0728	[-0.09261,0.192]	5700	1
varianza sin2	0.06447	0.0701	[-0.0722,0.2041]	4300	1
varianza cos2	0.2941	0.0691	[0.1575,0.4265]	4800	1
varianza sin3	-0.05373	0.0234	[-0.1,-0.008316]	4800	1
varianza cos3	0.02082	0.0221	[-0.02217,0.06418]	5400	1
varianza sin4	-0.05411	0.0136	[-0.08078,-0.02725]	9300	1
varianza cos4	0.04392	0.0135	[0.01738,0.06984]	9900	1
varianza year	-0.003733	0.000525	[-0.004781,-0.002702]	8700	1
varianza y.lag:sin1	0.0002816	0.000325	[-0.0003411,0.0009231]	8600	1
varianza y.lag:cos1	-0.0008901	0.000317	[-0.001516,-0.000265]	7300	1
varianza y.lag:sin2	-0.0006337	0.00033	[-0.00128,2.191e-05]	4900	1
varianza y.lag:cos2	-0.001341	0.000315	[-0.001953,-0.0007191]	5300	1

Cuadro B19: Tabla con el resumen de las distribuciones a posteriori de Pallaruelo y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Cueva Foradada</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-147.7	18.4	[-183.7,-111.2]	4800	1
media y.lag	0.7117	0.00462	[0.7027,0.7208]	8000	1
media sin1	-13.49	1.22	[-15.85,-11.05]	3600	1
media cos1	-28.85	1.25	[-31.35,-26.46]	3300	1
media sin2	5.86	0.388	[5.084,6.616]	3900	1
media cos2	0.579	0.383	[-0.1619,1.334]	3800	1
media year	0.1027	0.00936	[0.08417,0.121]	4700	1
media y.lag:sin1	0.03072	0.00633	[0.01838,0.04295]	3800	1
media y.lag:cos1	0.01333	0.00623	[0.001399,0.02584]	3300	1
varianza (Intercept)	-10.76	1.05	[-12.84,-8.712]	8300	1
varianza y.lag	0.002949	0.000239	[0.002485,0.003424]	10000	1
varianza sin1	0.1833	0.0644	[0.05537,0.3101]	6500	1
varianza cos1	0.3918	0.0667	[0.262,0.5262]	6100	1
varianza sin2	-0.05403	0.0204	[-0.09404,-0.01418]	6100	1
varianza cos2	-0.08971	0.0204	[-0.1302,-0.04938]	6700	1
varianza year	0.008388	0.000535	[0.007345,0.009447]	8200	1
varianza y.lag:sin1	0.0003521	0.000322	[-0.0002821,0.0009826]	6600	1
varianza y.lag:cos1	-0.001203	0.000326	[-0.00186,-0.0005712]	6300	1

Cuadro B20: Tabla con el resumen de las distribuciones a posteriori del embalse de Cueva Foradada y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<b>Sallent</b>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	-58.63	21.3	[-101.2,-17.38]	3800	1
media y.lag	0.7431	0.00441	[0.7346,0.7518]	8100	1
media sin1	-8.534	0.95	[-10.38,-6.675]	3700	1
media cos1	-25.55	1.04	[-27.56,-23.48]	3300	1
media sin2	3.969	0.412	[3.151,4.781]	4000	1
media cos2	2.257	0.408	[1.443,3.04]	4500	1
media year	0.04856	0.0108	[0.0278,0.06995]	3800	1
media y.lag:sin1	0.001537	0.00624	[-0.01063,0.01356]	3900	1
media y.lag:cos1	0.01979	0.0062	[0.007472,0.03175]	3700	1
varianza (Intercept)	20.87	1.02	[18.88,22.83]	8900	1
varianza y.lag	0.0001837	0.000205	[-0.0002238,0.0005814]	11000	1
varianza sin1	-1.279	1.41	[-4.031,1.51]	7900	1
varianza cos1	7.075	1.44	[4.304,9.963]	7800	1
varianza sin2	0.04166	0.0195	[0.003728,0.07945]	10000	1
varianza cos2	-0.05857	0.019	[-0.0953,-0.02152]	10000	1
varianza year	-0.007088	0.000515	[-0.008084,-0.006083]	8900	1
varianza y.lag:sin1	0.001434	0.000294	[0.0008562,0.002004]	11000	1
varianza y.lag:cos1	0.0008788	0.000282	[0.000324,0.001427]	10000	1
varianza sin1:year	0.0005763	0.000713	[-0.0008344,0.001963]	7900	1
varianza cos1:year	-0.003704	0.000728	[-0.005167,-0.0023]	7800	1

Cuadro B21: Tabla con el resumen de las distribuciones a posteriori de Sallent y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

<i>Yesa</i>	Media	$\sigma$	Intervalo de credibilidad	$N_{eff}$	$\hat{R}$
media (Intercept)	60.24	0.924	[58.44,62.04]	3400	1
media y.lag	0.6798	0.00486	[0.6703,0.6893]	3300	1
media sin1	-9.899	0.347	[-10.59,-9.212]	5500	1
media cos1	-31.51	0.556	[-32.59,-30.4]	3700	1
media sin2	5.646	0.308	[5.052,6.254]	7600	1
media cos2	-0.8002	0.294	[-1.364,-0.225]	9500	1
varianza (Intercept)	16.78	1.03	[14.76,18.79]	8800	1
varianza y.lag	0.001509	0.000205	[0.001107,0.001911]	10000	1
varianza sin1	0.1417	0.0396	[0.06329,0.2182]	9000	1
varianza cos1	-0.1733	0.0408	[-0.2546,-0.09356]	10000	1
varianza year	-0.005106	0.000521	[-0.006119,-0.00409]	8900	1
varianza y.lag:sin1	-0.0001073	0.000196	[-0.0004941,0.0002839]	8600	1
varianza y.lag:cos1	-4.363e-05	0.000191	[-0.0004194,0.0003363]	9800	1

Cuadro B22: Tabla con el resumen de las distribuciones a posteriori de *Yesa* y los valores del tamaño efectivo y el  $\hat{R}$  de cada parámetro.

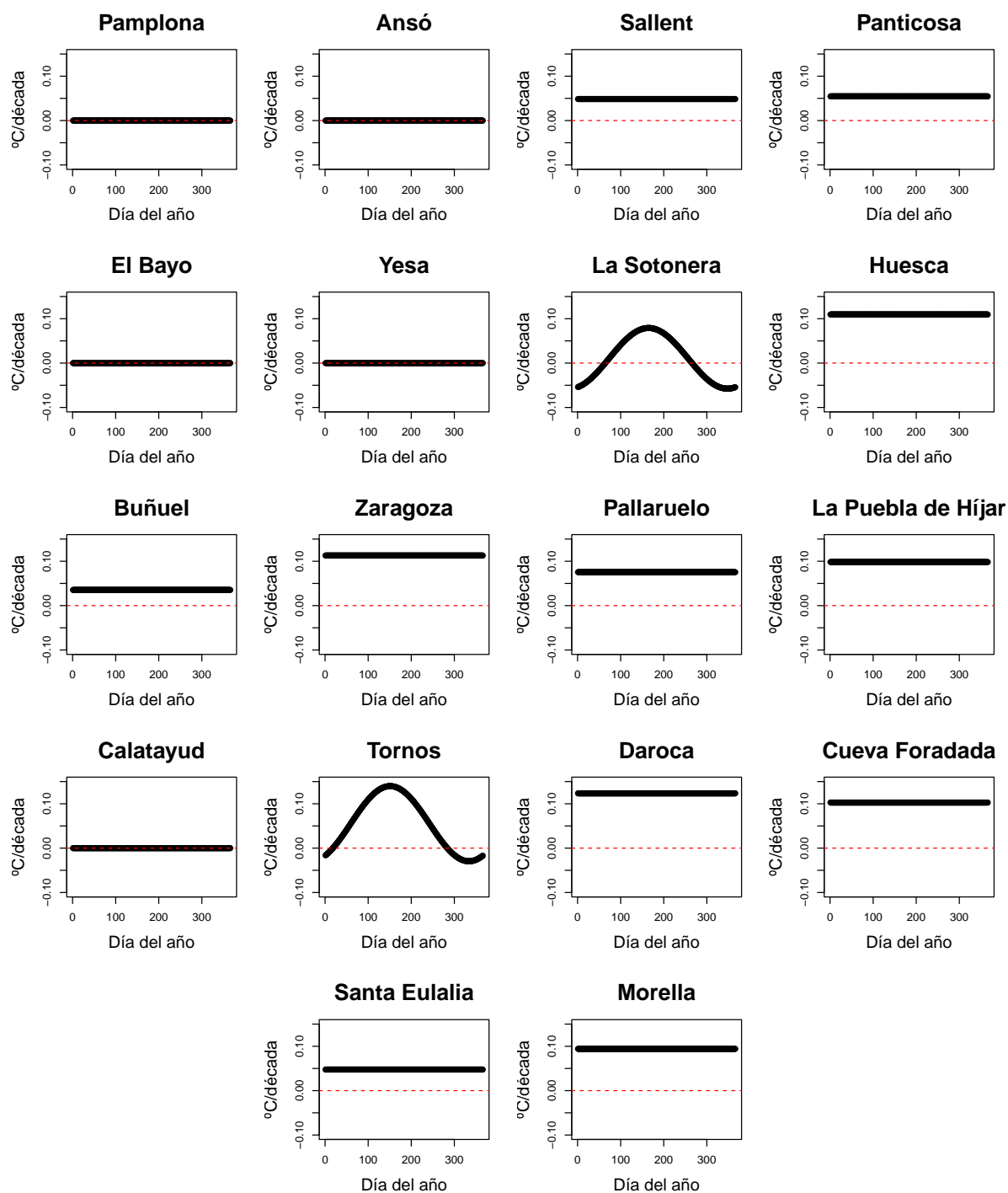


Figura B5: Media a posteriori del término de la tendencia a lo largo del año en cada lugar en el submodelo de la media.

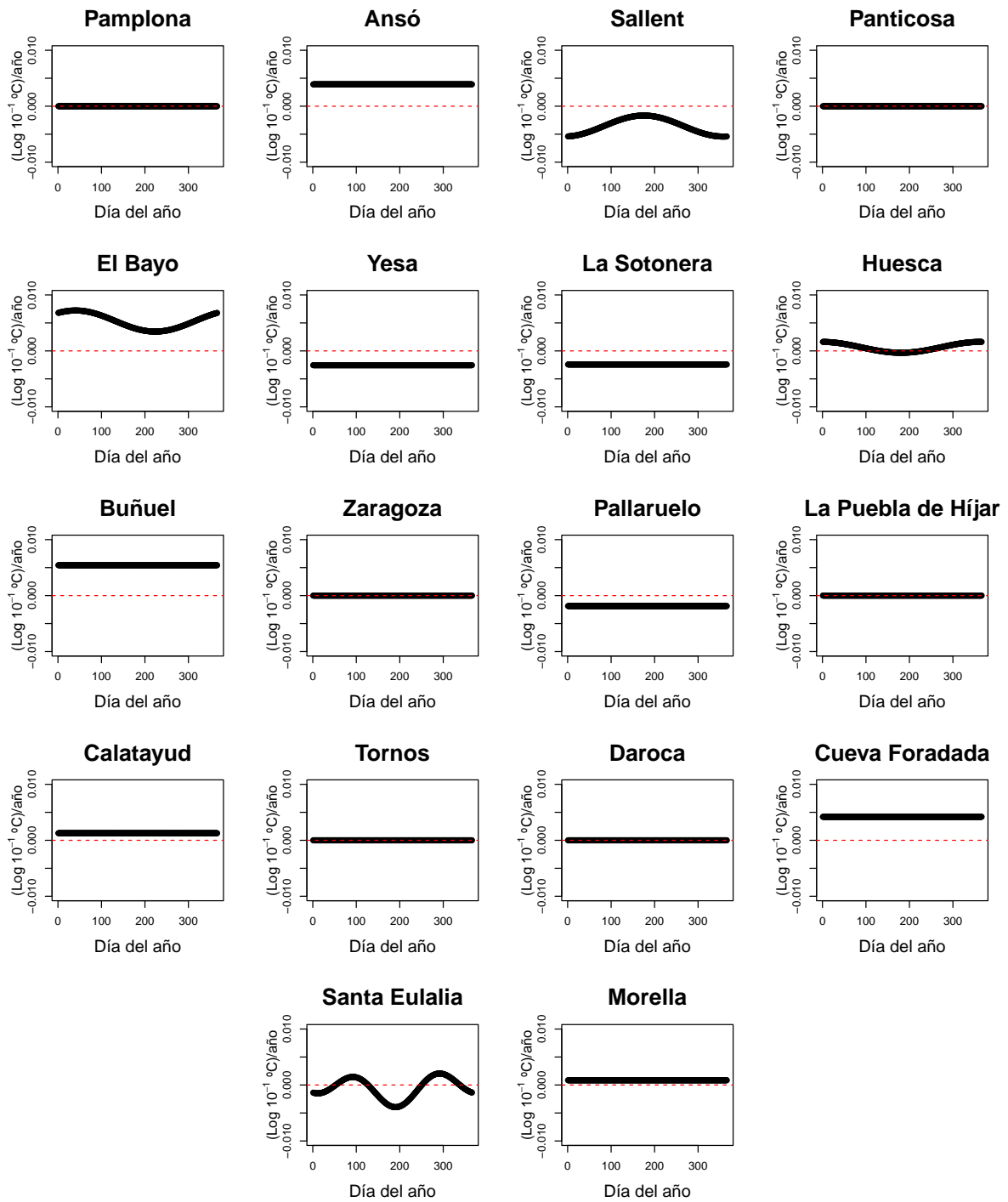


Figura B6: Media a posteriori del término de la tendencia a lo largo del año en cada lugar en el submodelo de la desviación típica (escala logarítmica).

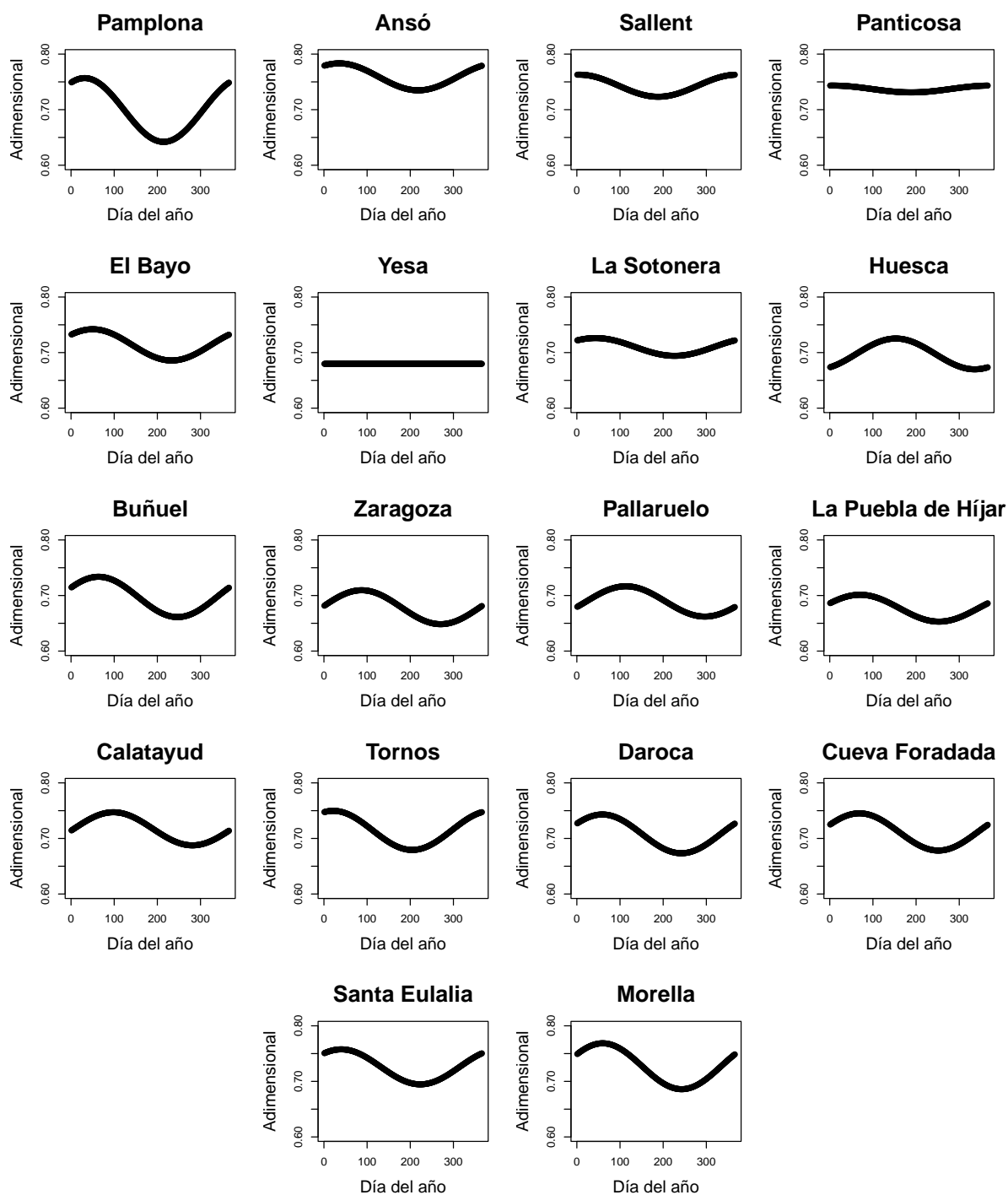


Figura B7: Media a posteriori del término autorregresivo a lo largo del año en cada lugar en el submodelo de la media.



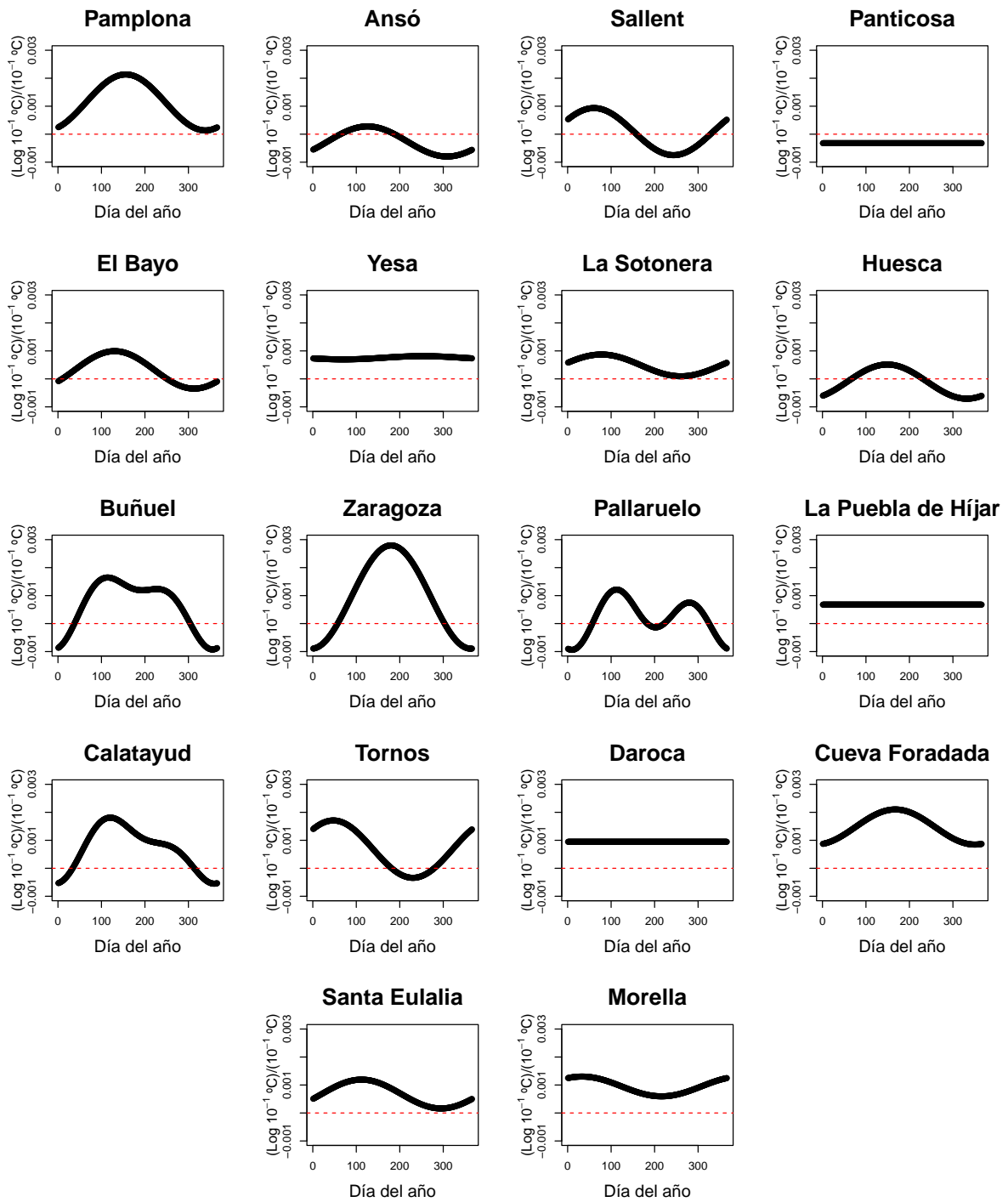


Figura B8: Media a posteriori del término autorregresivo a lo largo del año en cada lugar en el submodelo de la desviación típica (escala logarítmica).