Lecture 15

Interpolation of Spatial Data II

DSA 8020 Statistical Methods II

Whitney Huang
Clemson University

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

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

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

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

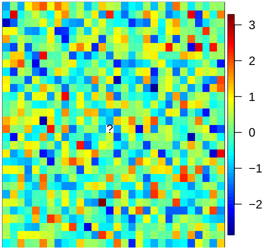
where

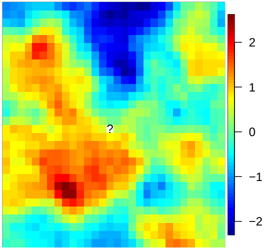
$$\begin{aligned} m_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} &= m_0 + k^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{m}) \\ \sigma^2_{Y_0|\boldsymbol{Y}=\boldsymbol{y}} &= \sigma_0^2 - k^{\mathrm{T}}\Sigma^{-1}k \end{aligned}$$

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:

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

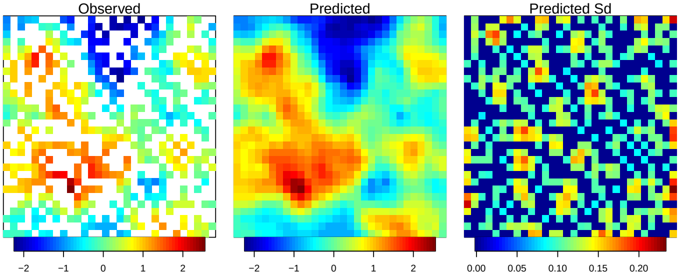
Interpolating Multiple Points in Space

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

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

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

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

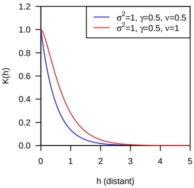
Outline

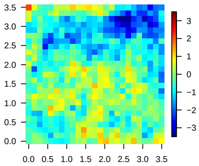
Recap: Gaussian Process

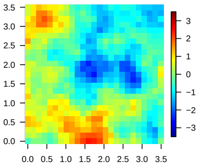
Assume $\{y(\boldsymbol{s}_i)\}_{i=1}^{n}$ is one partial realization of a spatial stochastic process $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$.

- 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

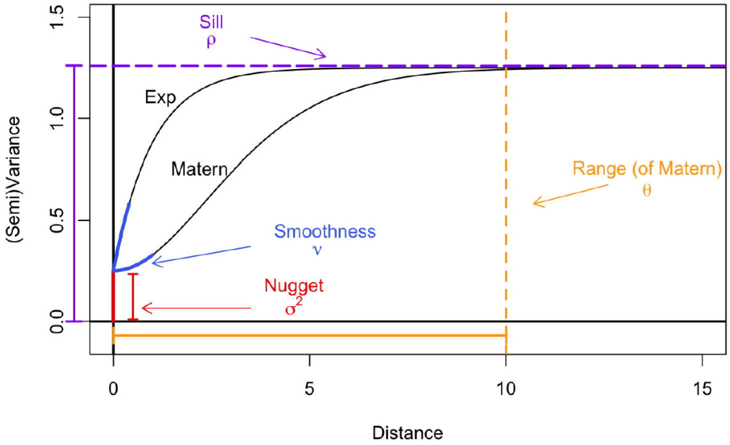
Variogram:

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

Semivariogram and covariance function:

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

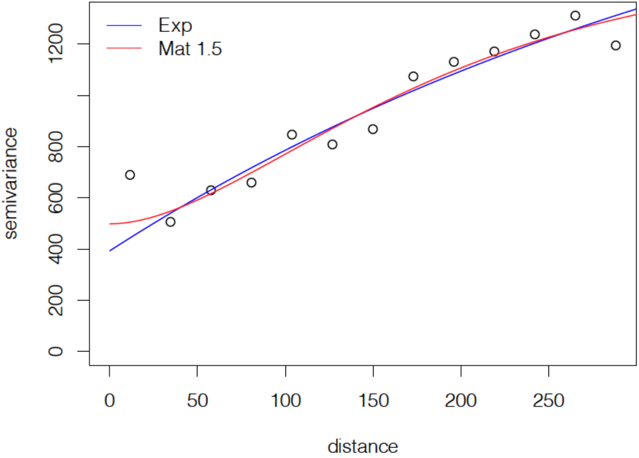
Semivariogram $\{\frac{1}{2}\mathrm{Var}\left(\varepsilon\left(\boldsymbol{s}_i\right)-\varepsilon\left(\boldsymbol{s}_j\right)\right)\}_{i,j}$



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 $\boldsymbol{\beta}$ 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

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

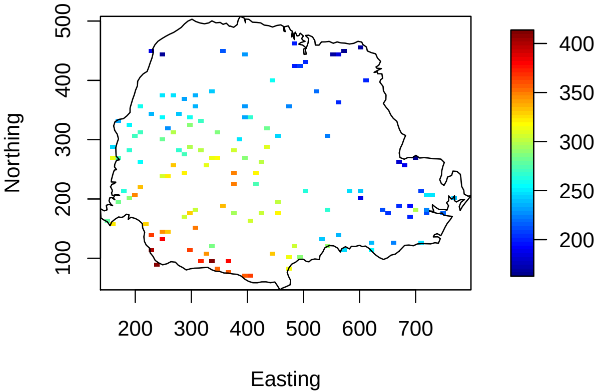
Solve the maximization problem above to get the MLE

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.

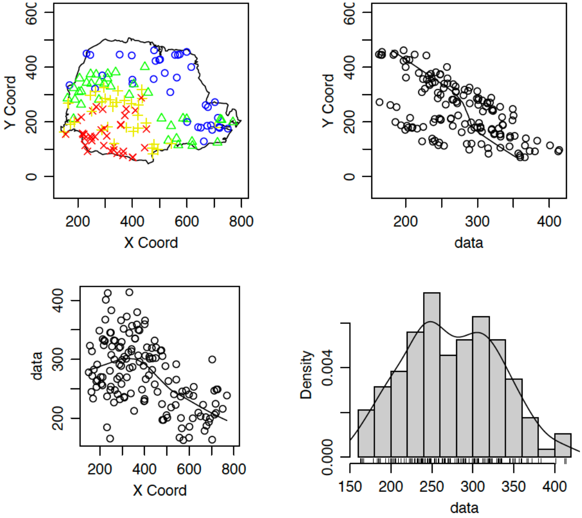
Paraná State Precipitation Data

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



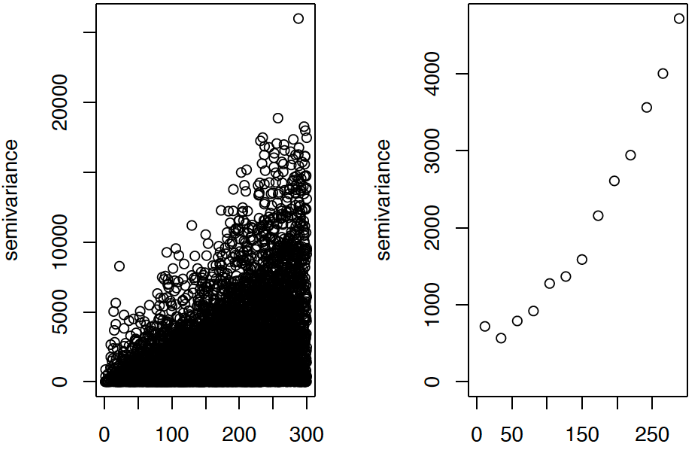
Goal: To interpolate the values in the spatial domain

Exploratory Data Analysis



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

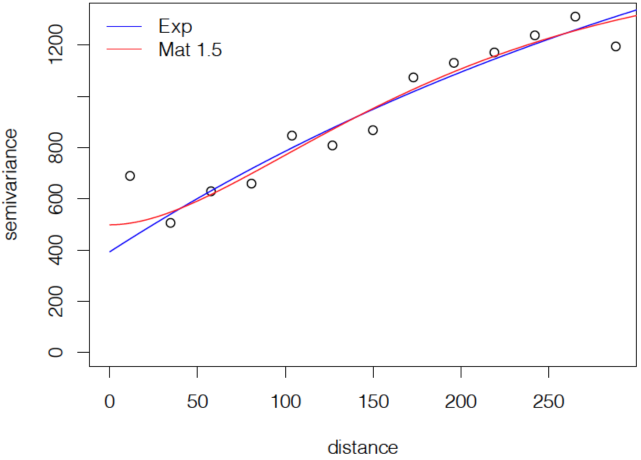
Variogram Analysis



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)
```



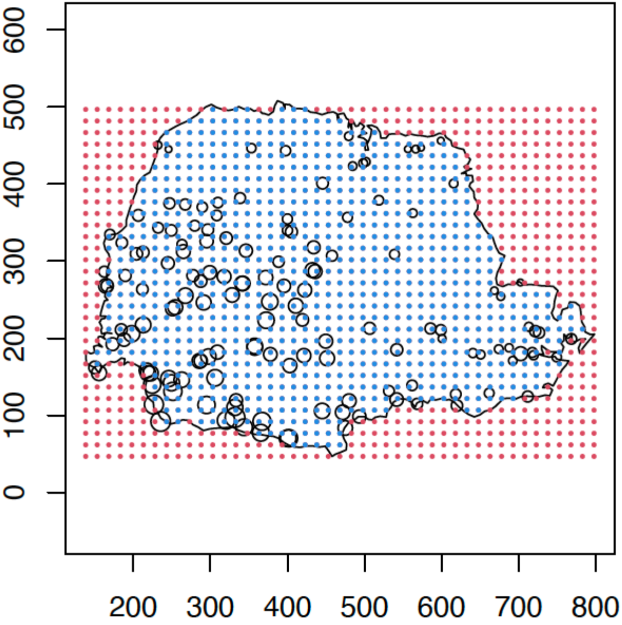
Maximum Likelihood Estimation of Paraná Rainfall

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

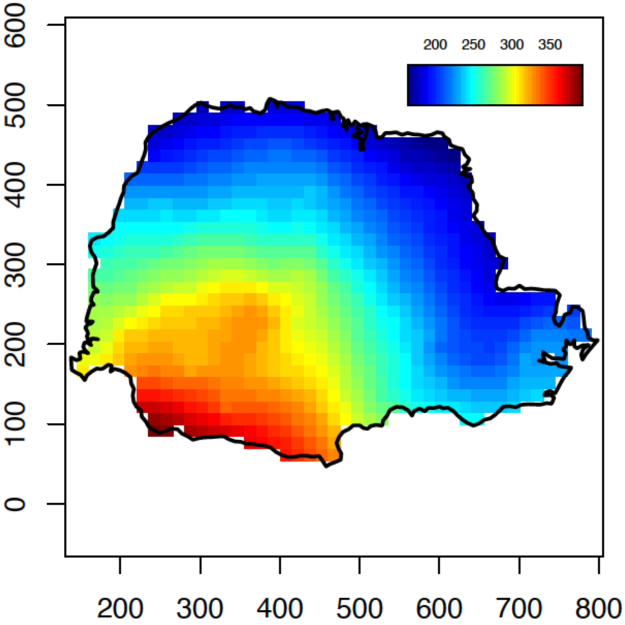
```
## ---------------------------------------------------------------
## likfit: likelihood maximisation using the function optim.
## likfit: Use control() to pass additional
##          arguments for the maximisation function.
##         For further details see documentation for optim.
## likfit: It 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

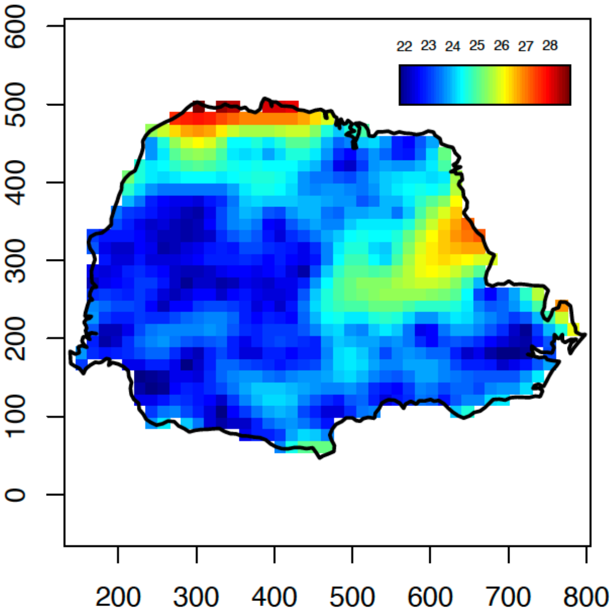
Setting Up the Spatial Grids for Prediction



Spatial Predicted Map



Prediction Uncertainty Map



Summary

These slides cover:

- Parameter Estimation for Gaussian Process Spatial Models
- Spatial predictions using Gaussian Process Spatial Models

`R` functions to know:

- `quilt.plot` (under the package `fields`) for visualizing irregularly distributed spatial data
- `vgram` and `variog` (under the package `fields` and `geoR`, respectively) for visualizing spatial dependence
- `variofit` and `likfit` from the package `geoR` for condcuting weighted least squares and maximum likelihood estimation