Spatial Statistical Analysis of Rainfall Patterns

Matteo Vantaggio

Statistical Methods and Applications

Professor: Giovanna Jona Lasinio

01-07-2024

Contents

List of Figures

List of Tables

Abstract

This report aims to analyze and model rainfall data using different geostatistical techniques. The objective is to interpolate rainfall measurements over a geographic region by employing methods that account for spatial dependencies and variations in the data, as well as to assess which one is best suited given the context. The report begins with an introduction to the geographical context and a theoretical background on which the statistical techniques to be used are based. It then proceeds with an Exploratory Data Analysis (EDA) to understand both non-spatial and spatial characteristics of the data, including variogram and trend analysis. Following this, the report covers the Inverse Distance Weighting (IDW) technique, discussing its implementation, parameter selection, and results. The focus then shifts to Kriging, presenting both classical and Bayesian approaches, detailing their estimation and prediction methods, and comparing the precision of the estimates.

Chapter 1

Introduction

1.1 Geographical overview

Calabria and Basilicata, regions located in the southern part of Italy, are characterized by diverse and complex rainfall patterns influenced by both geographical and atmospheric factors. These regions experience a Mediterranean climate with hot, dry summers and mild, wet winters, which significantly impacts their rainfall distribution.

Calabria, with its oblong shape and a coastline extending over 738 kilometers along the Ionian and Tyrrhenian seas, exhibits substantial spatial variability in rainfall. The western side of the region, influenced by moist air masses from the Tyrrhenian Sea, typically receives more rainfall compared to the eastern side, which is shielded by mountain ranges. The average annual rainfall in Calabria is about 1150 mm, but this can vary greatly depending on the specific location within the region. Heavy rainfall events in Calabria are often associated with synoptic weather systems, convective storms, and occasionally Mediterranean tropical-like cyclones (Medicanes), leading to intense and localized precipitation.

Basilicata, bordering Calabria to the north, also demonstrates notable rainfall variability. The region's rainfall patterns are influenced by broader climatic phenomena such as the North Atlantic Oscillation, as well as local geographic features. Rainfall in Basilicata tends to be higher in the mountainous areas, particularly in the Apennines, with a notable decrease in the plains and coastal areas. Historical data indicates that extreme rainfall events have become more frequent and intense, contributing to a higher incidence of floods and landslides in these regions.

Rainfall in both regions varies significantly based on elevation and geographical location. For instance, higher elevations tend to receive more rainfall due to orographic lift, where moist air rises over mountains and cools, leading to precipitation. Conversely, lower elevations and plains often experience less rainfall.

In this context, spatial statistics and geostatistical methods are essential tools for analyzing and predicting rainfall distribution in these regions. By incorporating geographical and climatic factors, these methods enable a more comprehensive understanding of the spatial variability in rainfall patterns, which is crucial for effective water resource management, agricultural planning, and disaster preparedness in Calabria and Basilicata.

1.2 Theoretical background

The fondamental tool underlying the application of the statistical techniques which will be discussed in this report is the $random field$

$$\{Y(s) : s \in D\}$$

where $D \subset \mathbb{R}^2$ in our case study. The data consists of a partial realization of the random field, namely measurements at a finite set of locations $s = \{s_1, s_2, ..., s_n\}$.

The basic setup includes the locations and the corresponding measurements, but it can be expanded to include possible *covariates* $\{X(s) : s \in D\}$ available at any location within D.

Our task is to make inference about $Y(s)$ and predict at unobserved locations, given the available data.

In general we deal with processes having the property of **second order stationarity** (also called weak stationarity) which translates into two assumptions:

- the mean of the process is constant over the (measured) surface;
- the covariance $C(Y(s), Y(s+u))$ depends only on the separation vector u and not on the spatial location s.

The second assumption tells us that $C(u) = C(Y(s+u), Y(s))$ because the spatial location s does not influence the relationship between two observations.

The covariance function must be **symmetric** and **positive-definite**

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(Y(s_i), Y(s_j)) \geq 0$$

for any finite set of locations s and any real numbers $\{a_1, a_2, ..., a_n\}$.

On top of this, it is convenient to introduce here the definition of **isotropy**: a spatial process is **isotropic** if the covariance between values at two locations depends only on the distance between the locations, namely if $C(s, s') = C(u)$ with $u = ||s - s'||$.

To assess the degree of spatial correlation of the Gaussian process, the **variogram** is employed.

Assuming that $\mathbb{E}[(Y(s) - Y(s+u)] = 0$, the semivariogram is defined as a function of the spatial separation (lag distance) between pairs of observations:

$$\gamma(u) = \frac{1}{2}\mathbb{E}[(Y(s) - Y(s+u))^2] \tag{1.1}$$

where $Y(s)$ is the value of the RF at location s and u is the lag distance between pairs of points.

After basic algebra we can define the link between variograms and covariances. Note that

$$\gamma(u) = \frac{1}{2}[\mathbb{E}[Y(s)^2] + \mathbb{E}[Y(s+u)^2] - 2\mathbb{E}[Y(s)Y(s+u)]]. \tag{1.2}$$

Assuming that $Y(s)$ is weakly stationary and isotropic[1], then:

$$\mathbb{E}[Y(s)^2] = \mathbb{E}[Y(s+u)^2] = C(0) \tag{1.3}$$

[1]By assuming stationarity and isotropy, the covariance function simplifies significantly. Instead of needing a unique covariance value for every pair of locations, we only need to consider the distance between locations.

and

$$\mathbb{E}[(Y(s)Y(s+u))] = C(u). \tag{1.4}$$

Substituting these into the semivariogram expression:

$$\gamma(u) = \frac{1}{2}[2C(0) - 2C(u)] = C(0) - C(u). \tag{1.5}$$

Moving from the RF to the actual data, an **empirical variogram** can be defined as

$$\hat{\gamma}(u) = \frac{1}{2|N(u)|} \sum_{(i,j)\in N(u)} |y_i - y_j|^2 \tag{1.6}$$

where $N(u)$ is the set of pairs i, j, and $|N(u)|$ is the number of pairs in the set.

The covariance functions addressed in this report are the following

- **Spherical**:

$$C(u) = \begin{cases} 0 & \text{if } u \geq \frac{1}{\phi} \\ \sigma^2\left(1 - \frac{3}{2}\phi u + \frac{1}{2}(\phi u)^3\right) & \text{if } 0 < u \leq \frac{1}{\phi} \\ \tau^2 + \sigma^2 & \text{if } u = 0 \end{cases}$$

- **Exponential**

$$C(u) = \begin{cases} \sigma^2 \exp(-\phi u) & \text{if } u > 0 \\ \tau^2 + \sigma^2 & \text{if } u = 0 \end{cases}$$

- **Matérn**:

$$C(u) = \begin{cases} \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\phi u)^\nu K_\nu(\phi u) & \text{if } u > 0 \\ \tau^2 + \sigma^2 & \text{if } u = 0 \end{cases}$$

where ϕ, τ^2 and σ^2 are also referred to as *covariance parameters*[2]:

- ϕ, the **range parameter**, determining the decay of the covariance function.
- τ^2, the **nugget effect**, representing measurement error or small-scale spatial variation.
- σ^2, the **partial sill**, the asymptotic value of the covariance function at large distances.

[2]For the Matérn covariance function there is also the smoothing parameter ν.

Chapter 2

Exploratory Data Analysis

This chapter delves into the exploratory data analysis (EDA) of the dataset, which includes rainfall measurements across Southern Italy, specifically in the regions of Basilicata and Calabria. The data spans over 12 years (from 1970 to 1982) and includes coordinates in UTM and Longitude and Latitude formats. The primary objective of this analysis is to understand the distribution, trends, and spatial characteristics of the rainfall data before proceeding to more advanced spatial modeling techniques. In the following we will focus on the year 1979.

2.1 Non spatial EDA

We begin with a summary of the dataset to get a general sense of its structure and contents. The dataset comprises 181 data points, detailing regions, geographic coordinates, elevation, and the annual total rainfall. Table 2.1 provides an explanation of the covariates included in the dataset.

Covariate	Description
ID	A unique identifier for each data entry.
Nome_Regione	The name of the region where the data was collected.
Lat10	Latitude of the location, in decimal degrees.
Lon10	Longitude of the location, in decimal degrees.
YUTM	UTM (Universal Transverse Mercator) northing coordinate.
XUTM	UTM (Universal Transverse Mercator) easting coordinate.
Elevation	The elevation of the location above sea level, measured in meters.
Rainfall amount	The total amount of rainfall recorded at the location, measured in millimeters.

Table 2.1: Description of the covariates included in the dataset.

To address our task, we need to consider the response variable *Rainfall amount*, along with the geographical coordinates and *Elevation*. While we assume that rainfall varies with geographical coordinates, we must also consider the relationship between rainfall and elevation. Therefore, we first examine whether there is a correlation between rainfall amounts and elevation.

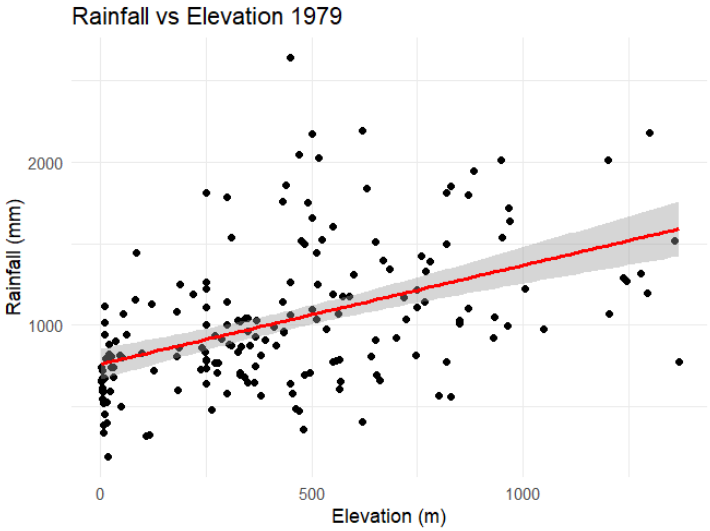

Figure 2.1: Rainfall amount measured in mm as a function of elevation of rain gauge measured in m

By fitting a regression line, we observe a positive linear relationship between the rainfall amount and elevation, further confirmed by the correlation coefficient in table 2.2

Covariate	**Correlation Coefficient**
Elevation	0.4638229

Table 2.2: Correlation coefficient between elevation and rainfall amount.

From a spatial perspective, figure 2.2 illustrates the spatial distribution of rainfall values along with elevation information.

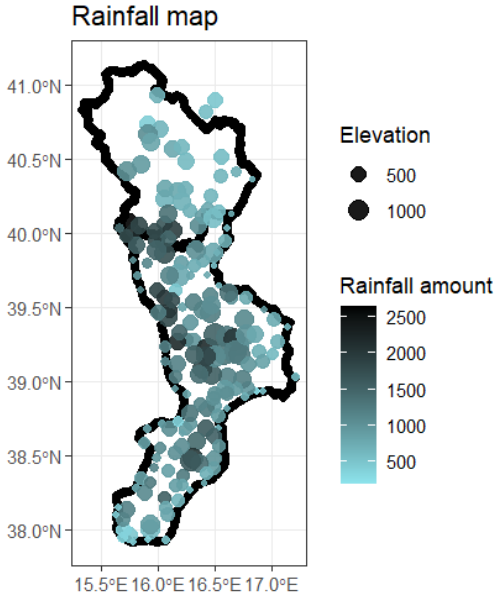

Figure 2.2: Rainfall map showing the intensity of rainfall with color coding. Circle size denotes elevation, with larger circles representing higher elevations.

The rainfall map reveals significant spatial variability in rainfall amounts. The color gradient indicates varying rainfall intensities, with darker colors representing higher rainfall amounts. The map shows that higher rainfall tends to be concentrated in certain areas, particularly towards the left and center of the region.

The circle sizes in the map correspond to elevation levels, with larger circles indicating higher elevations. This dual representation allows us to observe the relationship between elevation and rainfall. Generally, areas with higher elevations, depicted by larger circles, tend to receive more rainfall, as indicated by the darker shades. Conversely, lower elevations, shown by smaller circles, often correspond to areas with lighter shades, indicating lesser rainfall.

One notable aspect of the map is the small area on the top left that lacks any recorded rainfall measurements. This absence of data poses challenges for interpolation, as the model will need to estimate rainfall values without actual observations in that region, potentially leading to inaccuracies.

Another important aspect to investigate is the data distribution, specifically the hypothesis that the data follow a Gaussian distribution. A preliminary and straightforward method to check this is by performing a Shapiro-Wilk test. The result of the Shapiro-Wilk test in table 2.3 confirms that the rainfall data is not normally distributed.

Test	**Statistic (W)**	**p-value**
Shapiro-Wilk normality test	0.94075	8.327e-07

Table 2.3: Results of the Shapiro-Wilk normality test for rainfall amounts.

To further illustrate the distribution's asymmetry, we can construct a histogram, which reveals right-skewness of 0.9366826 (number derived from the computation of the skewness index).

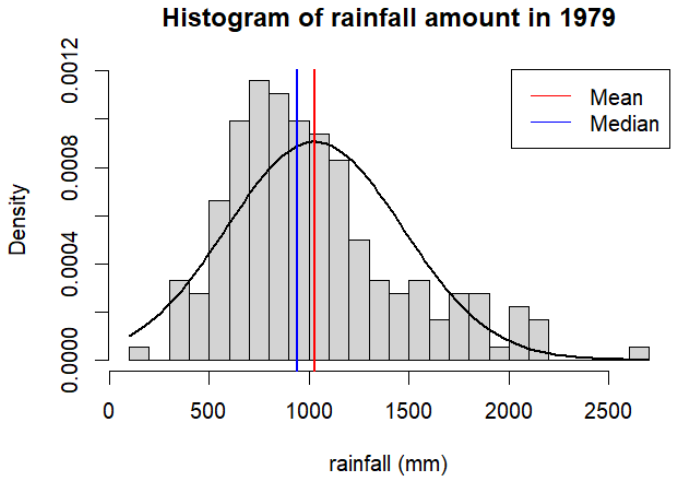

Figure 2.3: Histogram of the data.

In general, the stochastic variation in a physical quantity is often not adequately captured by a Gaussian distribution. One straightforward approach to enhance the Gaussian model is to consider transformations that can render the data more normally distributed.

For response variables that are positive in nature, a particularly useful class of transformations is the Box-Cox family:

$$Y^* = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log Y & \text{if } \lambda = 0 \end{cases}$$

Here, Y represents the original response variable, Y^* denotes the transformed variable, and λ is a parameter that determines the type of transformation applied. When $\lambda = 0$, the transformation simplifies to a logarithmic form.

For the available rainfall data, we used a function that computes profile log-likelihoods for different values of lambda in the Box-Cox power family of transformations[1].

The optimal λ is 0.2.

2.2 Spatial EDA

Given what we said about the geography of the phenomenon under study and the available data plotted in 2.2, we believe a trend is present.

2.2.1 Trend Analysis

In practical applications, it is often necessary to transform data to make it more Gaussian and to account for departures from stationarity, allowing us to apply statistical methods that rely on these assumptions. For the former, as we have seen, the Box-Cox transformation is very useful. For the latter, a common tool used is trend analysis.

The main purpose of spatial trend analysis is to decompose a regional variable into subcomponents such as trends and stochastic error terms.

The most widely used procedure is the global trend surface search. For this task, the spatial variable is approached by a polynomial expansion of the geographic coordinates. The coefficients of the polynomial function are estimated from available measurements using the least squares method, which minimizes the sum of the squared deviations from the trend surface. After identifying the trend surface, the sum of the trend surface value at a site and the residual is equal to the measurement value

$$Y(s) = m(s) + \epsilon(s). \tag{2.1}$$

An interesting type of spatial trend arises when the mean function can be modeled using spatially referenced covariates. In our case, we have elevation data for the recorded rainfall measurements, enabling us to incorporate it into the trend model. This approach helps to explain, rather than merely describe, the spatial variation in the response variable. Hence for the rainfall data, we model the trend as a polynomial function of the three available coordinates (say, $x = XUTM$, $w = YUTM$, $z = Elevation$).

The first order trend can be expressed as

$$m(s) = f(x, w, z) = b_0 + b_1 x + b_2 w + b_3 z. \tag{2.2}$$

[1]By varying lambda over a range of values, the function identifies the lambda that maximizes the log-likelihood, indicating the optimal transformation that makes the data most normally distributed.

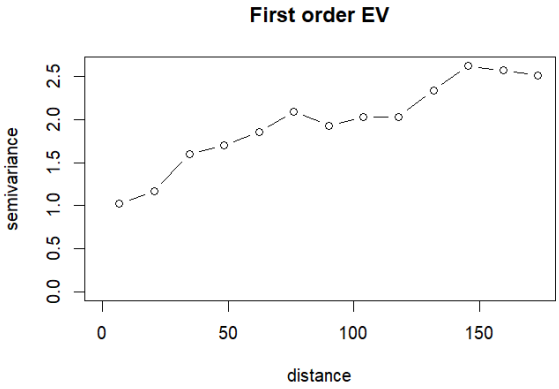

Figure 2.4: Empirical variogram when a first order trend is assumed.

The second-order trend model can be expressed as

$$\begin{aligned} m(s) = f(x, w, z) = & b_0 + b_1 x + b_2 w + b_3 z + b_4 x^2 \\ & + b_5 w^2 + b_6 z^2 + b_7 xw \\ & + b_8 xz + b_9 wz. \end{aligned} \tag{2.3}$$

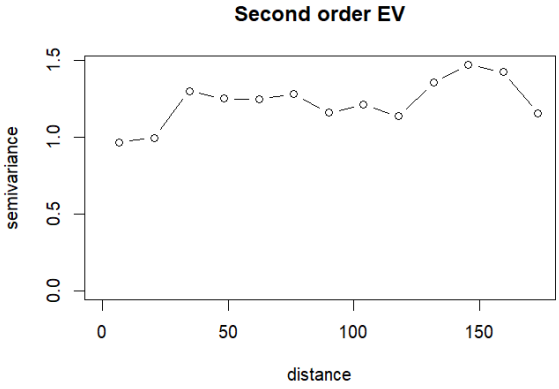

Figure 2.5: Empirical variogram when second order trend is assumed.

After plotting empirical variograms with both linear and quadratic trends, we observed that the quadratic trend had several issues. It exhibited a very short range, reached the sill too quickly, and was generally too sensitive to small-scale variations.

Therefore, we preferred to use a first order trend surface to model the mean of the random field.

We have then fitted some of the most common theoretical variograms, as spherical, exponential and Matérn. In order to choose among those theoretical variograms, we ran *Leave-one-out Cross Validation*. Leave-One-Out Cross-Validation (LOOCV) is a specific form of cross-validation where, for a dataset with n observations, the model is trained n times.

2.3 Leave-One-Out Cross-Validation

In LOOCV, one data point is removed at a time, and its value is predicted using the remaining data. Then the prediction error $e_i = y_i - \hat{y}_i$ is recorded.

To do so, we create a training subset by excluding the i-th point, fit a variogram model (e.g., exponential, spherical or Matérn) to this subset, and apply kriging to predict y_i at location s_i. This process is repeated for each data point, resulting in a series of errors $\{e_i\}_{i=1}^n$.

The model's accuracy is then assessed using the metric

$$\text{CV} = \frac{1}{n} \sum_{i=1}^{n} \frac{e_i^2}{\sigma_{k[i]}^2},$$

where $\sigma_{k[i]}$ is the predicted standard deviation from kriging, excluding the i-th observation. A CV value of 1 means that, on average, the prediction errors are equal to the predicted standard deviations from kriging. Values greater than 1 indicate underestimation, while values less than 1 indicate overestimation.

Although the Matérn model with a smoothing parameter $\nu = 0.4$ performs comparably to the exponential model, the latter is preferred due to its simplicity, with a CV of 1.001.

Model	**CV score**
Spherical	1.0750501
Exponential	1.0013454
Matern, $\nu = 0.6$	0.9839283
Matern, $\nu = 0.4$	1.0178751

Table 2.4: CV score for different models.

Chapter 3

Inverse Distance Weighting

3.1 Introduction to IDW technique

3.1.1 Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) is an (*exact*) deterministic interpolation method used to estimate unknown values at specific locations based on known values at surrounding locations. All interpolation methods are developed based on the theory that points closer to each other have more correlations and similarities than those farther apart. In the IDW method, it is assumed that the rate of correlation and similarity between neighbors is inversely proportional to the distance between them. This can be defined as a distance-reverse function of every point from neighboring points. In particular, the weight w_i assigned to each known value y_i is inversely proportional to its distance u_i from the interpolation point, raised to a power parameter p:

$$w_i = \frac{1}{u_i^p}$$

where:

- u_i is the distance from the i-th known data point to the point of interest.
- p is a positive real number that controls the significance of distant points in the interpolation.

The interpolated value $\hat{y}(x_0)$ at a location x_0 is then computed as a weighted average of the known values:

$$\hat{y}(x_0) = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

where n is the total number of known data points.

3.2 Choosing the optimal p parameter

To determine the optimal theoretical variogram model for fitting the data, we use a similar cross-validation approach. A grid of power values p is created, ranging from 0 to 16. Then, within a *for loop*, a leave-one-out cross-validation (LOOCV) is conducted to

evaluate each power value. The optimal p value, 3 in this case, is identified as the one that results in the smallest root mean square error (RMSE).

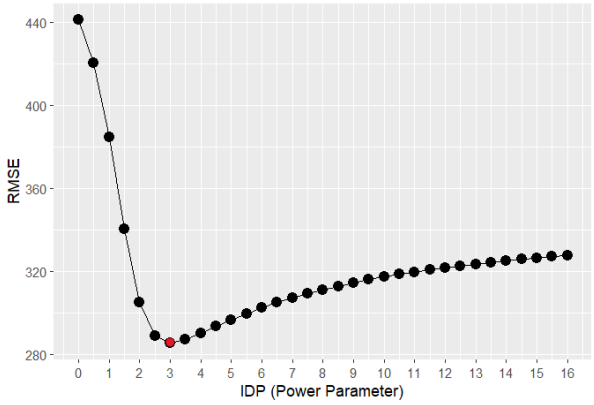

Figure 3.1: Optimal value for the power p marked in red

3.3 Mapping rainfall with IDW

Once the optimal power value p has been selected, we can proceed with the IDW interpolation. This interpolation is performed over a dense grid of location points within the study region, where elevation data has been previously recorded.

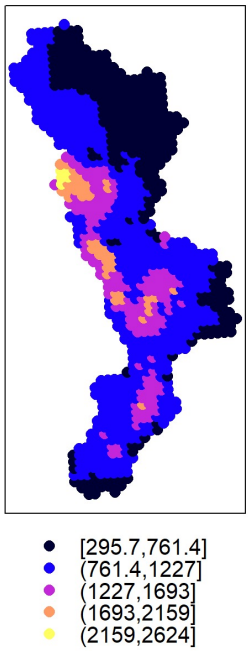

Figure 3.2: IDW interpolated surface

3.4 Comments on the results

The regions marked in yellow, orange and pink indicate areas of the highest rainfall intensity and we can see they are relatively sparse and concentrated in specific zones within the central part of the study area.

While the blue and purple regions represent moderate rainfall intensity, which are more widespread and form a gradient surrounding the high rainfall zones.

The dark blue areas show the lowest rainfall intensity, which are prevalent in the northern and southern extremes of the study area, coherently with what we know about the data.

Inverse Distance Weighting has the advantage that it is very easy to implement and understand, given that it's based on the principle that points closert to the target location have more influence than those further away. Also, the only one parameter we need to control is the power parameter p, making the fine-tuning of the model straightforward.

Ultimately, unlike Kriging, which requires assumptions about the statistical distribution of data and the spatial structure, IDW does not require such assumptions.

However we should mention some of the limitations of this technique.

First off, it is sensitive to the choice of power parameter p and the density of the known data points: sparse data can lead to less reliable interpolations.

Secondly, the output surface is sensitive to clustering and the presence of outliers. Thirdly, it doesn't provide prediction standard errors, which often makes the technique problematic to justify.

Chapter 4

Kriging

4.1 Introduction to Kriging technique

Here we assume that the random vector $Y(s) = (Y(s_1), Y(s_2), ..., Y(s_n))$ depends not only on the covariates $X(s)$ but also on an underlying latent spatial process $W(s)$.

This leads to the following formulation, called **spatial Gaussian mixed model**:

$$Y(s) = X(s)\beta + W(s) + \epsilon(s) \tag{4.1}$$

where $X(s)\beta$[1] models the deterministic part of $Y(s)$ that is explained by the covariates, while $W(s)$ and $\epsilon(s)$ capture the spatially structured random effects and the unstructured random noise, respectively.

In particular, $\epsilon(s)$ models the measurement errors since it is a spatially independent Gaussian process with zero mean vector and variance $\tau^2 I_n$, while $W(s)$ is responsible for the spatial variability since it is a spatially correlated stationary Gaussian process

$$W(s) \sim \mathcal{N}_n\left(0, \sigma^2 H_{11}(\phi)\right)$$

where H_{11} is the correlation matrix. We assume also that $W(s)$ and $\epsilon(s)$ are independent, so that $Y(s)$ is independent and normally distributed given $X(s)$.

Integrating the spatial effect out of 4.1 we get the **marginal model**

$$Y(x) \sim \mathcal{N}\left(X(s)\beta, \Sigma_{11}\right) \tag{4.2}$$

where $\Sigma_{11} = \sigma^2 H_{11}(\phi) + \tau^2 \mathbb{I}_n$.

4.1.1 ML Estimation

Letting $\omega = (\sigma^2, \tau^2, \phi)$ represent the covariance parameters, we obtain the likelihood $L(\beta, \omega; y)$ where y contains the vector of realizations of the RF. Considering Σ_{11} known (conditionally on it), the ML estimator of the mean parameter β is the BLUE and coincides with the usual GLS estimator (writing $X = X(s)$ for short)

$$\hat{\beta} = \left(X'\Sigma_{11}^{-1}X\right)^{-1} X'\Sigma_{11}^{-1}y. \tag{4.3}$$

However, considering Σ_{11} as fixed, the covariance of the estimator of β results to be underestimated.

[1] Usually $X(s)$ is a $n \times p$ matrix and β a $p \times 1$ vector of trend parameters.

An unbiased alternative approach which maximizes a likelihood function based on the residuals $y - X\hat{\beta}$ can be employed. It can be shown that the residuals and $\hat{\beta}$ are independent, so that the likelihood can be factorized into two parts: one that depends on the residuals and one that depends on the fixed effects

$$L(\omega, \beta; y) = L(\omega; y - X\hat{\beta})L(\omega, \beta; \hat{\beta}). \tag{4.4}$$

Hence one can compute $L(\omega; y - X\hat{\beta})$, namely the REML likelihood.

Given the independence between residuals, and fixed effects, $\hat{\beta}$, we can maximize REML likelihood with respect to the covariance parameters and update $\hat{\beta}$ and ω until the estimates become numerically stable.

REML likelihood is preferred for its power to correct for the loss of degrees of freedom due to estimating β.

4.1.2 ML Prediction

Prediction-wise, we consider the same model as in 4.1 also for the m-dimensional vector of unobserved locations x_0[2], namely $Y_0 = X_0\beta + W_0 + \epsilon_0$ and represent the marginal model for observed and unobserved locations as

$$\begin{bmatrix} Y \\ Y_0 \end{bmatrix} \sim \mathcal{N}_{n+m}\left(\begin{bmatrix} X\beta \\ X_0\beta \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{10} \\ \Sigma_{10}' & \Sigma_{00} \end{bmatrix} \right) \tag{4.5}$$

where $\Sigma_{10} = \sigma^2 H_{10}(\phi)$ is the covariance between observed and unobserved locations, Σ_{11} is the already defined covariance between observed locations and $\Sigma_{00} = \sigma^2 H_{00}(\phi) + \tau^2 \mathbb{I}_m$ is the covariance between unobserved locations.

The best linear predictor for this model is the expected value of the predictor conditionally on the observed values

$$\hat{Y}_0 = X_0\hat{\beta} + \Sigma_{10}'\Sigma_{11}^{-1}(y - X\hat{\beta}). \tag{4.6}$$

4.2 Rainfall interpolation with Kriging

Returning to our dataset, we implemented the kriging technique following the specification of the trend surface and the estimation of the covariance matrix using the exponential model. Figure 4.1 shows what the predicted surface looks like

[2]An added assumption has been included here, namely that the random component ϵ_0 at the unobserved location x_0 is independent of both W_0 and ϵ, simplifying Σ_{10}, the covariance between observed and unobserved locations.

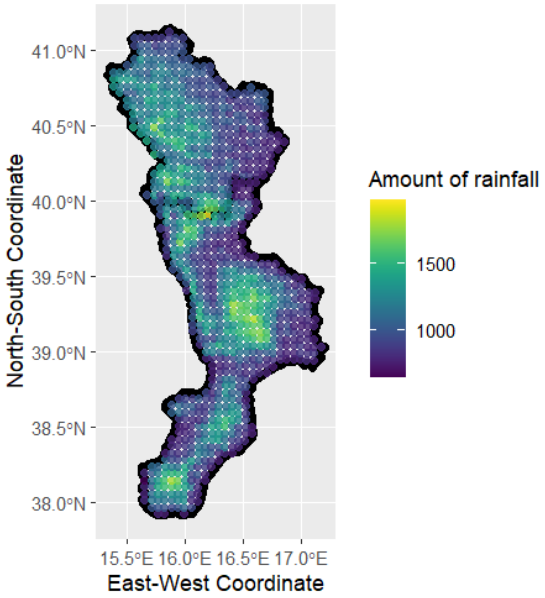

Figure 4.1: Spatial distribution of rainfalls in 1979. Predictions has been made through *Conventional Kriging* and values are mapped to the original scale.

One metric we are going to employ to assess the *accuracy* of the model is the **Root Mean Square Error (RMSE)**, which compare forecasting errors of different models

$$RMSE(\hat{Y}_0) = \sqrt{MSE(\hat{Y}_0)}$$

From this Kriging model we obtained a RMSE of 256.01.

The predicted values accurately capture the assumed spatial trend and reflect the variations present in the original data. However, caution is advised for the upper left part of the map, where no actual data were available, and the Kriging estimator had to rely on extrapolation.

Comments on uncertainty. In a frequentist approach, which we've been using, a parameter is viewed as a fixed yet unknown value to be estimated. This parameter characterizes the underlying processes that produce the data. On the other hand, a prediction target is the observed outcome of a random variable linked to those processes.

Our current methodology involves formulating a model, estimating its parameters, and subsequently using these estimated values in theoretical prediction equations under the assumption that they are exact representations of reality.

However, ignoring the uncertainty inherent in these parameters can lead to overly optimistic evaluations of predictive precision.

This is why, before evaluating the accuracy of the implemented Kriging method using *confidence intervals* and *interval scores*, we believe it would be beneficial to introduce its **Bayesian** counterpart.

Following this, we will proceed with a comparative analysis between the two approaches.

Chapter 5

Bayesian Kriging

In the Bayesian framework, both the response variable Y and parameters β, ω are treated as random quantities, each having their own distribution and sharing together a joint distribution. The likelihood of Y is given by

$$p(y|\beta,\omega) = \int p(y|W,\beta,\tau^2)p(W|\sigma^2,\phi)dW. \tag{5.1}$$

Hence we combine the prior knowledge about parameters in $\pi(\beta,\omega) = \pi(\beta)\pi(\tau^2)\pi(\sigma^2,\phi)$ with 5.1 and obtain the posterior distribution $\pi(\beta,\omega|y)$ through *Bayes theorem*

$$\pi(\beta,\omega|y) = \frac{\int p(y|W,\beta,\tau^2)p(W|\sigma^2,\phi)\pi(\beta)\pi(\tau^2)\pi(\sigma^2,\phi)\,dW}{\int\cdots\int p(y|W,\beta,\tau^2)p(W|\sigma^2,\phi)\pi(\beta)\pi(\tau^2)\pi(\sigma^2,\phi)\,d\beta\,dW\,d\tau^2\,d\sigma^2\,d\phi} \tag{5.2}$$

Bayesian prediction is made through the *predictive distribution*

$$p(y_0|y) = \iint p(y_0|y,\beta,\omega)\pi(\beta,\omega|y)d\beta d\omega$$

which can be regarded as a weighted average of the "plug-in" predictions, on which conventional Kriging predictor's distribution is based on, namely $(y_0|y) \sim p(y_0|y,\hat{\beta},\hat{\omega})$ and the weights are given by the posterior $\pi(\beta,\omega|y)$.

A first considerable difference between conventional Kriging and Bayesian Kriging resides in that the former gives a point estimate of each predicted observation given the maximum likelihood estimate of parameters, whereas the latter gives a distribution of values for each predicted observation.

As with the ML Kriging, we also consider the marginal model here as the main tool for inference and prediction

$$Y|\beta,\phi,\sigma^2,\tau_R^2 \sim \mathcal{N}(X\beta,\sigma^2\Sigma(\phi,\tau_R^2)))$$

although we use the *relative nugget effect* parameterization, $\tau_R^2 = \frac{\tau^2}{\sigma^2}$, because it is more computationally convenient than learning about ϕ, τ^2, and σ^2 simultaneously.

5.1 Implementation of Bayesian Kriging

For the implementation of this technique, we started with a preliminary model using default priors provided by the function *krige.bayes*, although having specified a reasonable support for the parameter τ_R^2 and a uniform prior for it.

This initial step has been made to exploit a first, vague prior to posterior update and utilize the posteriors as priors for the next, more precise model.

In figures 5.1 and 5.2 there are the prior and posterior distributions for ϕ and τ_R^2 respectively.

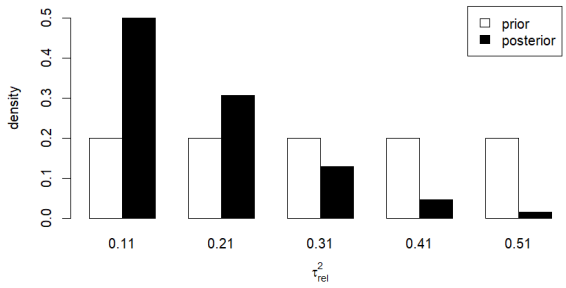

Figure 5.1: (First model) Prior distribution and posterior update of ϕ

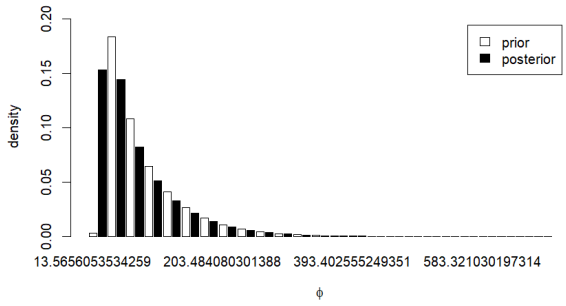

Figure 5.2: (First model) Prior distribution and posterior update of τ_R^2

At this point we fit a second Bayesian model using the posterior updates of the first model's prior distributions[1] and obtained 500 samples from posterior and predictive distributions.

In principle, the prior distribution for ϕ should have continuous support, but in practice we use a discrete prior, obtained by discretising its distribution in equal width intervals. This required to specify the range of the prior for ϕ.

We then did the same for τ_R^2.

For the former, a discrete distribution of 5 points in the interval $[14, 66]$ was placed and a discrete distribution of 5 points in the interval $[0.11, 0.51]$ for the latter.

[1] A flat prior for β was necessary to increase computational efficiency, although its posterior distribution obtained from the first model is Gaussian.

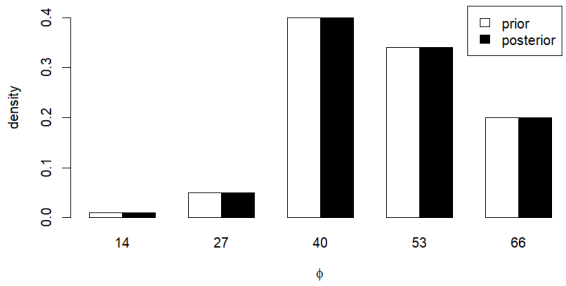

Figure 5.3: (Second model) Prior distribution and posterior update of ϕ

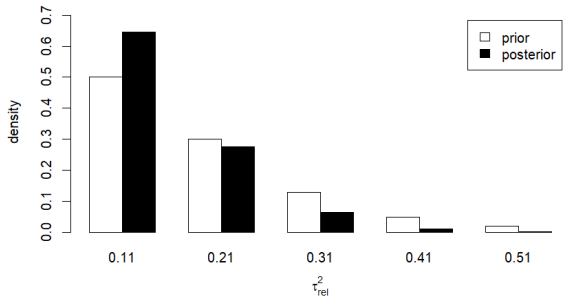

Figure 5.4: (Second model) Prior distribution and posterior update of τ_R^2

After having set up a grid of 1001 points to predict our response, we ran Bayesian Kriging interpolation using the prior specifications in figures 5.3 and 5.4.

Parallel to this routine, we applied the same interpolation method using only the 181 datapoints of rainfall measurements. We constructed a Leave-One-Out Cross-Validation (LOOCV) loop to extract the expected value and the corresponding variance of each predictive distribution.

To

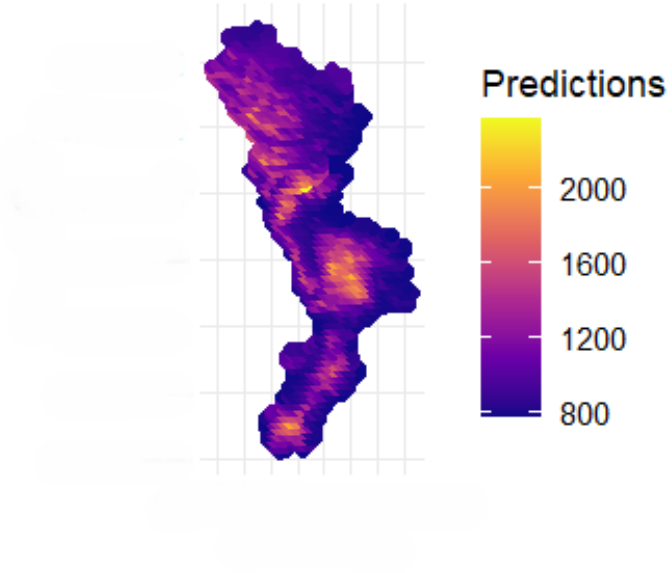

Figure 5.5: Bayesian predictions of rainfall.

At this point we were able to compute RMSE and compare it with the one obtained with the previous model (conventional Kriging). It appears that the two methods are

Model	RMSE
ML Kriging	256.0111
Bayesian Kriring	257.1825

Table 5.1: RMSE for the two models

similar in terms of RMSE. Therefore, we believe that further comparisons are necessary to draw a definitive conclusion.

5.2 Evaluation of estimate precision and comparison with classical Kriging results

In order to conduct a comparison between ML Kriging and Bayesian Kriging, we should introduce one broad class of evaluation metrics for forecasts, namely the **Scoring rules**.

We will focus on one particular subset, the *interval score*.

Just to get the intuition, the interval scores rewards forecasters for issuing narrow prediction intervals but penalizes them if the interval does not cover the actual outcome.

This balance ensures that forecasters provide intervals that are both informative and reliable.

The interval score can be derived from the general formulation of scoring rules for quantile forecasts. For a central $(1-\alpha)\times 100\%$ prediction interval with lower and upper endpoints l and u (the $\alpha/2$ and $1-\alpha/2$ quantiles), the scoring rule for the interval forecast $S_\alpha(l,u;x)$ when the actual outcome is x is defined as:

$$S_\alpha(l,u;x) = (u-l) + \frac{2}{\alpha}(l-x)\cdot \mathbb{I}_{x<l} + \frac{2}{\alpha}(x-u)\cdot \mathbb{I}_{x>u} \tag{5.3}$$

Here, $\mathbb{I}$ is the indicator function, which equals 1 if the condition is true and 0 otherwise.

- $u-l$ rewards the forecaster for providing a narrow prediction interval.
- $\frac{2}{\alpha}(l-x)\cdot \mathbb{I}_{x<l}$ penalizes the forecaster if the actual outcome x is less than the lower bound l of the prediction interval.
- $\frac{2}{\alpha}(x-u)\cdot \mathbb{I}_{x>u}$ penalizes the forecaster if the actual outcome x is greater than the upper bound u of the prediction interval.

$S_\alpha(l,u;x)$ is a proper scoring rule as it incentivizes the forecaster to provide an interval that genuinely reflects their uncertainty about the outcome.

We produced interval proper scores using 5.3 for all points in our dataset and then took the average and the median of them for both models.

For ML Kriging we used the basic 95% confidence intervals, while for Bayesian kriging we built credible intervals using the 2.5th and 97.5th percentiles of the predictive distributions.

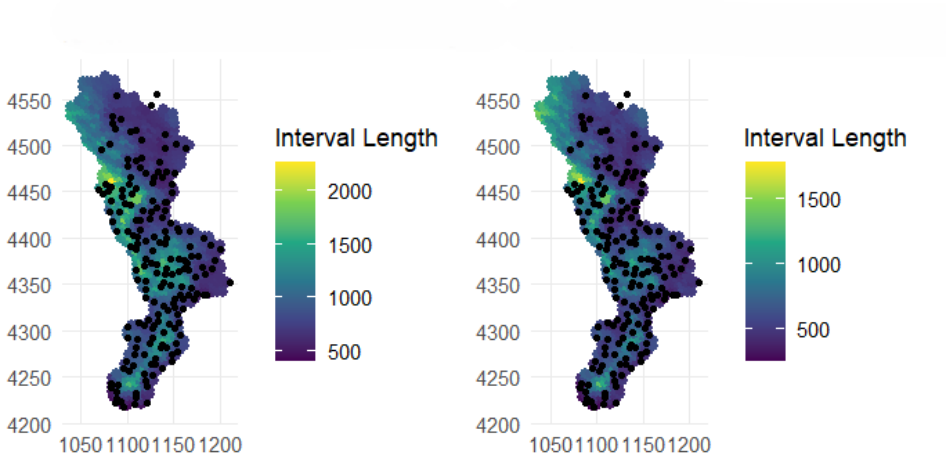

Figure 5.6: The left plot illustrates ML confidence intervals lengths. The right plot displays Bayesian credibility intervals lengths. In both plots, darker colors indicate narrower intervals, and lighter colors signify wider intervals. The black dots represent the actual rainfall data measurements.

Although the two models produce similar results, the right plot in Figure 5.6 (Bayesian intervals) shows that in the central region of the map, the intervals are narrower (darker color) compared to the ML intervals, which appear wider (lighter color). However, in the area where we have no observations, the Bayesian intervals seem wider than the ML intervals, reflecting more accurately the uncertainty due to the lack of measurements.

Further inspections are conducted and from table 5.2 we observe that, while the mean interval score of the Bayesian model is higher than that of the ML model, the median interval score is lower. This suggests that the Bayesian model could potentially be more precise than the maximum likelihood model.

Model	**Mean of Interval Proper Score**	**Median of Interval Proper Score**
ML Kriging	1342.142	929.3842
Bayesian Kriging	1704.788	684.3635

Table 5.2: Average and median interval proper score

Noticeable spikes in Figure 5.7 are attributed to certain observations being penalized significantly more by the Bayesian model compared to the MLE model. This suggests that while the Bayesian model tends to produce narrower intervals, it also incurs higher penalties for predictions that fall outside. In fact, we checked that Bayesian proper score exceeds the MLE proper score in only about 1/6 of the total instances.

Consequently, despite the lower frequency of higher proper scores, the severity of these penalties contributes to a higher overall mean for the Bayesian proper scores compared to the MLE proper scores.

This indicates a trade-off between the interval width and the penalties, suggesting that the Bayesian model's narrower intervals might come at the cost of higher penalties when predictions are off. As a matter of fact, all the Bayesian upper bounds are systematically lower than their ML counterparts, and all the Bayesian lower bounds are higher than those of the ML model.

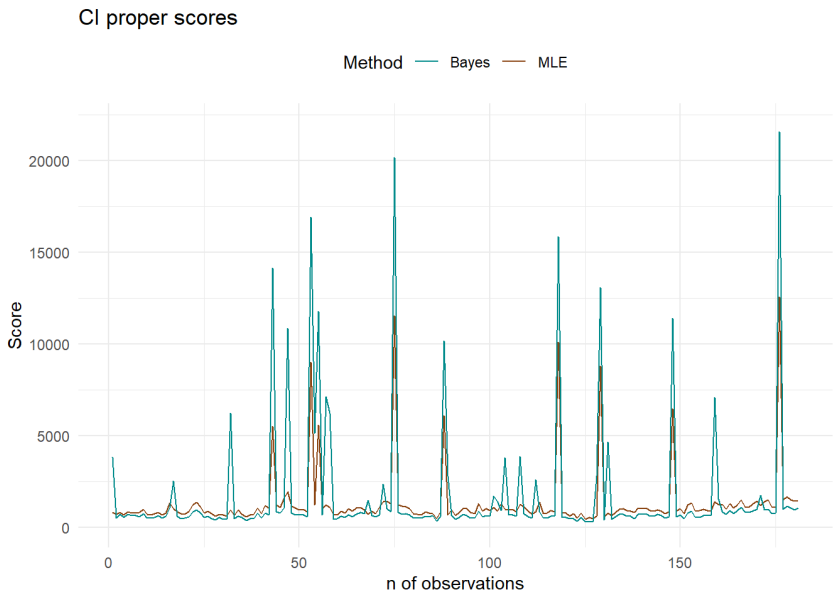

Figure 5.7: Interval Scores. The *x-axis* represents the number of observations, while the *y-axis* indicates the interval proper score.

At this point, we can conclude the analysis of the accuracy of the two models computing the **empirical coverage** which is a measure used to evaluate the performance of prediction intervals.

The empirical coverage is the proportion of times that the true values fall within the prediction intervals out of the total number of intervals considered. Formally, for a set of n observations, let y_i denote the true value and (l_i, u_i) denote the lower and upper bounds of the prediction interval for the i-th observation. The empirical coverage can be defined as:

$$\text{Empirical Coverage} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(l_i \leq y_i \leq u_i).$$

Method	**Empirical Coverage**
MLE	0.9447514
Bayesian	0.8618785

Table 5.3: Empirical Coverage for ML and Bayesian Kriging.

Table 5.3 shows that the MLE method has a higher empirical coverage compared to the Bayesian method, indicating that the prediction intervals from the MLE method are more reliable in containing the true values.

Chapter 6

Conclusions

In conclusion, the analysis revealed that Bayesian Kriging and Maximum Likelihood (ML) Kriging offer distinct advantages and limitations, making the choice between them highly dependent on the specific problem at hand.

Bayesian Kriging demonstrated higher prediction precision by incorporating prior information, enabling a probabilistic framework that effectively quantifies uncertainty. However, this method's precision can also be a drawback if the priors are not chosen carefully, leading to subjectivity and potential biases in the results.

On the other hand, ML Kriging tends to produce less precise predictions but offers larger, more consistent prediction intervals. This method's simplicity and ease of implementation make it a practical choice for many applications, especially when computational resources are limited or when a straightforward approach is preferred. The consistency of ML Kriging's broader intervals can be advantageous in providing a more conservative range of predictions, which can be crucial in risk-averse scenarios.

The study also employed the Inverse Distance Weighting (IDW) method at the beginning as a deterministic approach to interpolate rainfall values. While IDW is straightforward and easy to implement, its deterministic nature implies that it does not account for spatial variability and uncertainty, leading to potential oversimplification of complex spatial patterns. Unlike Kriging methods, IDW does not provide confidence intervals, making it less robust for applications requiring uncertainty quantification.