CLEMSON UNIVERSITY

Lecture 14

State-Space Models and Geostatistics

Reading: SS17 Chapter 6.2-6.4, Chapter 6.12; BD Chapter 9.4-9.7

MATH 8090 Time Series Analysis
Week 14

Whitney Huang
Clemson University

Agenda

1. Kalman Recursions for Filtering, Forecasting, and Smoothing
2. Estimating the State-Space Model Parameters
3. Gaussian Process Spatial Model
4. Spatial Interpolation
5. Parameter estimation
6. A Case Study of Paraná State Precipitation Data

Review: Kalman Recursions for Filtering/Forecasting

1. Compute innovation:

$$U_t = Y_t - Y_t^{t-1} = Y_t - \mu_t^f$$

2. Compute MSE for Y_t^{t-1}:

$$\Sigma_t^f + \sigma_W^2 \stackrel{\text{def}}{=} F_t$$

3. Compute new filtered value:

$$\mu_t^a = \mu_t^f + K_t U_t,$$

where $K_t \stackrel{\text{def}}{=} \Sigma_t^f / F_t$ is the so-called Kalman gain

4. Compute MSE for new filtered value:

$$\Sigma_t^a = \Sigma_t^f (1 - K_t)$$

Kalman Recursions for Filtering/Forecasting (Cont'd)

5. Compute new forecast:

$$\mu_{t+1}^{f} = \mu_t^{f} + K_t U_t = \mu_t^{a}$$

6. Compute MSE for new forecast:

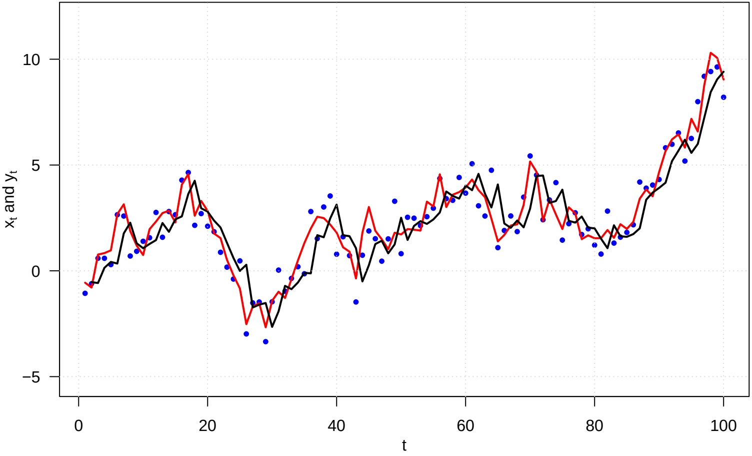
$$\Sigma_{t+1}^{f} = \Sigma_t (1 - K_t) + \sigma_V^2 = \Sigma_t^{a} + \sigma_V^2$$

Recursions are carried out for $t = 0, \cdots, n$ with inputs $\mathrm{E}[X_0] = \mu_0$, $\mathrm{Var}(X_0) = \sigma_0^2$ and $Y_t's$

Simulated Example: Local Level Model with SNR $= 2$

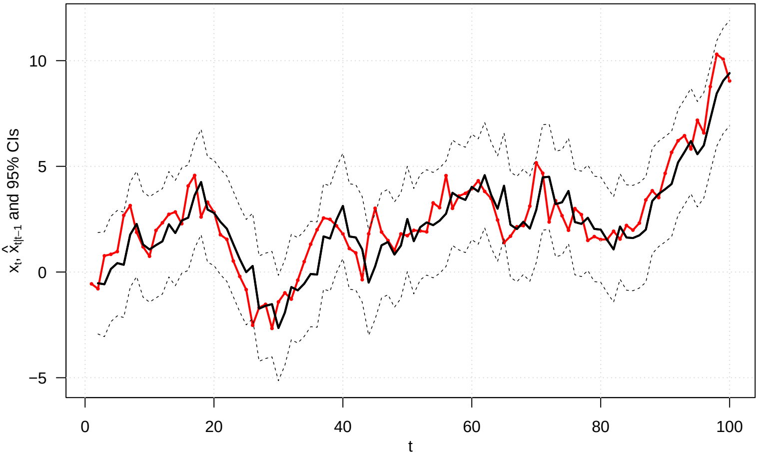
Setup: $\mu_0 = 0$, $\sigma_0^2 = 1 = \sigma_V^2$, $\sigma_W^2 = 0.5$

Time series Y_t, states X_t, and forecasts μ_t^f



Simulated Data from Local Level Model with SNR $= 2$

States X_t, forecasts μ_t^f, and 95% CIs based on Σ_t^f

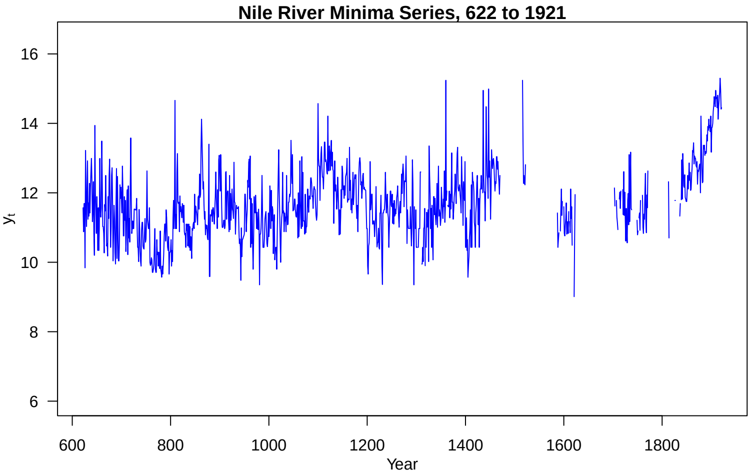


Kalman Recursions for Time Series with Missing Values: I

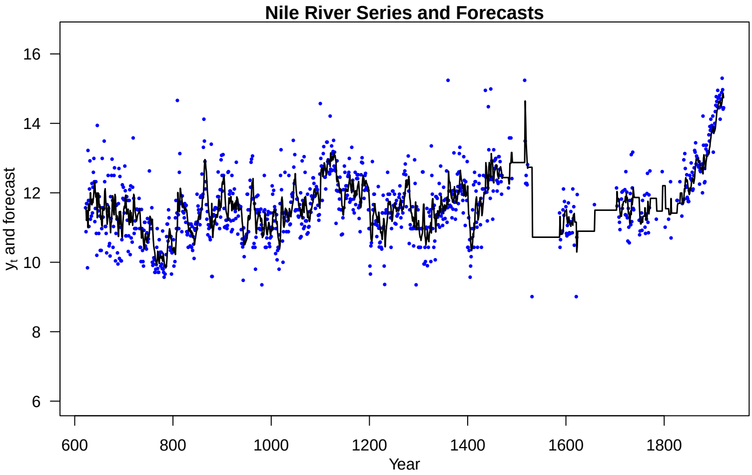
One of the strengths of state-space formulation is the capability to handle time series with missing values. Suppose $Y_1, \cdots, Y_t$ and Y_{t+3} are observed, but not Y_{t+1} and Y_{t+2}:

- use modified recursion (i.e., skip the calculation of the innovation when data is missing)
  - use $\mu^f_{t+1} \stackrel{\text{def}}{=} X^t_{t+1}$ and $\Sigma^f_{t+1} \stackrel{\text{def}}{=} \Sigma^t_{t+1}$ for X^t_{t+2} and Σ^t_{t+2}
  - use X^t_{t+2} and Σ^t_{t+2} for X^t_{t+3} and Σ^t_{t+3}
- take X^t_{t+3}, Σ^t_{t+3}, and Y_{t+3} into usual recursion to obtain $\mu^a_{t+3} = X^{t+3}_{t+3}$ and $\Sigma^a_{t+3} = \Sigma^{t+3}_{t+3}$ and $\mu^f_{t+4} = X^{t+3}_{t+4}$ and $\Sigma^f_{t+4} = \Sigma^{t+3}_{t+4}$
- need to interpret "given $t+3$" as conditioning on everything available at time $t+3$, i.e., $Y_1, \cdots, Y_t$ and Y_{t+3}

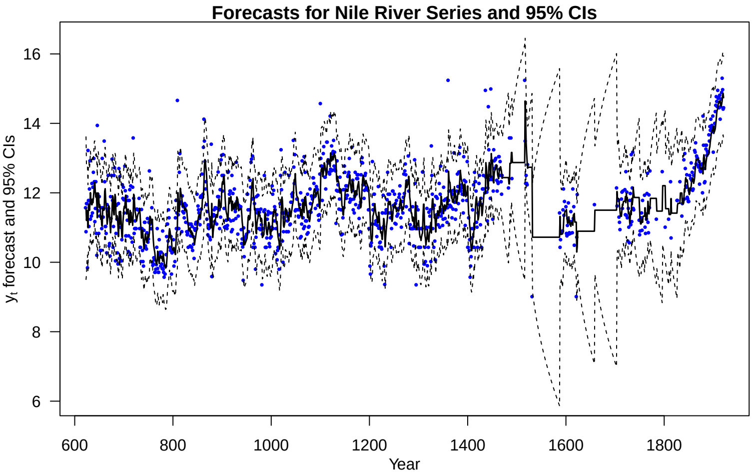
Example: Nile River Annual Minima Series



Nile River Annual Minima Series with Missing Values Imputed



Nile River Annual Minima Series Forecasts with 95 % CI



Kalman Recursions for Smoothing: I

Given time series $Y_1, \cdots, Y_n$, Kalman filter recursions give us $\mu_t^a = X_t^t$ for $t = 1, \cdots, n$

- Regression lemma says solution to smoothing problem is

$$\mu_t^s \stackrel{\text{def}}{=} \mathrm{E}[X_t | \boldsymbol{Y}_{1:n}] = \mu_t + \Sigma_{t,n}^T \Sigma_{Y,n}^{-1} (\boldsymbol{Y}_{1:n} - \boldsymbol{\mu}_{1:n})$$

- MSE for predictor, i.e., $\mathrm{E}\left[(X_t - \mu_t^s)^2\right]$, is

$$\Sigma_t - \Sigma_{t,n}^T \Sigma_{Y,n}^{-1} \Sigma_{t,n} \stackrel{\text{def}}{=} \Sigma_t^s,$$

where $\Sigma_t \stackrel{\text{def}}{=} \mathrm{Var}[X_t]$

Kalman Recursions for Smoothing: II

Using innovation U_t, innovation variance F_t, Kalman gains K_t, forecasts $\mu_t^f \stackrel{\text{def}}{=} X_t^{t-1}$ and associated MSEs $\Sigma_t^f \stackrel{\text{def}}{=} \Sigma_t^{t-1}, t = 1, \cdots, n$ computed by Kalman filter recursions, Kalman smoother recursions allow efficient computation of $\mu_t^s, t = 1, \cdots, n$

The first two steps yield desired predictor μ_t^s

1. Manipulate innovations: starting with $r_n = 0$, recursively form
$$r_{t-1} = \frac{U_t}{F_t} + (1 - K_t) r_t, \quad t = n, \cdots, 1$$
2. Combine manipulated innovations and forecasts:
$$\mu_t^s = X_t^t + \Sigma_t^{t-1} r_{t-1}, \quad t = 1, \cdots, n$$

Kalman Recursions for Smoothing: III

Next two steps yield MSE for predictor X_t^n:

3. Manipulate innovation variances: starting with $N_n = 0$, recursively form

$$N_{t-1} = \frac{1}{F_t} + (1 - K_t)^2 N_t, \quad t = n, \cdots, 1$$

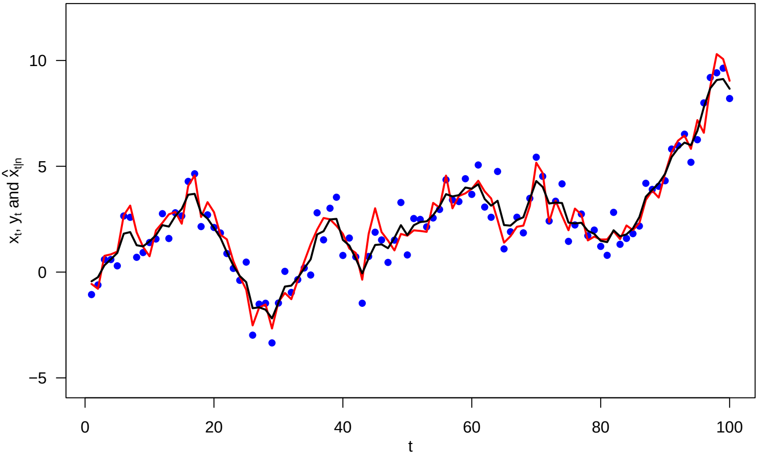
4. Combine manipulated innovation variances and forecast MSEs:

$$\Sigma_t^n = \Sigma_t^{t-1} - \left(\Sigma_t^{t-1}\right)^2 N_{t-1}, \quad t = 1, \cdots, n,$$

where Σ_t^n is the desired MSE

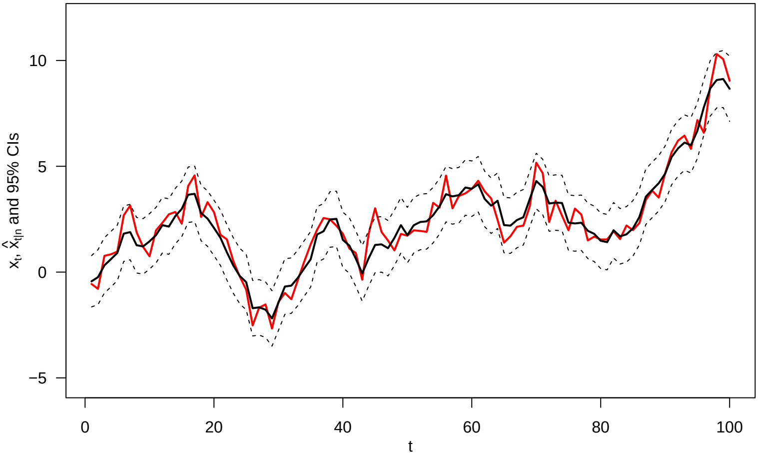
Simulated Example: Local Level Model with SNR $= 2$

Time series Y_t, states X_t, and smooths μ_t^s



Simulated Data from Local Level Model with SNR $= 2$

States X_t, smooths X_t^n, and 95% CIs based on Σ_t^s



Estimating the State-Space Model Parameters

So far, we've assumed that the parameters $\boldsymbol{\theta} = (\sigma_V^2, \sigma_W^2, \mu_0, \sigma_0^2)$ are known. In practice, we need to estimate from the data

This requires maximizing the marginal likelihood of the data $\boldsymbol{y}$, having integrated the latent time series $\boldsymbol{x}$ out. This is given by:

$$f(\boldsymbol{y}|\sigma_V^2, \sigma_W^2, \mu_0, \sigma_0^2) = \int f(\boldsymbol{y}|\boldsymbol{x}, \sigma_W^2) f(\boldsymbol{x}|\mu_0, \sigma_0^2, \sigma_V^2)$$

Maximizing over an integral can be difficult ☹

Direct Maximum Marginal Likelihood

Fortunately, our normal distribution facts tell us that the marginal distribution of $\boldsymbol{y}$ is

$$\boldsymbol{y} \sim \mathrm{N}(\mathrm{E}(\boldsymbol{x}), \mathrm{Var}(\boldsymbol{x}) + \sigma_W^2 I_n).$$

However, the direct evaluation of the marginal likelihood can be challenge due to $n \times n$ matrix inversions

Alternative, we use the innovations $U_t = Y_t - Y_t^{t-1}$ to compute the likelihood:

$$\ell(\boldsymbol{\theta}) \propto f(u_1) \prod_{t=2}^{n} f(u_t | \boldsymbol{y}_{1:t-1}).$$

We can do the following iteratively:

- Pick an initial guess $\hat{\boldsymbol{\theta}}^0$ and run the Kalman filter to get a set of innovations
- Maximizing $\boldsymbol{\theta}$ (e.g., via Newton–Raphson) with $\boldsymbol{u}$ to obtain new estimate of $\boldsymbol{\theta}$

Expectation-Maximization (EM) Maximum Marginal Likelihood

Another way to compute maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is to use the expectation-maximization (EM) algorithm [Dempster, Laird, and Rubin, 1977]

- Initialize by choosing starting value $\boldsymbol{\theta}^0$, and compute the incomplete likelihood
- Perform the E-step to obtained X_t^n, Σ_t^n
- Perform M-step to update the estimate $\boldsymbol{\theta}$ using the complete likelihood
- Recompute the incomplete likelihood
- Repeat until convergence, i.e., $|\hat{\boldsymbol{\theta}}^N - \hat{\boldsymbol{\theta}}^{N-1}| < \epsilon$

Bayesian Estimation

Markov Chain Monte Carlo (MCMC) methods, such as the Gibbs sampler [Gelfand and Smith, 1990] or the Metropolis-Hastings algorithm [Metropolis et al., 1953; Hastings, 1970], are commonly used for Bayesian inference in state space models

Gibbs Sampler for State Space Models

1. Draw $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{x}_{0:n}, \boldsymbol{y}_{\mathbf{1:n}})$, where

$$p(\boldsymbol{\theta}|\boldsymbol{x}_{0:n}, \boldsymbol{y}_{1:n}) \propto \pi(\boldsymbol{\theta}) p(x_0|\boldsymbol{\theta}) \prod_{t=1}^{n} p(x_t|x_{t-1}, \boldsymbol{\theta}) p(y_t|x_t, \boldsymbol{\theta})$$

2. Draw $\boldsymbol{x}_{0:n}$ from $p(\boldsymbol{x}_{0:n}|\boldsymbol{y}_{1:n}, \boldsymbol{\theta})$, where

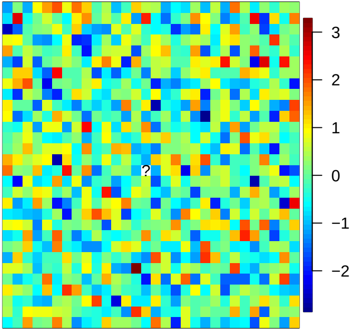
$$p(\boldsymbol{x}_{0:n}|\boldsymbol{y}_{1:n}, \boldsymbol{\theta}) = p(x_n|\boldsymbol{y}_{1:n}, \boldsymbol{\theta}) p(x_{n-1}|x_n, \boldsymbol{y}_{1:n-1}, \boldsymbol{\theta}) \cdots p(x_0|x_1, \boldsymbol{\theta})$$

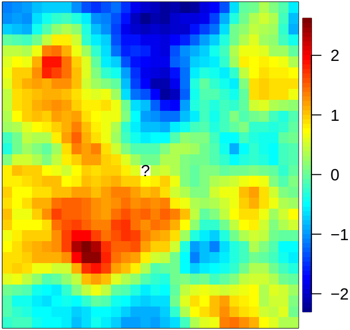
   Use forward-filtering, backward sampling (FFBS) algorithm to sequentially simulating the individual states backward

Geostatistics

Toy Examples of Spatial Interpolation

Let's consider two spatial images, each with a missing pixel

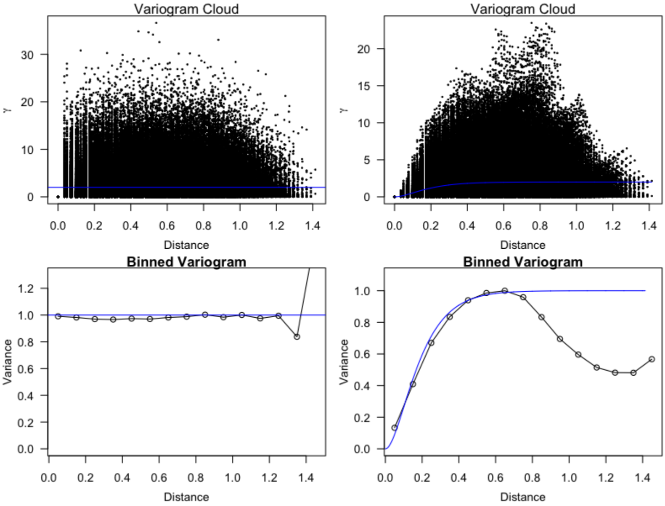




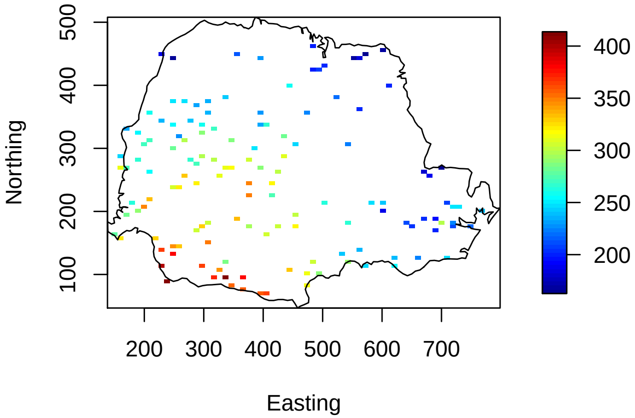
Question: What is your best guess of the value of the missing pixel, denoted as $Y(s_0)$, for each case?

Visualizing Spatial Dependence Structure

Similar to time series analysis, we can compute the covariance between data points in space to examine the degree of spatial dependence.



Interpolating Paraná State Precipitation Data



Goal: To interpolate the values in the spatial domain

The Spatial Interpolation Problem

Given observations of a spatially varying quantity Y at n spatial locations

$$y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \cdots, y(\boldsymbol{s}_n), \qquad \boldsymbol{s}_i \in \mathcal{S}, \, i = 1, \cdots, n$$

We want to estimate this quantity at any **unobserved location**

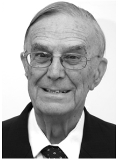
$$Y(\boldsymbol{s}_0), \qquad \boldsymbol{s}_0 \in \mathcal{S}$$

Applications

- Mining: ore grade
- Climate: temperature, precipitation, $\cdots$
- Remote Sensing: CO_2 retrievals
- Environmental Science: air pollution levels, $\cdots$

Some History of Spatial Statistics

- Mining (Krige 1951) Matheron (1960s), Forestry (Matérn 1960)



- More recent work: Cressie (1993) Stein (1999)


Outline

Linear Interpolation

The best guess (in a statistical sense) should be based on the conditional distribution $[Y(\boldsymbol{s}_0) | \boldsymbol{Y} = \boldsymbol{y}]$ where

$$\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$$

- Calculating this conditional distribution can be difficult
- Instead we use a linear predictor:

$$\hat{Y}(\boldsymbol{s}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i y(\boldsymbol{s}_i)$$

- The best linear predictor is completely determined by the mean and covariance of $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S}\}$

Next, we will introduce a class of spatial model where the distribution is fully determined by its mean and covariance

Gaussian Process (GP) Spatial Model

We assume that the observed data $\{y(\boldsymbol{s}_i)\}_{i=1}^{n}$ is one partial realization of a (continuously indexed) spatial GP $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$.

Model:

$$Y(\boldsymbol{s}) = m(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$$

where

- Mean function:

$$m(\boldsymbol{s}) = \mathrm{E}\left[Y(\boldsymbol{s})\right] = \boldsymbol{X}^T(\boldsymbol{s})\boldsymbol{\beta}$$

- Covariance function:

$$\{\epsilon(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}} \sim \mathrm{GP}\left(0, K(\cdot,\cdot)\right), \quad K(\boldsymbol{s}_1,\boldsymbol{s}_2) = \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1),\epsilon(\boldsymbol{s}_2)\right)$$

Assumptions on Covariance Function

In practice, the covariance must be estimated from the data $(y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$. We need to impose some structural assumptions

- Stationarity:

$$\begin{aligned} K(\boldsymbol{s}_1, \boldsymbol{s}_2) &= \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2)\right) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2) \\ &= \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1 + \boldsymbol{h}), \epsilon(\boldsymbol{s}_2 + \boldsymbol{h})\right) \end{aligned}$$

- Isotropy:

$$K(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{Cov}\left(\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2)\right) = C(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)$$

A Valid Covariance Function Must Be Positive Definite!

A covariance function is positive definite (p.d.) if

$$\sum_{i,j=1}^{n} a_i a_j C(\boldsymbol{s}_i - \boldsymbol{s}_j) \geq 0$$

for any finite locations $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n$, and for any constants a_i, $i = 1, \cdots, n$

Question: what is the consequence if a covariance function is NOT p.d.? $\Rightarrow$ We can get a negative variance

Question: How to guarantee a $C(\cdot)$ is p.d.?

- Using a parametric covariance function (see some examples in next slide)
- Using Bochner's Theorem to construct a valid covariance function

Bochner's Theorem

A complex-valued function C on $\mathbb{R}^{\mathrm{d}}$ is the covariance function for a weakly stationary mean square contituous complex-valued random process on $\mathbb{R}^d$ if and only if it can be represented as

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^d} \exp(i\omega^{\mathrm{T}}\boldsymbol{h})F(d\boldsymbol{\omega}),$$

with F a positive finite measure. When F has a density with respect to Lebesgue measure, we have the spectral density f and

$$f(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}^d} \exp(-i\omega^{\mathrm{T}}\boldsymbol{h})C(\boldsymbol{h})\,d\boldsymbol{h}$$

Some Commonly Used Covariance Functions

- Powered exponential:

$$C(h) = \sigma^2 \exp\left(-(\frac{h}{\rho})^\alpha\right), \qquad \sigma^2 > 0,\ \rho > 0,\ 0 < \alpha \le 2$$

- Spherical:

$$C(h) = \sigma^2 \left(1 - 1.5\frac{h}{\rho} + 0.5\left(\frac{h}{\rho}\right)^3\right) 1_{\{h \le \rho\}}, \qquad \sigma^2,\ \rho > 0$$

  Note: it is only valid for 1,2, and 3 dimensional spatial domain.

- Matérn:

$$C(h) = \sigma^2 \frac{\left(\sqrt{2\nu} h/\rho\right)^\nu \mathcal{K}_\nu\left(\sqrt{2\nu} h/\rho\right)}{\Gamma(\nu) 2^{\nu-1}}, \qquad \sigma^2 > 0,\ \rho > 0,\ \nu > 0$$

"Use the Matérn model" – Stein (1999, pp. 14)

1-D Realizations from Matérn Model with Fixed σ^2, ρ

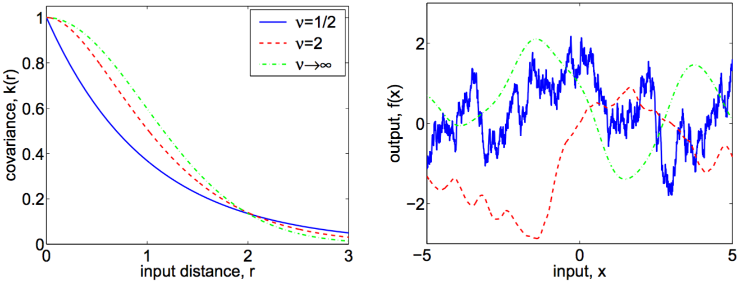
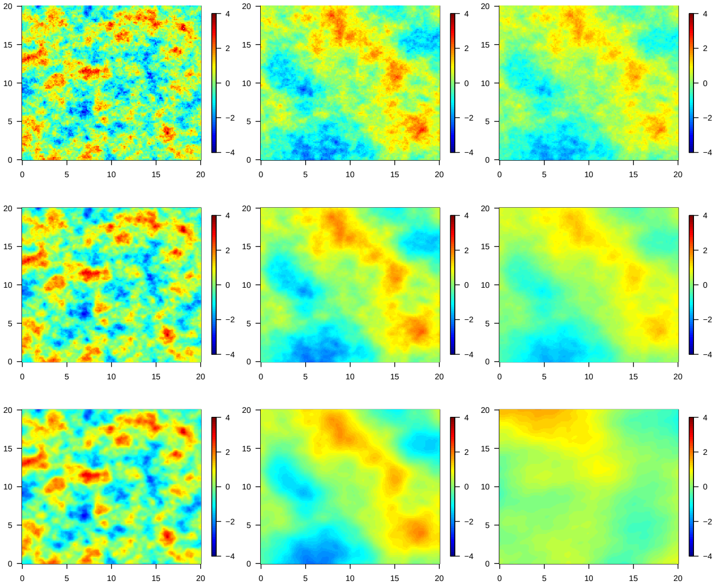


Figure: courtesy of Rasmussen & Williams 2006

The larger ν is, the smoother the process is

2-D Realizations from Matérn Model with Fixed σ^2



Outline

Conditional Distribution of Multivariate Normal

If

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

Then

$$[\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2] \sim \mathrm{N}\left(\boldsymbol{\mu_{1|2}}, \Sigma_{1|2}\right)$$

where

$$\begin{aligned} \boldsymbol{\mu_{1|2}} &= \boldsymbol{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}\left(\boldsymbol{y}_2 - \boldsymbol{\mu_2}\right) \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

GP-Based Spatial Interpolation: Kriging

If $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ follows a GP, then

$$\begin{pmatrix} Y_0 \\ \boldsymbol{Y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} m_0 \\ \boldsymbol{m} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & k^{\mathrm{T}} \\ k & \Sigma \end{pmatrix}\right)$$

We have

$$[Y_0 | \boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}, \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

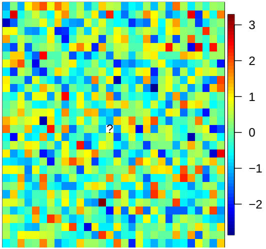
where

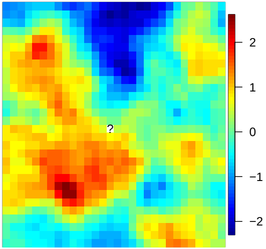
$$\begin{aligned} m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} &= m_0 + k^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{m}) \\ \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} &= \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k \end{aligned}$$

Next, we are going to revisit our toy examples

Toy Examples Revisited

For simplicity, we assume $m(s) = 0$ for $s \in \mathcal{S}$, the spatial covariance only depends on distance





$$m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = 0 + k^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{y} - \boldsymbol{0}\right), \quad \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k$$

Spatial uncorrelated field:

- $m_{Y_0|\boldsymbol{Y}} = 0$
- $\sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2$

Spatial correlated field:

- $m_{Y_0|\boldsymbol{Y}} = k^{\mathrm{T}}\Sigma^{-1}\boldsymbol{y}$
- $\sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} = \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k$

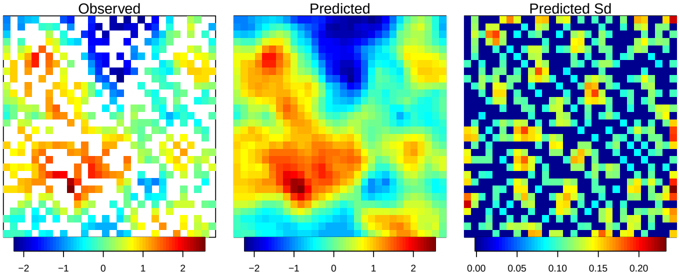
Interpolating Multiple Points in Space

In practice, we would like to predict the values at many locations. The Gaussian conditional distribution formula can still be used:

$$[\boldsymbol{Y}_0 | \boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}, \Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

where

$$\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \boldsymbol{m}_0 + \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{m})$$
$$\Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \Sigma_0 - \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{k}$$



GP-Based Spatial Interpolation: Kriging

If $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ follows a GP, then

$$\begin{pmatrix} \boldsymbol{Y}_0 \\ \boldsymbol{Y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{m}_0 \\ \boldsymbol{m} \end{pmatrix}, \begin{pmatrix} \Sigma_0 & \boldsymbol{k}^{\mathrm{T}} \\ \boldsymbol{k} & \Sigma \end{pmatrix}\right)$$

We have

$$[\boldsymbol{Y}_0|\boldsymbol{Y} = \boldsymbol{y}] \sim \mathrm{N}\left(\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}, \Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}}\right)$$

where

$$\boldsymbol{m}_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \boldsymbol{m}_0 + \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{m})$$
$$\Sigma_{\boldsymbol{Y}_0|\boldsymbol{Y}=\boldsymbol{y}} = \Sigma_0 - \boldsymbol{k}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{k}$$

Question: what if we don't know $m(s;\boldsymbol{\beta}), c(h;\boldsymbol{\theta})$?

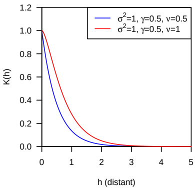
⇒ We need to estimate the mean and covariance from the data $\boldsymbol{y}$.

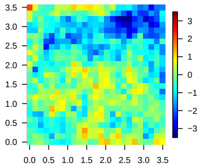
Outline

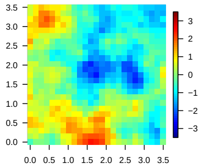
Recap: Gaussian Process

Assume $\{y(\boldsymbol{s}_i)\}_{i=1}^{n}$ is one partial realization of a spatial stochastic process $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$.

- Gaussian Processes $\mathrm{GP}\left(m\left(\cdot\right), K\left(\cdot,\cdot\right)\right)$ are widely used in modeling spatial stochastic processes, where the covariance $K\left(\cdot,\cdot\right)$ is typically assumed to be a stationary and isotropic covariance function $C(h)$ that depends on spatial distance h only
- Spatial statisticians often focus on the covariance function. e.g. $C(h) = \sigma^2 \frac{\left(\sqrt{2\nu}h/\gamma\right)^{\nu}\mathcal{K}_{\nu}\left(\sqrt{2\nu}h/\gamma\right)}{\Gamma(\nu)2^{\nu-1}}$







Variogram, Semivariogram, and Covariance Function

Under the stationary and isotropic assumptions

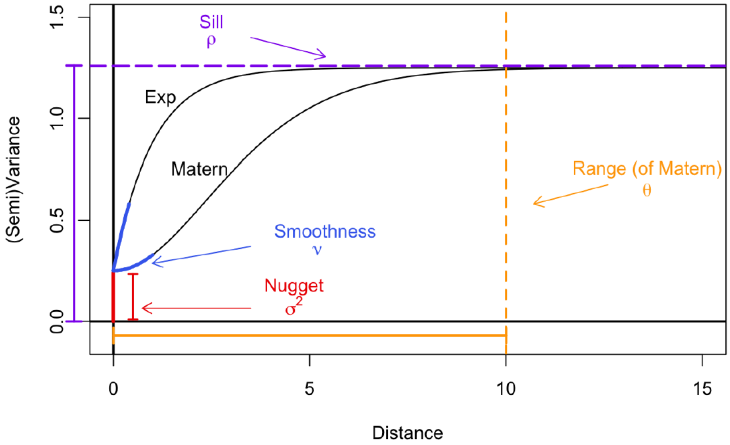
Variogram:

$$
\begin{aligned}
2\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) &= \mathrm{Var}\left(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)\right) \\
&= \mathrm{E}\left\{\left((Y(\boldsymbol{s}_i) - \mu(\boldsymbol{s}_i)) - (Y(\boldsymbol{s}_j) - \mu(\boldsymbol{s}_j))\right)^2\right\} \\
&= \mathrm{E}\left\{\left(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)\right)^2\right\} \\
&= 2\gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) = 2\gamma(h)
\end{aligned}
$$

Semivariogram and covariance function:

$$\gamma(h) = C(0) - C(h)$$

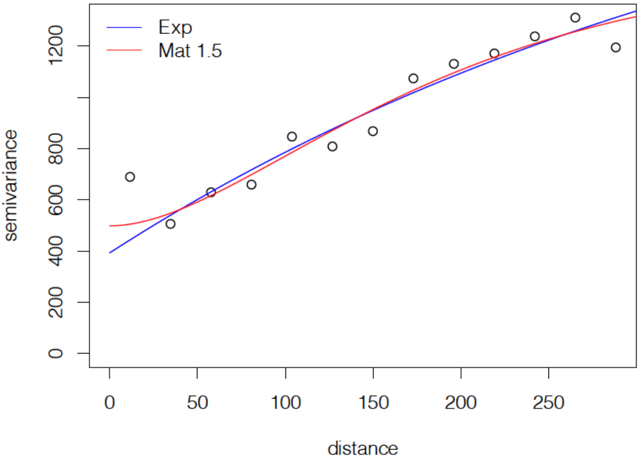
Semivariogram $\left\{\frac{1}{2}\mathrm{Var}\left(\varepsilon\left(\boldsymbol{s}_i\right)-\varepsilon\left(\boldsymbol{s}_j\right)\right)\right\}_{i,j}$



Source: `fields` vignette by Wiens and Krock, 2019

Estimation: Weighted Least Squares Method

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{u \in \mathcal{U}} \frac{N(h_u)}{[\gamma(h_u;\boldsymbol{\theta})]^2} \left[\hat{\gamma}(h_u) - \gamma(h_u;\boldsymbol{\theta})\right]^2$$



Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}}\left[\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right]^{-1}_{n\times n}(\boldsymbol{y} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(i,j) = \sigma^2\rho_{\rho,\nu}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) + \tau^2 1_{\{\boldsymbol{s}_i = \boldsymbol{s}_j\}}, i, j = 1, \cdots, n$

Maximum Likelihood Estimation (MLE)

Log-likelihood:

Given data $\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^{\mathrm{T}}$

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}}\left[\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right]_{n\times n}^{-1}(\boldsymbol{y} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(i,j) = \sigma^2\rho_{\rho,\nu}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) + \tau^2 1_{\{\boldsymbol{s}_i = \boldsymbol{s}_j\}}, i, j = 1, \cdots, n$

for any fixed $\boldsymbol{\theta}_0 \in \Theta$ the unique value of β that maximizes ℓ_n is given by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^{\mathrm{T}}\Sigma_{\boldsymbol{\theta}_0}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\Sigma_{\boldsymbol{\theta}_0}\boldsymbol{y}$$

Then we obtain the profile log likelihood

$$\ell_n(\boldsymbol{\theta}; \boldsymbol{y}) \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}P(\boldsymbol{\theta})\boldsymbol{y}$$

where

$$P(\boldsymbol{\theta}) = \Sigma_{\boldsymbol{\theta}}^{-1} - \Sigma_{\boldsymbol{\theta}}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\Sigma_{\boldsymbol{\theta}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\Sigma_{\boldsymbol{\theta}}$$

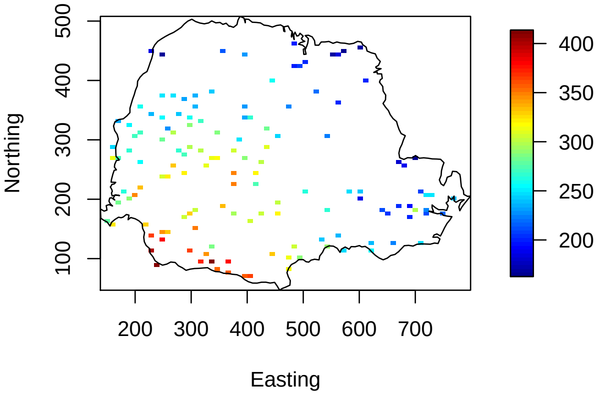
Solve the maximization problem above to get the MLE

Remarks on Likelihood-based estimation

- Maximizing $\ell_n(\boldsymbol{\theta}; \boldsymbol{y})$ involves solving a constrained nonlinear optimization problem, necessitating numerical methods for obtaining ML estimates.
- Alternatively, Restricted (or residual) maximum likelihood (REML) can be employed.
- Likelihood-based estimation poses computational challenges with large spatial datasets, primarily due to the significant computational complexity, requiring $\mathcal{O}(n^3)$ operations and $\mathcal{O}(n^2)$ memory.

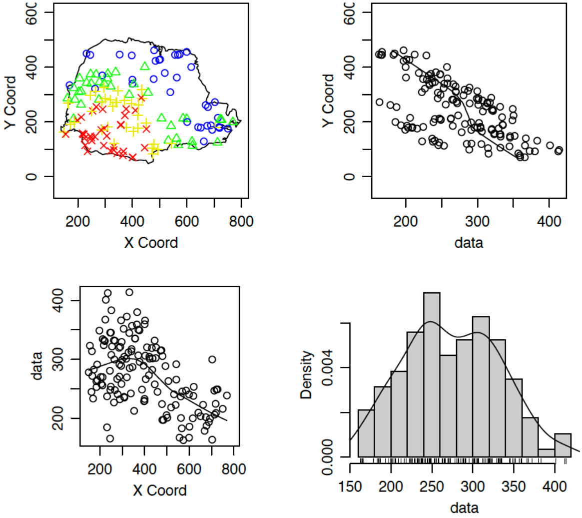
Paraná State Precipitation Data

We look at the average winter (May-June, dry season) rainfall at 143 locations throughout Paraná, Brazil



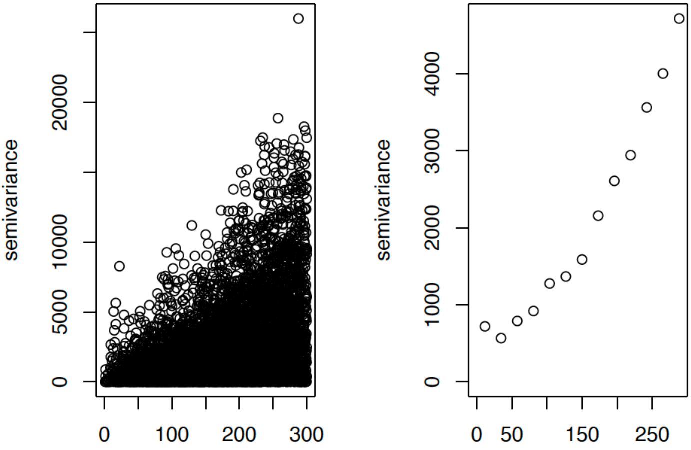
Goal: To interpolate the values in the spatial domain

Exploratory Data Analysis



A linear trend in space (both longitude and latitude) may be suitable to characterize the large-scale spatial trend

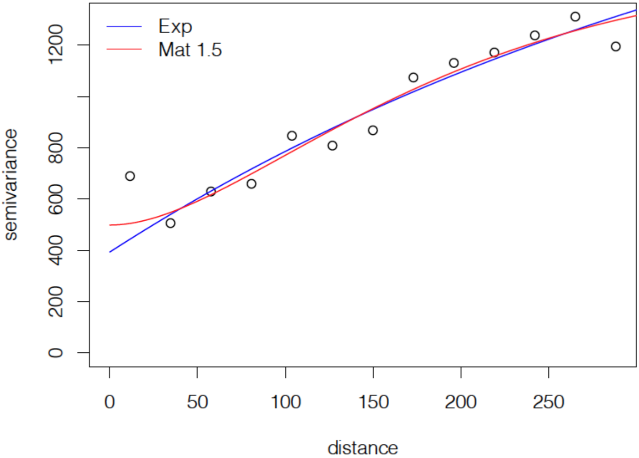
Variogram Analysis



An increasing variogram pattern suggests a positive spatial dependence structure.

Estimating Spatial Covariance Via Varogram

```
parana.vtfit.exp <- variofit(parana.variot)
parana.vtfit.mat1.5  <- variofit(parana.variot, kappa = 1.5)
```



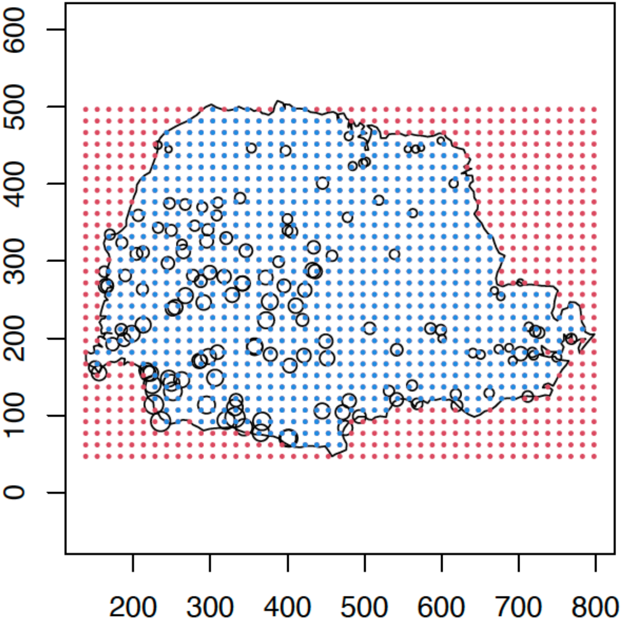
Maximum Likelihood Estimation of Paraná Rainfall


```
(parana.ml1 <- likfit(parana, trend = "1st", ini = c(1000, 50), nug = 100))
```

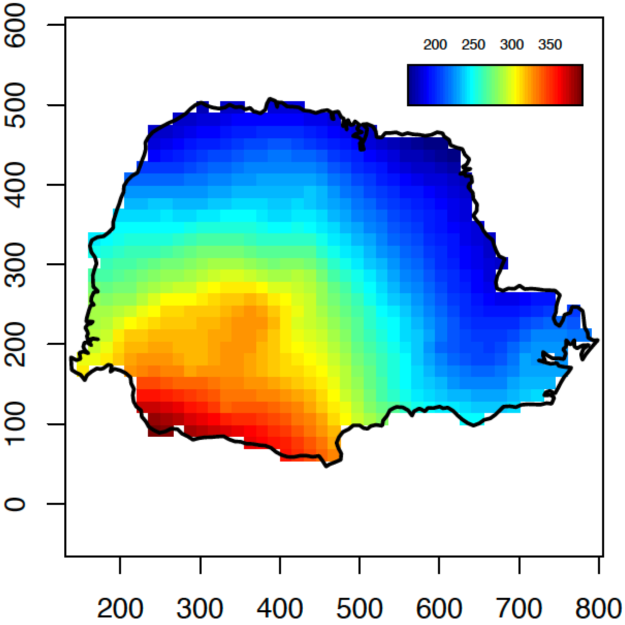
```
## ---------------------------------------------------------------
## likfit: likelihood maximisation using the function optim.
## likfit: Use control() to pass additional
##          arguments for the maximisation function.
##         For further details see documentation for optim.
## likfit: It is highly advisable to run this function several
##         times with different initial values for the parameters.
## likfit: WARNING: This step can be time demanding!
## ---------------------------------------------------------------
## likfit: end of numerical maximisation.

## likfit: estimated model parameters:
##      beta0      beta1      beta2      tausq    sigmasq        phi
## "416.4984" " -0.1375" " -0.3997" "385.5180" "785.6904" "184.3863"
## Practical Range with cor=0.05 for asymptotic range: 552.3719
##
## likfit: maximised log-likelihood = -663.9
```

Next, we will use these information to conduct spatial interpolation

Setting Up the Spatial Grids for Prediction



Spatial Predicted Map



Prediction Uncertainty Map