Comunicaciones en Estadística
Diciembre 2023, Vol. 16, No. 2, pp. 2–20

Bayesian Hierarchical Modeling of Spatiotemporal Data

Modelamiento Jerárquico Bayesiano de Datos Espaciotemporales

Juan Sosa[a]
jcsosam@unal.edu.co

Resumen

In this paper, we illustrate in-depth Bayesian hierarchical statistical modeling approaches. Bayesian hierarchical modeling provides a robust framework for analyzing spatial data, accommodating complex dependencies, making possible incorporating external knowledge into the analysis. To do so, we consider a dataset from 80 stations in the Venezuelan state of *Guárico* consisting of accumulated monthly rainfall in a time span of 16 years. The spatial correlation is modeled by using a Matérn correlation function with a fixed smoothness parameter. Following Banerjee et al. (2014), we examine two fully Bayesian parametric approaches: One of them static, based on a hierarchical model with latent variables; and the other spatiotemporal, based on the dynamic framework given in West and Harrison (2006). Both alternatives are sensible ones, but due to the nature of the data, the dynamic model is more appealing since it gives a complete spatiotemporal characterization of the response variable.

***Palabras clave*:** Bayesian estimation; dynamic models; hierarchical modeling; Markov chain Monte Carlo; spatial statistics.

Abstract

En este documento, ilustramos enfoques de modelado estadístico jerárquico Bayesiano en profundidad. El modelado jerárquico Bayesiano proporciona un marco sólido para analizar datos espaciales, acomodando dependencias complejas, haciendo posible la incorporación de conocimiento externo al análisis. Para hacerlo, consideramos un conjunto de datos de 80 estaciones en el estado venezolano de Guárico, asociado con la precipitación mensual acumulada en un período de 16 años. La correlación espacial se modela utilizando una función de correlación de Matérn con un parámetro de suavidad fijo. Siguiendo a Banerjee et al. (2014), examinamos dos enfoques paramétricos completamente Bayesianos: uno de ellos

[a]Departamento de Estadística, Universidad Nacional de Colombia

estático, basado en un modelo jerárquico con variables latentes; y el otro espaciotemporal, basado en el marco dinámico de West y Harrison (1997). Ambas alternativas son importantes, pero debido a la naturaleza de los datos, el modelo dinámico es más atractivo ya que proporciona una caracterización espaciotemporal completa de la variable respuesta.

***Keywords*:** Estimación Bayesiana; modelos dinámicos; modelado jerárquico; Cadenas de Markov de Monte Carlo; estadística espacial.

1. Introduction

Bayesian hierarchical modeling of spatial data has wide-ranging applications, from environmental science, where it is used to model phenomena like rainfall and pollution levels (Gelfand et al., 2010), to epidemiology, for analyzing the spread of diseases over geographic regions (Lawson, 2018). By incorporating prior knowledge and accommodating complex dependencies, Bayesian hierarchical modeling offers a powerful tool for analyzing and interpreting spatial data, providing insights that are both statistically rigorous and practically relevant (Cressie, 2015).

Such an alternative to modeling data constitutes a robust statistical approach that leverages a multi-stage structure to analyze complex spatial structures. At its core, this methodology involves several stages, each representing different sources of variability. The first stage deals with the observed data, while the second stage (typically including several layers) captures the underlying spatial process, often incorporating latent variables to model unobserved factors (Banerjee et al., 2014). Finally, the third stage includes hyperparameters that influence the distributions at the process level. By organizing the model hierarchically, this approach allows for a nuanced representation of spatial dependencies and uncertainties (Gelman et al., 2013).

Under the Bayesian parading, the primary objective is to derive the posterior distribution of the model parameters, which combines prior information with the likelihood of the observed data (Hoff 2009, Gelman et al. 2013, Reich and Ghosh 2019). Bayesian inference techniques are then used to draw conclusions about the parameters themselves, and most importantly, the underlying spatial processes. A key feature of spatial data is the presence of spatial correlation, where observations that are closer in proximity tend to be more similar (Cressie, 2015). This spatial dependence is modeled using covariance functions like the Matérn covariance function and Gaussian processes, which offer a flexible framework for capturing these correlations (Williams and Rasmussen 2006, Stein 2012). We refer the reader to (Banerjee et al., 2014, Chap. 2) for a exhaustive survey on covariance functions.

To perform the necessary Bayesian computations, Markov chain Monte Carlo (MCMC) methods are employed, allowing for sampling from the posterior distribution of the model parameters (Robert et al. 2004, Gamerman and Lopes 2006). In scenarios involving spatiotemporal data, dynamic models, including state-space

models, are used to account for both spatial and temporal dependencies, providing a comprehensive characterization of the data (West and Harrison 2006, Cressie 2015). Model validation is an essential component of Bayesian hierarchical modeling, with techniques such as posterior predictive checks and spatial cross-validation used to assess model fit and predictive performance (Gelman et al. 2013, Banerjee et al. 2014). Even though there is a humongous literature on spatiotemporal modeling, we strongly recommend the following books about the matter: Gelfand et al. (2010), Diggle (2013), Banerjee et al. (2014), Cressie (2015), Wikle et al. (2019), Oyana (2020), Sahu (2022), and Kent and Mardia (2022).

In this paper, we illustrate in-depth a couple of Bayesian hierarchical statistical modeling approaches. To do so, we consider a dataset from 80 stations in the Venezuelan state of *Guárico* consisting of accumulated monthly rainfall in a time span of 16 years. The spatial correlation is modeled by using a Matérn correlation function with a fixed smoothness parameter. Following Banerjee et al. (2014), we examine two fully Bayesian parametric approaches: One of them static, based on a hierarchical model with latent variables; and the other spatiotemporal, based on the dynamic framework given in West and Harrison (2006). Both alternatives are sensible ones, but due to the nature of the data, the dynamic model is more appealing since it gives a complete spatiotemporal characterization of the response variable.

The rest of the document is structured as follows: Section 2 provides an exhaustive description of the data. Then, Section 3 analyzes the data in a cross-sectional way using a static hierarchical model. Next, Section 4 does the same but dynamically using dynamic spatiotemporal model. Finally, Section 5 discusses our main findings and some alternatives for future research.

2. The Guárico dataset

Guárico is one of the 23 states of *Venezuela* covering a total surface area of 64,986 km^2 (7.1 % of Venezuela). Guárico's geographical location is Latitude 7^o 39'; 10^o 02' North, Longitude 64^o 45'; 68^o 02' West (UTM[1] zone 19) (**?**). Left panel in Figure 1 shows the location of Guárico in Venezuela.

[1]Universal Transverse Mercator.

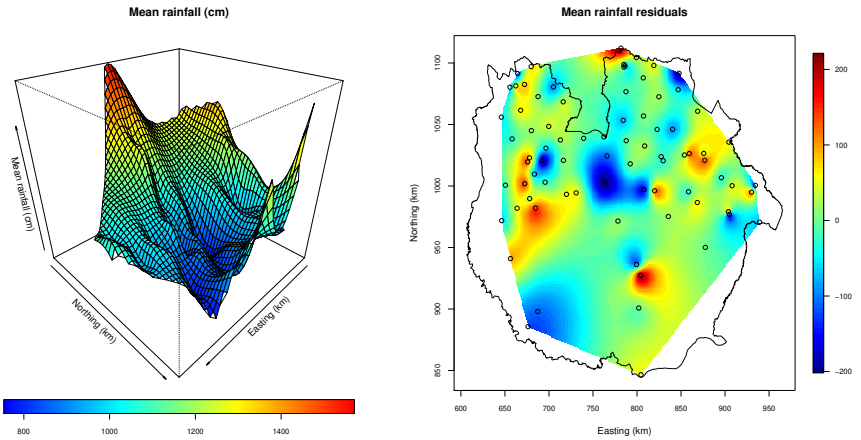


Figure 1: Left panel: Location of *Estado de Guárico* (in red), Venezuela. Right panel: State of Guárico (in yellow) and location of the 80 stations (in red).

We consider a dataset of $n = 80$ different stations in the Venezuelan state of Guárico, consisting of accumulated monthly rainfall (in centimeters) in a time span of 16 years from 1968 to 1983. Right panel in Figure 1 shows the location of the 80 stations on a map of Guárico in which the elevation is also perceptible through the satellite image –map downloaded from the Google server using package `RgoogleMaps` in `R` (Loecher, 2012). The stations are irregularly scattered along Guárico's territory. Assuming that the different years are independent replicates of the same random field and focusing only on the spacial structure of the data, we firstly consider the *mean rainfall* as response variable, and the planar coordinates (using the UTM projection system in kilometers) and the altitude (in kilometers) as covariates.

First of all, we perform a graphical exploration of the rainfall as a function of the location and the altitude in order to visually explore the trend function to be consider in the models (see Sections 3 and 4 for details). Figure 2 displays the 80 locations in Guárico along with the mean rainfall intensity. This plot strongly suggests that location is a fundamental factor in characterizing mean rainfall features (as expected) because precipitation is clearly more intense towards Western (on the border) and Southern Guárico than in the North-East of the state –the *Cordillera de la Costa* in Venezuela might have a fundamental impact on the precipitation activity of northern Guárico. The previous behavior in precipitations is also apparent in Figure 3 where a interpolation of the mean rainfall was carried out by means of multilevel B-splines (Banerjee et al., 2014, p.46,47); a quadratic trend in relation to location is recognizable due to the parabolic shape of the surface.

Finally, Figure 4 makes also clear the quadratic association between the mean rainfall and the altitude where the corresponding station is located at. Figures exhibited in Section 4 and further exploration (see Section 3) confirm that it is

appropriate to consider quadratic terms in the trend. We use powerful `R` libraries (R Core Team et al., 2013), such as `fields`, `geoR`, `maps`, `Rgoolemaps`, and in particular `spBayes` (Finley et al., 2007) in this work in order to fit the data and produce the corresponding plots.

3. Static hierarchical modeling

Taking into account the descriptive characterization of the trend given in Section 2, we consider a trend function $\mu(\boldsymbol{s})$ including the intercept, all the linear and quadratic terms of easting, northing and elevation (7 terms in total); this trend function produces a fit in which all terms are *significant* but elevation (see Table 1). That is why we consider the parsimonious trend function given by

$$\mu(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\top}\boldsymbol{\beta} = \sum_{j=0}^{6} \beta_j \, x_j(\boldsymbol{s}) \tag{1}$$

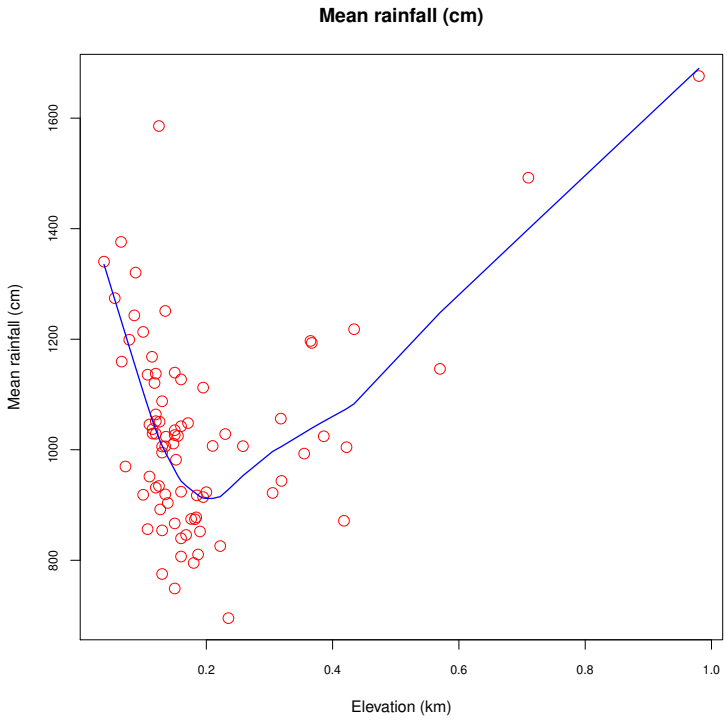


Figure 2: Guárico mean rainfall data. Sampling locations are shown as circles, with the radius of each circle proportional to the corresponding mean rainfall.

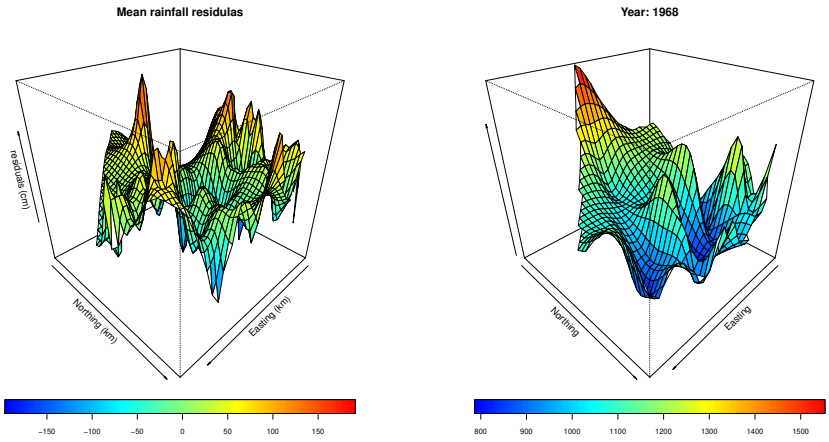


Figure 3: Interpolation of mean rainfall using multilevel B-splines.

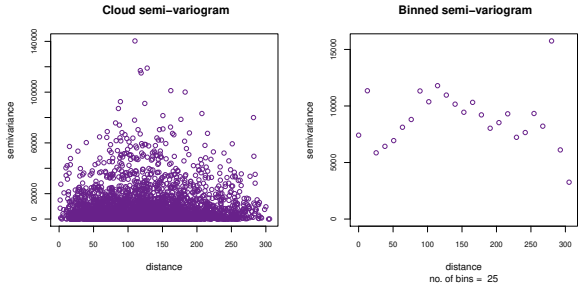


Figure 4: Guárico mean rainfall against elevation.

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_6]^\mathsf{T}$, and $\boldsymbol{x}(\boldsymbol{s}) = [x_0(\boldsymbol{s}), x_1(\boldsymbol{s}), \ldots, x_6(\boldsymbol{s})]^\mathsf{T}$, with

$$
\begin{aligned}
&x_0(\boldsymbol{s}) \equiv 1\,, \\
&x_1(\boldsymbol{s}) \equiv \text{easting}(\boldsymbol{s})\,, \qquad x_2(\boldsymbol{s}) \equiv \text{northing}(\boldsymbol{s})\,, \qquad x_3(\boldsymbol{s}) \equiv \text{elevation}(\boldsymbol{s})\,, \\
&x_4(\boldsymbol{s}) \equiv \text{easting}^2(\boldsymbol{s})\,, \qquad x_5(\boldsymbol{s}) \equiv \text{northing}^2(\boldsymbol{s})\,, \qquad x_6(\boldsymbol{s}) \equiv \text{elevation}^2(\boldsymbol{s})\,.
\end{aligned}
$$

Then, we dispose of $p = 7$ covariates in this case.

An ordinary linear regression of mean rainfall using the trend function (1) leads to the mean rainfall residuals shown in the left panel of Figure 5. Although the Q-Q plot in this Figure suggests that the distribution of the residuals is slightly left-skewed, no obvious deviation from a Gaussian assumption is detected. This is also suggested by a Shapiro-Wilk test whose corresponding p-value is 0.5152. Therefore, we do not perform any additional transformations to the data in order to achieve normality.

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
β_0 intercept	20929.94	3366.24	6.22	0.0000
β_1 easting	-13.56	2.54	-5.34	0.0000
β_2 northing	-26.94	6.25	-4.31	0.0001
β_3 elevation	127.23	334.82	0.38	0.7050
β_4 I(easting^2)	0.01	0.00	5.09	0.0000
β_5 I(northing^2)	0.01	0.00	3.88	0.0002
β_6 I(elevation^2)	683.50	306.40	2.23	0.0288

Table 1: Results for the regression with linear predictor given in (1). All terms are *significant* but the one corresponding to elevation. Residual standard error: 87 on 73 degrees of freedom. Multiple R-squared: 0.784. Adjusted R-squared: 0.767. F-statistic: 44.3 on 6 and 73 degrees of freedom, p-value: $<$ 2E−16.

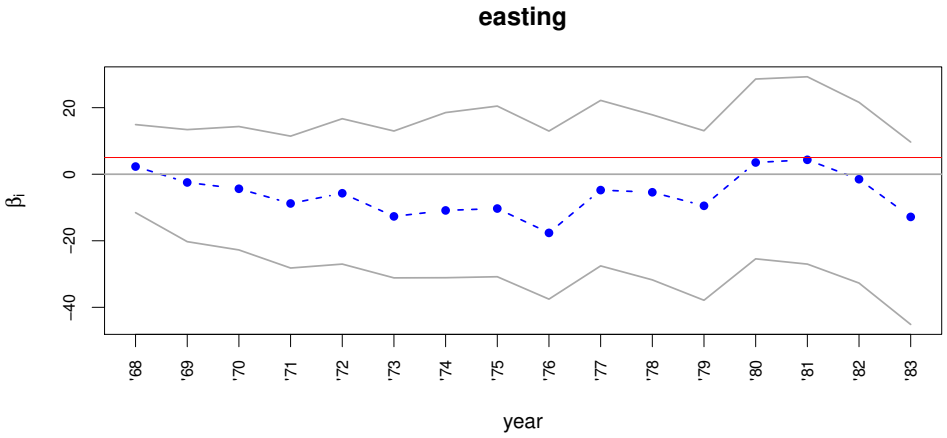


Figure 5: Values (left panel) and Q-Q normality plot (right panel) corresponding to the mean rainfall residuals using the trend function (1).

In addition, the residual image plot in Figure 6 suggests that there is spatial dependence even after accounting for the covariates. These patterns can be formally examined using empirical semi-variograms. Two versions of semi-variograms are offered in Figure 7. This figure indicates that the nugget effect considerably influences the variability of the mean rainfall, which will be confirmed later.

Additionally, directional semi-variograms of the detrended mean rainfall in the four cardinal directions (not shown) enable us to identify potential anisotropic patterns in spatial dependence. Although we observe jagged and erratic shapes, particularly in the 0^o directional semi-variogram, and some indications of sill anisotropies at distances of 200 and 300 km in the Northwest direction, we do not place significant emphasis on the implications of these directional semi-variograms. It is risky to overinterpret and attribute too much significance to directional semi-variograms.

According to Banerjee et al. (2014, p. 38), plots of data generated from a simple isotropic model will typically show differences of the magnitudes seen in the figure.

We consider binned semi-variograms and fitted semi-variograms in the Matérn[2] family for the mean rainfall residuals (not shown here), using four different values of the smoothness parameter ν, namely, $\nu = 0.5, 1.0, 1.5, 2.5$, showing the sill (σ^2), the nugget (τ^2), and the effective range (distance at which the correlation drops to 0.05). The Matérn family of covariance functions is preferred for its flexibility in modeling varying degrees of spatial smoothness, clear parameter interpretation, and ability to encompass other common covariance functions. Furthermore, we compute estimates of these quantities and the inverse of the spacial decay ϕ (not shown here), for all the considered values of ν. These values indicate a significant nugget effect. Furthermore, we observe that $\hat{\tau}$ increases with ν because an increase in the assumed smoothness of the Gaussian field (measured by its mean-square differentiability) is compensated by a corresponding increase in the estimated nugget variance $\hat{\tau}^2$. We notice also the non-orthogonality between ν and ϕ. As ν increases, $\hat{\phi}$ decreases. This again illustrates a general feature of the Matérn model, specifically that the interpretation of ϕ cannot be made independently of ν (Diggle and Ribeiro, 2007, p. 120).

Generally, the fitted semi-variogram is satisfactory, particularly for small distances where the sample variogram exhibits relatively high precision (Diggle and Ribeiro, 2007, p. 121). We see that there is not a "best fit" in the sense that all cases perform relatively similar. Recall that for the Matérn family, the corresponding process is k-times mean-square differentiable if and only if $\nu > k$. Matérn covariance functions are particularly simple when ν is a half-integer, represented as $\nu = p + 1/2$, where p is a non-negative integer (Rasmussen, 2006, p. 85). If a choice must be made for

[2]We consider the following parameterization of the Matérn family of covariance functions: $C(h) = (1/[2^{\nu-1}\,\Gamma(\nu)])\,(h/\phi)^{\nu} K_{\nu}(h/\phi)$, where K_{ν} is the modified Bessel function of order ν.

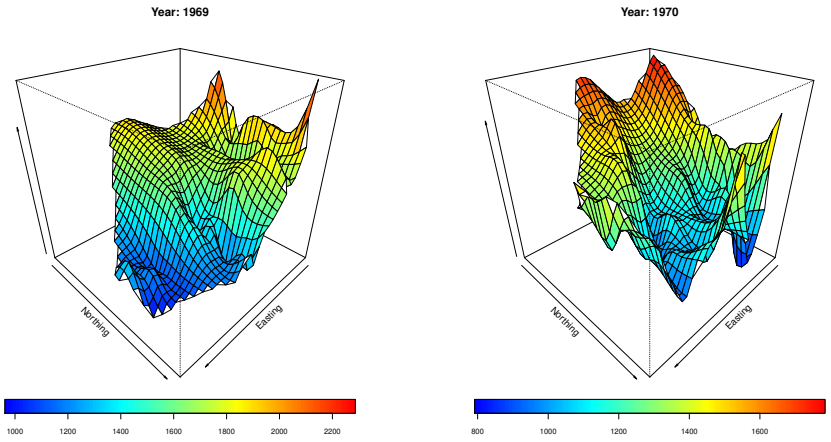


Figure 6: Interpolation of mean rainfall residuals using multilevel B-splines.

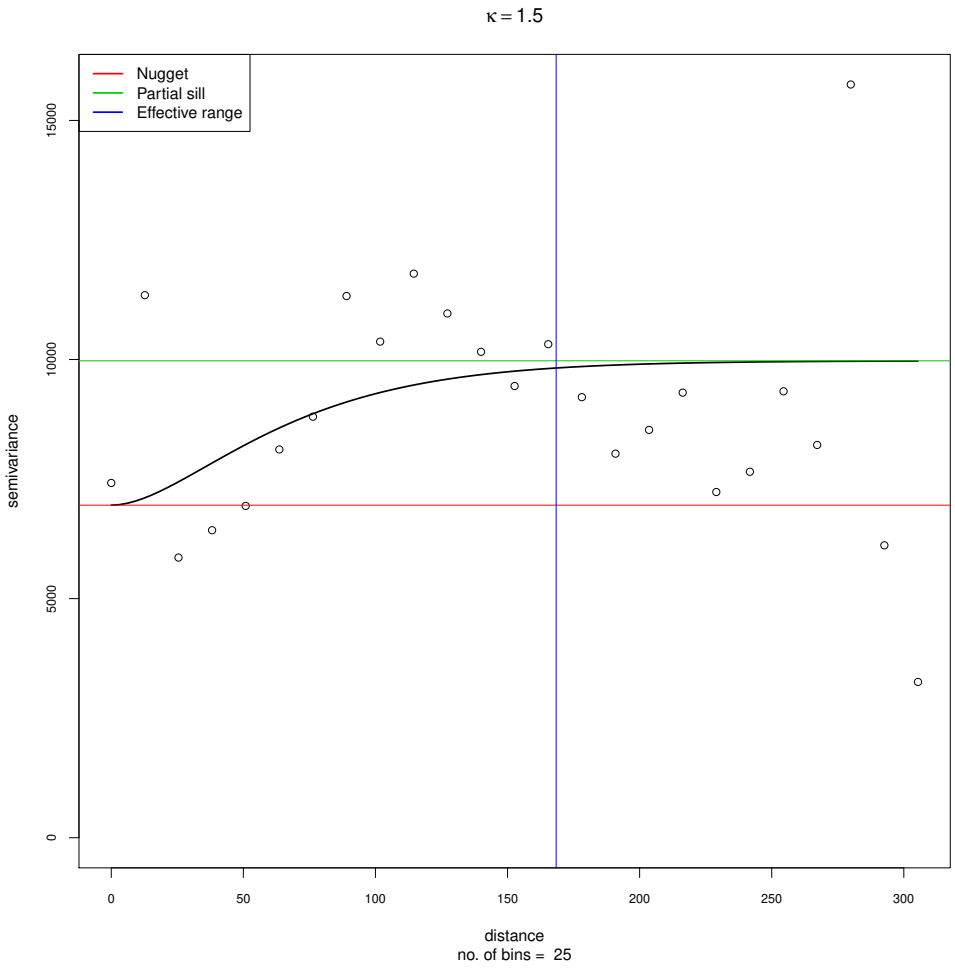


Figure 7: Cloud (isotropic) semi-variogram (left panel) and Binned (isotropic) semi-variogram for the mean rainfall residuals.

the smoothness parameter, $\nu = 2.5$ would be a more reasonable and parsimonious option.

Figure 8 displays binned semi-variograms and fitted semi-variograms in the Matérn family for the response variable and the residuals, using $\nu = 2.5$ as smoothness parameter. Here, the upper (green) and lower (blue) horizontal lines are the sill (σ^2) and the nugget (τ^2), respectively, and the vertical (red) line is the effective range. Furthermore, Table 2 exhibits the corresponding estimates (including the spacial decay ϕ). These estimates serve as starting values in the Markov chain Monte Carlo (MCMC) algorithms implemented below and in Section 4. This table also highlights the significant nugget effect present in this problem.

Quantity	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$	Eff. Range
Response	37,419.96	42.77	11,233.39	253.12
Residuals	1,464.91	22.25	5,815.35	131.69

Table 2: Estimates of the parameters in the Matérn family and the nugget, and effective range, using $\nu = 2.5$ as smoothness parameter.

Now, we consider the static model

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\beta} + w(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \tag{2}$$

where $\boldsymbol{x}(\boldsymbol{s})$ and $\boldsymbol{\beta}$ are given as in (1). The residual is partitioned into two pieces: The $w(\boldsymbol{s})$ are assumed to be realizations from a zero-centered stationary Gaussian spatial process, capturing residual spatial association, while the $\epsilon(\boldsymbol{s})$ represent uncorrelated pure error terms. Thus, the $w(\boldsymbol{s})$ introduce the sill (σ^2) and spatial decay (ϕ) parameters, while the $\epsilon(\boldsymbol{s})$ add the nugget effect (τ^2).

We have data $y(\boldsymbol{s}_i)$, for $i = 1, \ldots, n$. Let $\boldsymbol{y} = [y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)]^{\mathsf{T}}$ be the n-dimensional vector gathering the mean rainfall corresponding to the $n = 80$ locations. Collecting the entire collection of model parameters into a single vector $\boldsymbol{\theta} = [\boldsymbol{\beta}, \sigma^2, \tau^2, \phi]$, and the random effects into the vector $\boldsymbol{w} = [w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)]^{\mathsf{T}}$, parameter estimates can be obtained from the posterior distribution

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{w})\, p(\boldsymbol{w} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\,,$$

with

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{w}) = \mathsf{N}(\boldsymbol{y} \mid \mathbf{X}\boldsymbol{\beta} + \boldsymbol{w}, \tau^2\mathbf{I}_n), \qquad p(\boldsymbol{w} \mid \boldsymbol{\theta}) = \mathsf{N}(\boldsymbol{w} \mid \mathbf{0}, \sigma^2\mathbf{H}(\phi)),$$

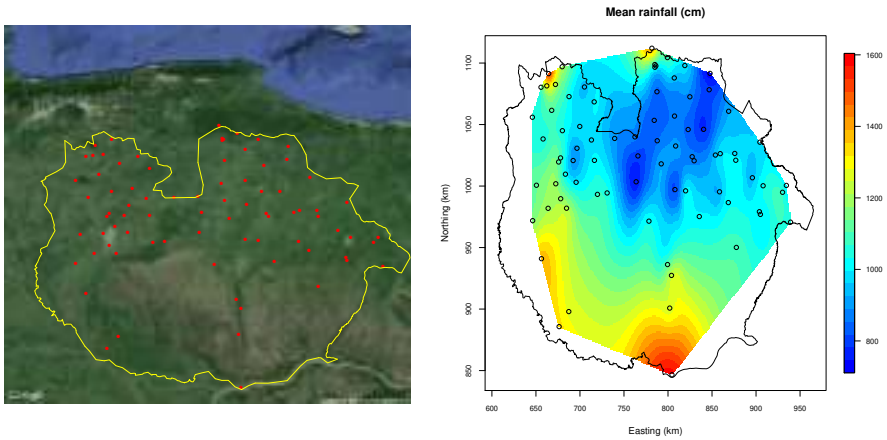


Figure 8: Binned (isotropic) semi-variograms and fitted semi-variograms in the Matérn family (solid black line) for the mean rainfall and the mean rainfall residuals, using $\nu = 2.5$ as smoothness parameter. The upper (green) and lower (blue) horizontal lines sill (σ^2) and the nugget (τ^2), respectively, and the vertical (red) line is the effective range.

where $\mathbf{X}$ is an $n \times p$ design matrix storing of all the covariate information given by the p-dimensional vectors $\boldsymbol{x}(\boldsymbol{s})$ in each one of the n locations, and $\mathbf{H}(\phi)$ is a correlation matrix with $H_{ij} = \rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \phi)$ and ρ is the Matérn correlation function on $\mathbb{R}^2$ (which is a valid correlation function) indexed by the spatial decay parameter ϕ (The smoothness parameter ν is fixed at a value of 2.5). Furthermore, independent priors are chosen for the remaining parameters, i.e., $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})\, p(\sigma^2)\, p(\tau^2)\, p(\phi)$, where

$$p(\boldsymbol{\beta}) = \mathsf{N}(\boldsymbol{\beta} \mid \mathbf{0}_p, b_{\boldsymbol{\beta}}\, \mathbf{I}_p)\,, \qquad p(\sigma^2) = \mathsf{IG}(\sigma^2 \mid a_{\sigma^2}, b_{\sigma^2})\,,$$
$$p(\tau^2) = \mathsf{IG}(\tau^2 \mid a_{\tau^2}, b_{\tau^2}), \qquad p(\phi) = \mathsf{U}(\phi \mid a_\phi, b_\phi)\,.$$

As stated before, the empirical semi-variogram estimates of the sill (σ^2), nugget (τ^2), and spatial decay (ϕ) in Table 2 offer valuable insights for setting the hyperparameters in the priors mentioned above, and serve as effective starting values for implementing the hierarchical model. We implement a MCMC algorithm with $40,000$ iterations to generate samples from the posterior distribution. Trace and autocorrelation plots (not shown here) of the regression parameters $\boldsymbol{\beta}$ show that the corresponding chains achieve convergence quickly and that there is no signs of autocorrelation issues. Similarly, chains of σ^2, τ^2, and ϕ displayed in Figure 9 show reasonably fast convergence, but they exhibit a notably volatile behavior, particularly the chain corresponding to σ^2. The overall Metropolis acceptance rate is 39.31 %.

We discard the initial 32,000 iterations (burn-in period) and, to ensure approximately independent draws, we select one sample from every 32 iterations, resulting in a total of 1000 posterior samples. Those 1000 samples are approximately independent and identically distributed according to the corresponding posterior distribution and form the basis of posterior inference. Table 3 summarizes the posterior distribution of all the parameters in the model. We present there the correspondent posterior medians and 95 % posterior credible intervals.

We also obtain the posterior mean for the spatial random effects of the model. These posterior means can then be interpolated across the domain to produce

Parameter	2.5 %	50 %	97.5 %
σ^2 sill	10,856.85	28,990.83	98,383.63
τ^2 nugget	5,203.77	7,514.75	10,219.66
$1/\phi$ decay	0.0100	0.0174	0.0295
β_0 intercept	-1,809.63	184.15	2,263.38
β_1 easting	-8.93	5.02	21.64
β_2 northing	-14.11	0.21	12.18
β_3 elevation	-462.98	211.49	975.6182
β_4 I(easting^2)	-0.0136	-0.0034	0.0053
β_5 I(northing^2)	-0.0071	-0.0010	0.0065
β_6 I(elevation^2)	6.18	595.30	1,187.41

Table 3: Posterior summary of the posterior distribution of $\boldsymbol{\theta} = [\boldsymbol{\beta}, \sigma^2, \tau^2, \phi]$.

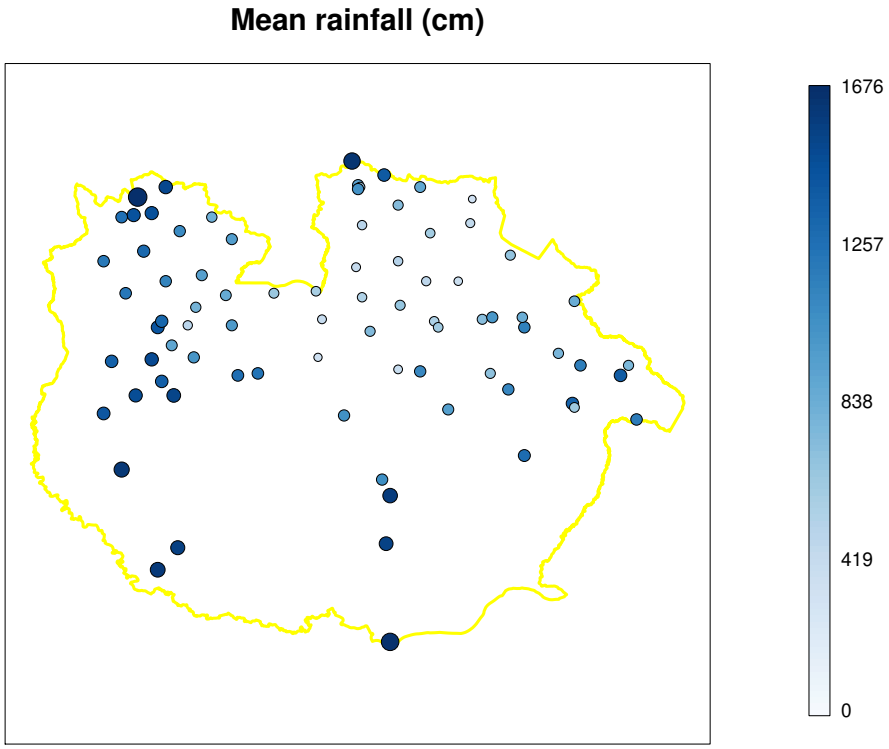


Figure 9: MCMC trace plots (left) and Monte Carlo approximations to the posterior densities (right) of σ^2, τ^2, and ϕ.

"maps" of spatial variables. We plot the corresponding contour and perspective plots side by side as shown in Figure 10.

4. Dynamic spatiotemporal modeling

The approach adopted here applies to the setting where space is viewed as continuous, but time is taken to be discrete. We view the data as a time series of spatial process realizations and work in the setting of dynamic models, building upon previous work by West and Harrison (2006) and **?**.

Let $\boldsymbol{y}_t(\boldsymbol{s})$ be the response at location $\boldsymbol{s}$ and time t. Following Banerjee et al. (2014, Sec. 4, p. 344), we model $\boldsymbol{y}_t(\boldsymbol{s})$ using a *measurement equation* that includes a regression specification with a space-time varying intercept, along with serially and spatially uncorrelated zero-centered Gaussian disturbances as measurement errors, denoted by $\epsilon_t(\boldsymbol{s})$.

Next, a *transition equation* introduces a p-dimensional coefficient vector $\boldsymbol{\beta}_t$, which

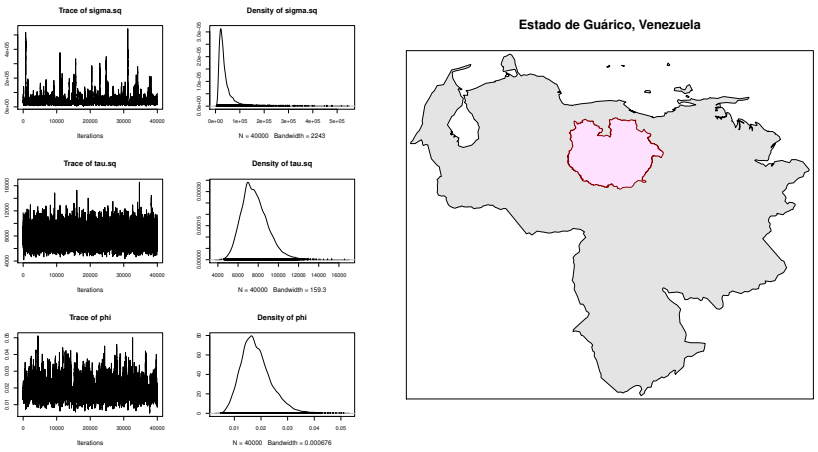


Figure 10: Interpolation of the mean of the random spatial effects in the hierarchical model using multilevel B-splines.

represents purely temporal components (i.e., time-varying regression parameters), along with a spatiotemporal component $u_t(\boldsymbol{s})$. Both terms are generated through transition equations, capturing their Markovian dependence in time. The overall model is written as

$$\begin{aligned} \boldsymbol{y}_t(\boldsymbol{s}) &= \boldsymbol{x}_t(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\beta}_t + u_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s})\,, & \epsilon_t(\boldsymbol{s}) &\overset{\text{ind}}{\sim} \mathsf{N}(0,\tau_t^2)\,,\\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t\,, & \boldsymbol{\eta}_t &\overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\eta}})\,,\\ u_t(\boldsymbol{s}) &= u_{t-1}(\boldsymbol{s}) + w_t(\boldsymbol{s}), & w_t(\boldsymbol{s}) &\overset{\text{ind}}{\sim} \mathsf{GP}(0, C_t(\cdot,\cdot;\boldsymbol{\theta}_t))\,, \end{aligned}$$

for $t = 1, \ldots, N_t$. For each time t, $\boldsymbol{x}_t(\boldsymbol{s})$ is a $p\times 1$ vector of predictors and $\boldsymbol{\beta}_t$ is a $p\times 1$ vector of regression parameters. The $\mathsf{GP}(0, C_t(\cdot,\cdot;\boldsymbol{\theta}_t))$ is a spatial Gaussian process with covariance function $C_t(\boldsymbol{s}_1,\boldsymbol{s}_2;\boldsymbol{\theta}_t) = \sigma_t^2\rho(\boldsymbol{s}_1,\boldsymbol{s}_2;\phi_t)$, where $\boldsymbol{\theta}_t = [\sigma_t^2,\phi_t]$ and $\rho(\cdot;\phi)$ is a valid correlation function –the Matérn correlation function with smoothness parameter $\nu = 2.5$ in this case. We also assume that $\boldsymbol{\beta}_0 \sim \mathsf{N}(\mathbf{m}_0,\boldsymbol{\Sigma}_0)$ and $u_0(\boldsymbol{s}) \equiv 0$, completing the prior specification for a well-identified Bayesian hierarchical model that yields reasonable dependence structures.

We also assume that the same locations are monitored at each time point, resulting in a space-time matrix where the rows index the locations and the columns index the time points, i.e. the (i,j)-th element is $\boldsymbol{y}_j(\boldsymbol{s}_i)$. The algorithm described in Banerjee et al. (2014, Sec. 5, p. 352) will accommodate the situation where some cells of the space-time data matrix may have *missing observations*, as is common in monitoring environmental variables.

We now consider again the Guárico data, which consisting of accumulated monthly rainfall (in centimeters) from $n = 80$ different stations in a time span of $N_t = 16$ years from 1968 to 1983. Figure 11 the interpolation of rainfall using multilevel B-splines in a time span of $N_t = 16$ years from 1968 to 1983. We also take into

account the same covariates as in Section 3 and the estimates of the parameters given there in order to initialize the algorithms implemented here. Even though the original dataset contains 190 missing values (out of 1280 observations, i.e., 14.8 % of the data points are missing), we set 8 (randomly selected) observations from station 5 (which has complete information) to NA in order to illustrate the predictive capabilities of this modeling framework. The true values of the holdout observations are retained for subsequent comparison.

Next, we specify appropriate priors for the model parameters. The variogram analysis conducted in Section 3 guides the specification of priors and hyperparameters, following a similar approach as detailed in Section 3. The MCMC treatment applied in this phase is also analogous to that used in the aforementioned section.

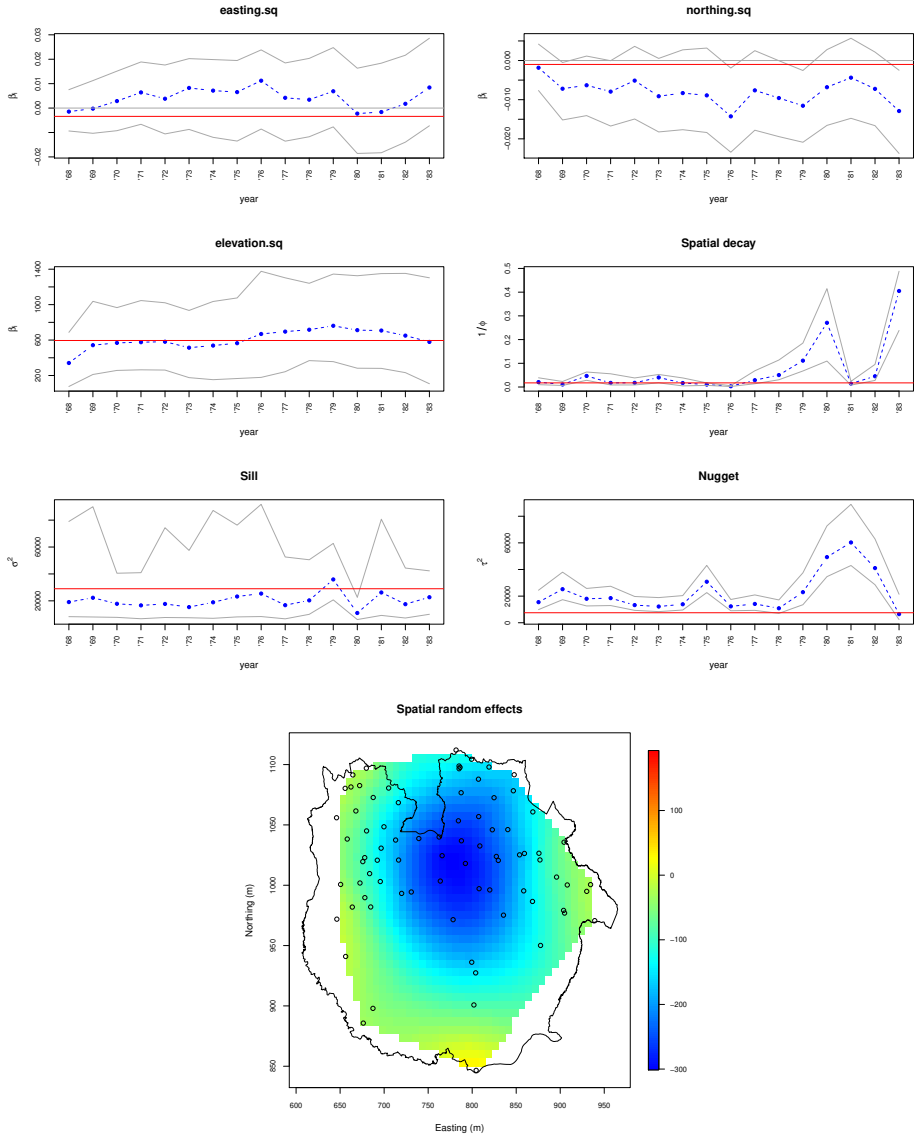
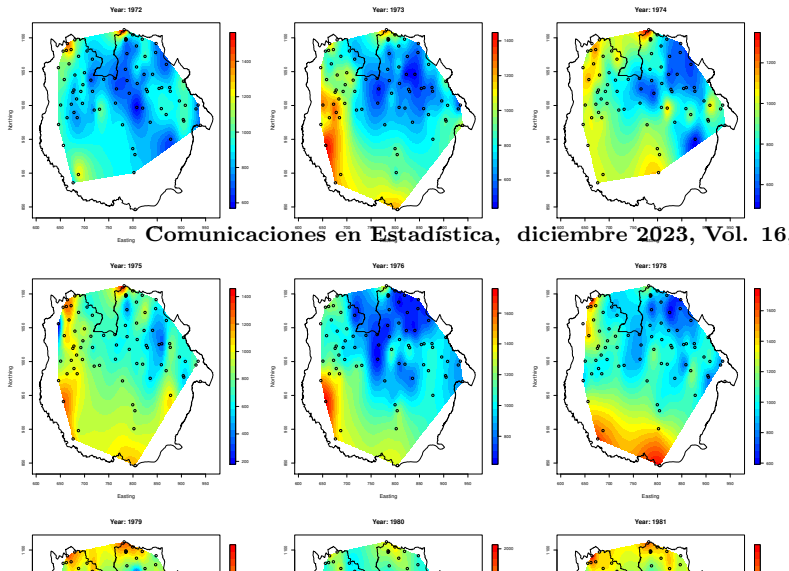


Figure 12: Posterior distribution medians (in blue) and 95 % credible intervals (in gray) for model intercept and predictors. The line in red corresponds to the estimate of the parameter in the static model.



Time series plots of posterior summary statistics for parameters are often useful for exploring their temporal evolution. For regression coefficients, these plots depict the time-varying trends in the response and the effects of covariates. Figure 12 displays the evolution for all the parameters in $\boldsymbol{\beta}$. In each frame of this Figure, we have also included 95 % credible intervals along with a red line corresponding to the estimate of the respective parameter in the static model.

Figure 13 shows the the evolution for σ^2, $1/\phi$, and τ^2. There are noticeable trends in the variance components over time. In addition, we see that the value of the spatial decay remains almost constant and equal to the value in the static model, but it has considerable increments in 1978, 1980, and 1983.

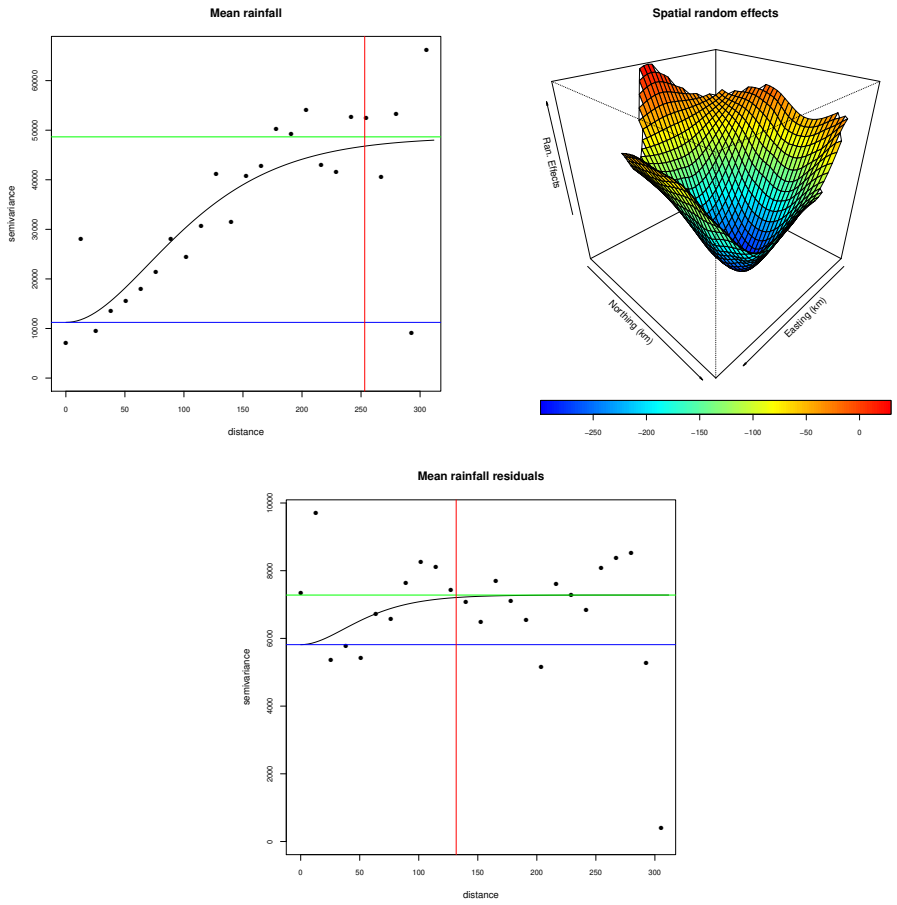


Figure 13: Posterior distribution medians (in blue) and 95 % credible intervals (in gray) for $\boldsymbol{\theta}$ and τ^2. The line in red corresponds to the estimate of the parameter in the static model.

Finally, in order to evaluate the goodness-of-fit and the predictive capacity of the

model, we plot the fitted values against the observed values, and the predictive values against the observed values of the randomly removed observations in station 5. These plots are provided in Figure 14. From the left panel of this Figure we see that the model fits reasonable well to the data. Only six (0.55 % of the observed values) credible intervals corresponding to observations with the lowest rainfalls miss the actual observed value. Furthermore, from the right panel of the Figure, we conclude that the predictive capacity of the model is rather promising since all the 95 % credible intervals of the observations being tested contain the true value of the rainfall.

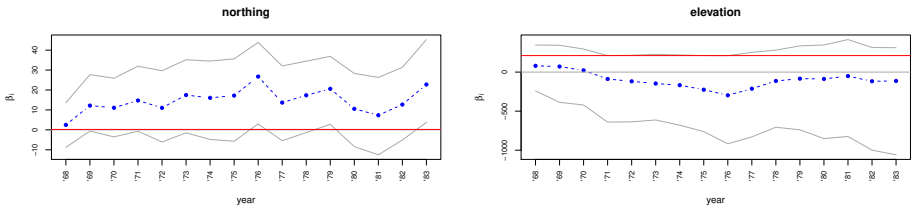


Figure 14: Fitted values (left) and holdout predicted values (right).

5. Discussion

There are many different kinds of spatiotemporal data and extensive statistical literature that addresses most common scenarios. The dynamic strategy elaborated here provides a rather flexible approach to characterize spatiotemporal settings, like this one, in which the time-varying evolution of rainfall is fully described in terms of the possibly changing impact of the covariates conforming the linear trend.

The fully Bayesian approach considered here allows us to handle properly missing values and to validate the predictive capabilities of the model. Other possible strategies include dynamic frameworks to model residual spatial and temporal dependence. These proposed frameworks are flexible and easily extended to accommodate non-stationary and multivariate outcomes(Banerjee et al., 2014, Sec. 11, p. 329).

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this art

Recibido: Octubre 10 de 2023
Aceptado: Diciembre 10 de 2023

Referencias

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2014.

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

P. J. Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press, 2013.

P. J. Diggle and P. J. Ribeiro. Model-based geostatistics. *Model-based Geostatistics*, 2007.

A. O. Finley, S. Banerjee, and B. P. Carlin. spbayes: an r package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of statistical software*, 19(4):1, 2007.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.

A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

P. D. Hoff. *A first course in Bayesian statistical methods*, volume 580. Springer, 2009.

J. T. Kent and K. V. Mardia. *Spatial analysis*, volume 72. John Wiley & Sons, 2022.

A. B. Lawson. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC, 2018.

M. Loecher. Rgooglemaps: Overlays on google map tiles in r. volume 1, 08 2012.

T. J. Oyana. *Spatial analysis with r: statistics, visualization, and computational methods*. CRC press, 2020.

R. R Core Team et al. R: A language and environment for statistical computing, 2013.

C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2006.

B. J. Reich and S. K. Ghosh. *Bayesian statistical methods*. Chapman and Hall/CRC, 2019.

C. P. Robert, G. Casella, and G. Casella. *Monte Carlo statistical methods*, volume 2. Springer, 2004.

S. Sahu. *Bayesian modeling of spatio-temporal data with R.* Chapman and Hall/CRC, 2022.

M. L. Stein. *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media, 2012.

M. West and J. Harrison. *Bayesian forecasting and dynamic models.* Springer Science & Business Media, 2006.

C. K. Wikle, A. Zammit-Mangion, and N. Cressie. *Spatio-temporal statistics with R.* Chapman and Hall/CRC, 2019.

C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

A. Notation

The cardinality of a set A is denoted by $|A|$. If P is a logical proposition, then $\mathbf{1}\{\mathrm{P}\} = 1$ if P is true, and $\mathbf{1}\{\mathrm{P}\} = 0$ if P is false. $\lfloor x \rfloor$ denotes the floor of x, whereas $[n]$ denotes the set of all integers from 1 to n, i.e., $\{1, \ldots, n\}$. The Gamma function is given by $\Gamma(x) = \int_0^\infty u^{x-1}\, e^{-u}\, \mathrm{d}u$. Matrices and vectors with entries consisting of subscripted variables are denoted by a boldfaced version of the letter for that variable. For example, $\boldsymbol{x} = (x_1, \ldots, x_n)$ denotes an $n \times 1$ column vector with entries $x_1, \ldots, x_n$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the column vector with all entries equal to 0 and 1, respectively, and $\mathbf{I}$ to denote the identity matrix. A subindex in this context refers to the corresponding dimension; for instance, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The transpose of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^\mathsf{T}$; analogously for matrices. Moreover, if $\mathbf{X}$ is a square matrix, we use $\mathrm{tr}(\mathbf{X})$ to denote its trace and $\mathbf{X}^{-1}$ to denote its inverse. The norm of $\boldsymbol{x}$, given by $\sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{x}}$, is denoted by $\|\,\boldsymbol{x}\|$.

Now, we present the form of some standard probability distributions used in this article:

- Multivariate normal:

  A $d \times 1$ random vector $\boldsymbol{X} = (X_1 \ldots, X_d)$ has a multivariate Normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its density function is

  $$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-d/2}\, |\boldsymbol{\Sigma}|^{-1/2}\, \exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

- Inverse Gamma:

  A random variable X has an Inverse Gamma distribution with parameters $\alpha, \beta > 0$, denoted by $X \mid \alpha, \beta \sim \mathsf{IG}(\alpha, \beta)$, if its density function is

  $$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{-(\alpha+1)}\, \exp\{-\beta/x\}, \quad x > 0\,.$$