TMA4300 Project 1

Jakob Heide, Bendik Waade

feb 18, 2024

Part 1: GRFs - model characteristics

Let X be a stationary GRF on $\mathcal{D} = [0, 50] \in \mathbb{R}$, with

$$\begin{aligned}
&\mathrm{E}[X(s)] = 0,\ s \in \mathcal{D}, \\
&\mathrm{Var}[X(s)] = \sigma^2,\ s \in \mathcal{D} \\
&\mathrm{Corr}[X(s), X(s')] = \rho(\|s - s'\|),\ s, s' \in \mathcal{D}
\end{aligned}$$

Correlation functions and semi-variograms

The correlation function is a *positive semi-definite* function if it satisfies

$$\forall m \in \mathbb{N}, \forall a_1, \ldots, a_m \in \mathbb{R}, \forall s_1, \ldots, s_m \in \mathcal{D}$$
$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \rho(\|s_i - s_j\|) \geq 0$$

This is a necessary requirement for any correlation function. To see this, we note that $\rho(\|s-s'\|) = c(s, s')/\sigma^2$, where σ^2 is the marginal variance. The requirement above becomes

$$\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j c(s_i, s_j) = \frac{1}{\sigma^2}\mathbf{a}^T \mathrm{Var}(\mathbf{X})\mathbf{a} = \frac{1}{\sigma^2}\mathrm{Var}\left(\sum_{i=1}^{m} a_i X(s_i)\right) \geq 0$$

where $\mathbf{a}^T = (a_1, \ldots, a_m)$ and $\mathbf{X} = (X(s_1), \ldots, X(s_m))^T$. Hence, the requirement on the correlation function to be positive semi-definite is the same as requiring that the covariance matrix of any set of points is positive semi-definite, or rather, that the variance of any linear combination is non-negative.

In Figure 1a, the isotropic powered exponential correlation function is shown, which has the form

$$\rho_0(h) = \exp(-(h/a)^\alpha),\ h > 0 \tag{1}$$

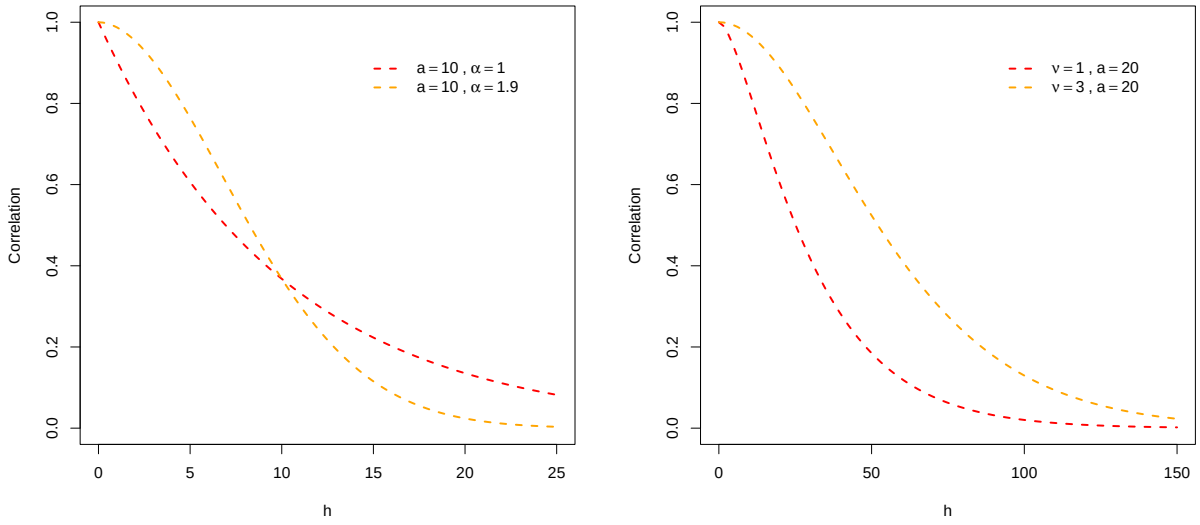
where $a > 0$ is called the spatial scale and $0 < \alpha \leq 2$ is called the power. In Figure 1b, the isotropic Matérn correlation function is shown, which has the form

$$\rho_0(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{h}{a}\right)^\nu K_\nu\left(\frac{h}{a}\right),\ h > 0 \tag{2}$$

where K_ν is the modified Bessel function of order ν. The parameter a is called the power and the parameter ν is called the smoothness. Note that there are several parametrizations of the Matérn, e.g.

$$\rho_0(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{8\nu}h}{a}\right)^\nu K_\nu\left(\frac{\sqrt{8\nu}h}{a}\right),\ h > 0$$

is another parametrization.



(a) Powered exponential

(b) Matérn

Figure 1: Correlation functions

For the powered exponential, we see that increasing the parameter α increases the correlation between points that are less than a away, but it decreases the correlation for points that are more than a away. For the Matérn, we see that increasing the parameter ν increases the correlation between all points. Another crucial feature of the spatial correlation function is the differentiability. The powered exponential correlation function is not differentiable for either of the parameter sets $(a, \alpha) \in \{(10, 1), (10, 1.9\}$. To see this, we write the isotropic correlation function as

$$\rho_0(|h|) = \exp(-(|h|/a)^\alpha), \ h \in \mathbb{R}$$

which has a sharp point in $h = 0$. On the other hand, a GRF with the Matérn correlation function is $\lceil \nu \rceil - 1$ times differentiable, so we expect the Matérn with $\nu = 3$ to produce realizations which are 2 times differentiable.

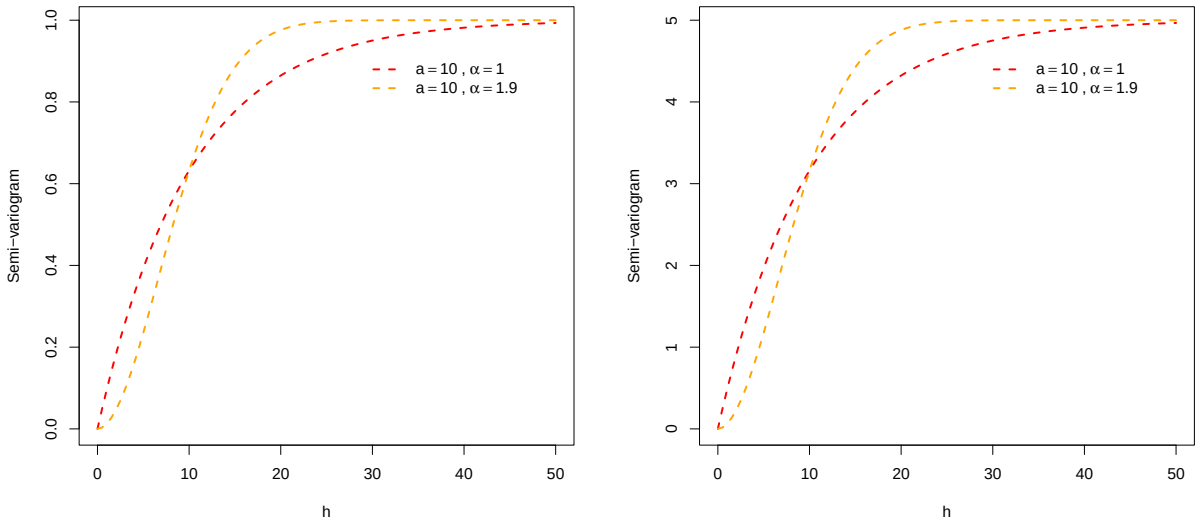
The *semi-variogram* is defined as

$$\gamma(h) = \frac{1}{2}\text{Var}(X(h) - X(0))$$

which we can express as

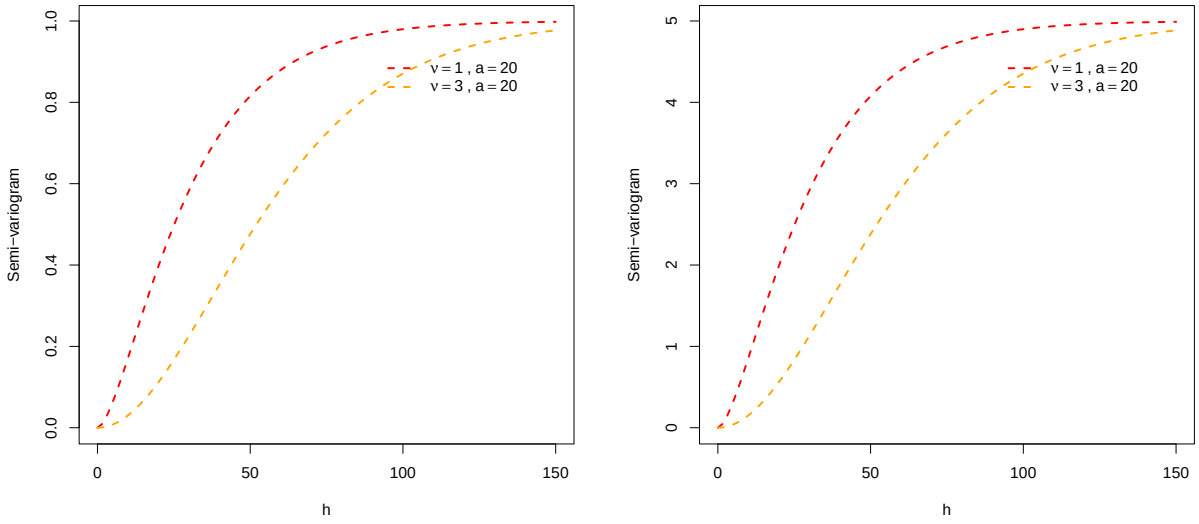
$$\gamma(h) = C(0) - C(h) = \sigma^2(1 - \rho_0(h)), \ h > 0$$

If we multiply the correlation functions in Equation (1) and (2) by the marginal variance σ^2, we obtain the corresponding stationary covariance functions. Let $\sigma^2 \in \{1, 5\}$, which gives in total eight semi-variograms for the functions in Figure 1, which are shown in Figure 2.



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 5$

Figure 2: Semi-variograms using the powered exponential



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 5$

Figure 3: Semi-variograms using the Matérn

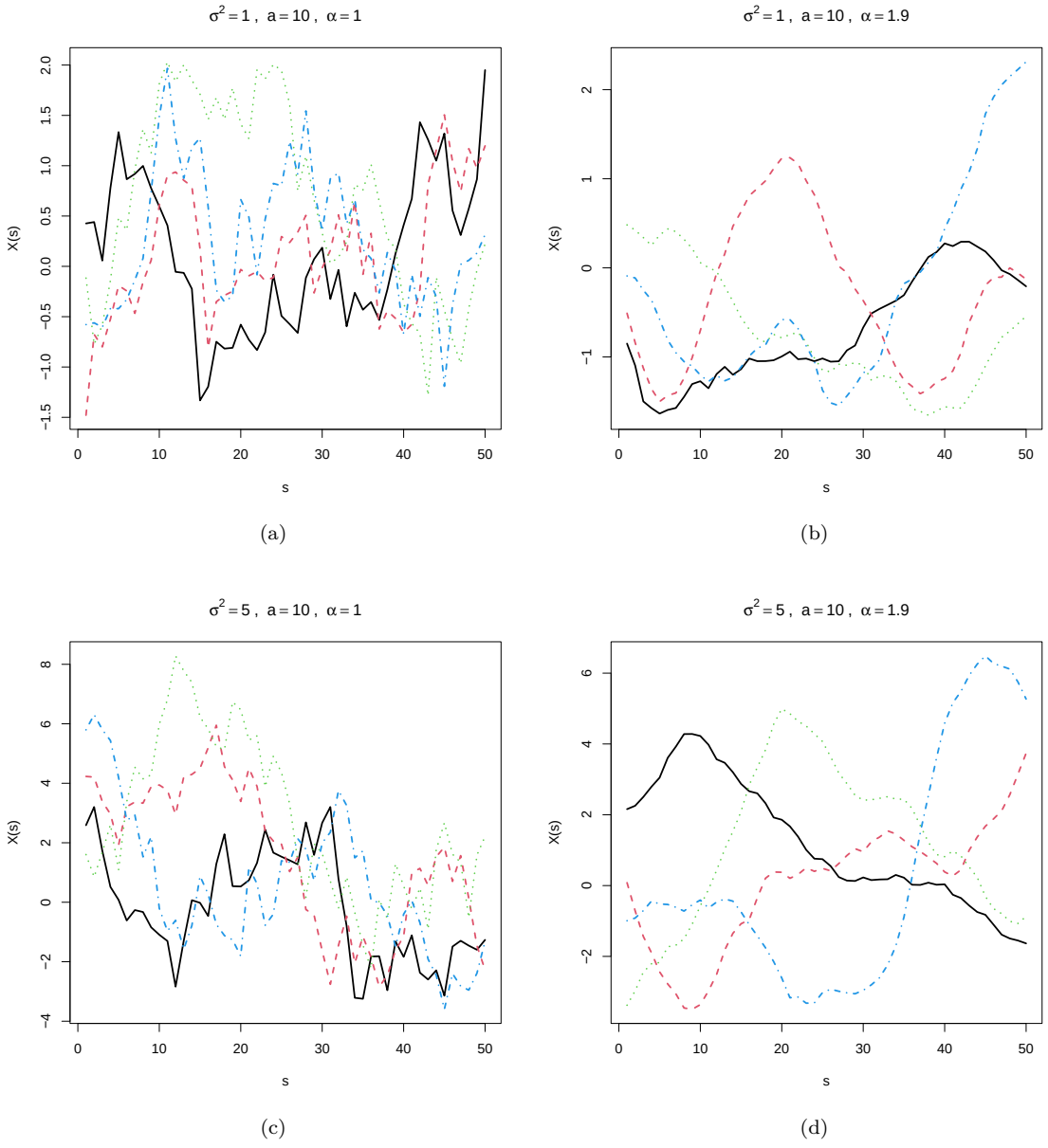
Semi-variograms are mostly used for exploratory analysis and can be used to estimate the sill, nugget and range. In the case of the both the powered exponential and the Matérn, we see that the semi-variograms grow toward the value of the marginal variance σ^2, which is also the sill in this case. Increasing the parameter α in the powered exponential increases the range, as seen in Figure 2a,b. Increasing the parameter ν in the Matérn also increases the range, as seen in Figure 3a,b.

Realizations

The expected value of $\mathbf{X} = (X(1), X(2), \ldots, X(50))^T$ is the vector $\mu = (0, \ldots, 0)^T$, since X is a centred GRF. We denote the variance of $\mathbf{X}$ as $\Sigma = \text{Var}(\mathbf{X})$, where $\Sigma_{ij} = C_0(\|i-j\|) = \sigma^2\rho(\|i-j\|)$, for $i, j = 1, \ldots, 50$. The distribution of $\mathbf{X}$ is thus

$$\mathbf{X} \sim N_{50}(\mathbf{0}, \Sigma)$$

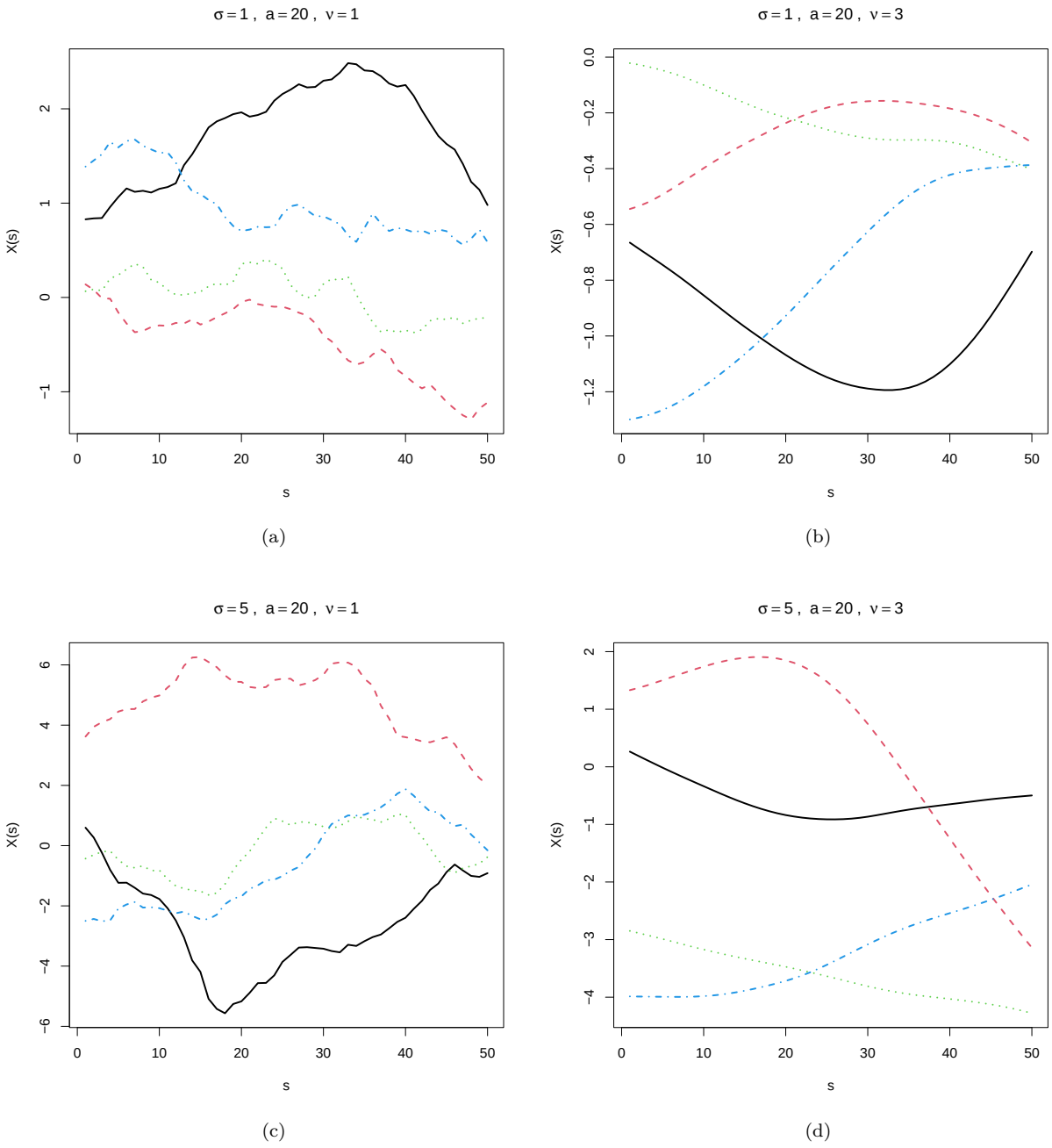
We can now simulate realizations of $\mathbf{X}$ with both the powered exponential covariance function and the Matérn covariance function, for the eight different sets of parameters used above. Four realizations for each of the sets of parameters are shown in Figure 4 (the powered exponential) and in Figure 5 (the Matérn).



(a) (b)

(c) (d)

Figure 4: Realizations using the powered exponential



(a) (b)

(c) (d)

Figure 5: Realizations using the Matérn

We see that the realizations in Figure 5b and 5d are the smoothest - the smoothness parameter $\nu = 3$ indicates that X should be 2 times differentiable. In Figure 4, we see the effect of the power parameter α. In 4a and 4c, the realizations fluctuate more over smaller intervals than in 4b and 4d. Increasing the parameter α will give a higher covariance with points that are less than a away, and smaller covariance with the points that are further than a away, which effectively restricts the small-scale fluctuations. An increase in the marginal variance parameter σ^2 increases the large-scale fluctuations.

Observation model distribution

Consider the observation model

$$Y_i = X(s_i) + \epsilon_i,\ i = 1, 2, 3,$$

where $\epsilon_i \sim N(0, \sigma_N^2)$ are i.i.d. and independent of X, and $(s_1, s_2, s_3) = (10, 25, 30)$. σ_N^2 is the nugget variance. The expected value of $\mathbf{Y} = (Y_1, Y_2, Y_3)^T$ is

$$\mathrm{E}[\mathbf{Y}] = (\mathrm{E}[X(s_1) + \epsilon_1], \mathrm{E}[X(s_2) + \epsilon_2], \mathrm{E}[X(s_3) + \epsilon_3])^T = (0, 0, 0)^T$$

and the variance of $\mathbf{Y}$ is

$$\mathrm{Var}(\mathbf{Y}) = \Sigma_{\mathbf{Y}} = [\mathrm{Cov}(Y_i, Y_j)]_{i,j=1,2,3}$$

where $\mathrm{Cov}(Y_i, Y_j) = \sigma^2 + \sigma_N^2$ for $i = j$ and $\mathrm{Cov}(Y_i, Y_j) = C_0(|s_i - s_j|) = \sigma^2 \rho_0(|s_i - s_j|)$ for $i \neq j$. Since $X(s_i)$ is normally distributed, Y_i is also normally distributed, and the distribution of $\mathbf{Y}$ is

$$\mathbf{Y} \sim N_3(\mathbf{0}, \Sigma_{\mathbf{Y}})$$

Conditional distribution and the best linear unbiased predictor

Since $\mathbf{X}$ and $\mathbf{Y}$ are normal, so is the conditional distribution $\mathbf{X}|\mathbf{Y} = \mathbf{y}$. The expected value of $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ is

$$\mathrm{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}$$

where we have constructed the multivariate normal vector

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix},\ \text{with } \mathrm{E}(\mathbf{Z}) = 0,\ \mathrm{Var}(\mathbf{Z}) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

The variance is

$$\mathrm{Var}(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Σ_{11} is the covariance matrix of $\mathbf{X}$, and Σ_{22} is the covariance matrix of $\mathbf{Y}$, both of which we have already seen how to calculate. To calculate Σ_{12} and Σ_{21}, we use the definition of covariance to find

$$\Sigma_{12} = \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathrm{E}(\mathbf{X}\mathbf{Y}^T) = [C_0(|i - s_j|)]_{i=1,\ldots,50,\ j=1,2,3,}$$

that is, a 50×3 matrix. Since the covariance function is isotropic, we note that $\Sigma_{12} = \Sigma_{21}^T$. To predict X based on the observations $\mathbf{Y} = \mathbf{y}$, we consider the *best linear unbiased predictor*, also known as the Simple Kriging predictor, which takes the form

$$\hat{\mathbf{X}} = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{y},\ \text{with } \mathrm{Cov}(\hat{\mathbf{X}} - \mathbf{X}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{3}$$

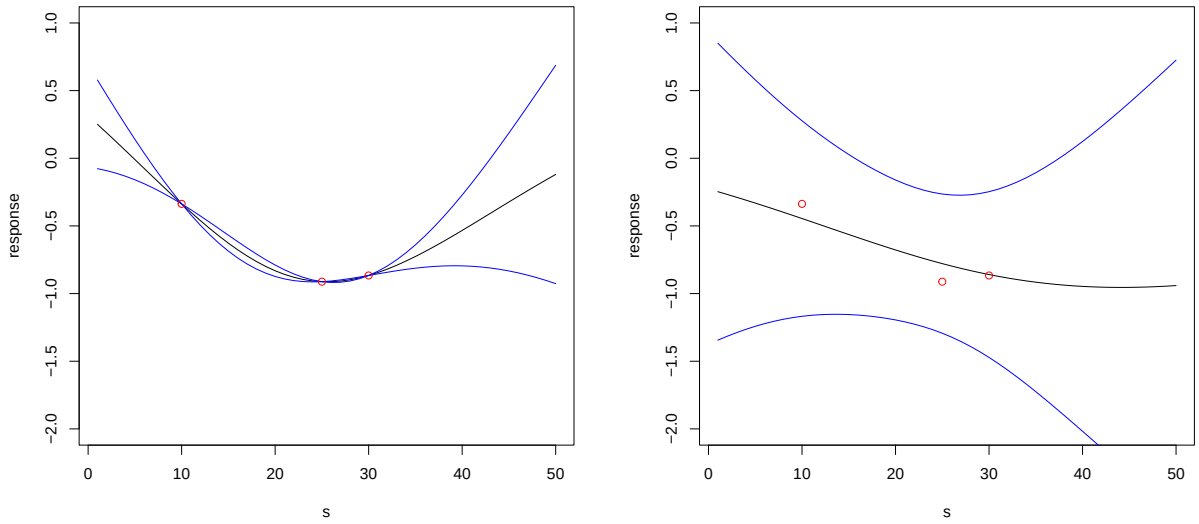
which coincides with the conditional expectation and variance - this is due to X being a GRF. We extract the values at s_1, s_2, s_3 from one of the realizations in Figure 5d and use these as our observations $\mathbf{y}$, i.e. we use the Matérn covariance function with $\sigma^2 = 5, a = 20$, and $\nu = 3$. Since we know the mean and variance, we can use that the test observer

$$Z = \frac{(\mathbf{X}|\mathbf{Y} = \mathbf{y}) - \mathrm{E}(\mathbf{X}|\mathbf{Y} = \mathbf{y})}{\sqrt{\mathrm{Var}(\mathbf{X}|\mathbf{Y} = \mathbf{y})}}$$

is standard normal, so the 90% prediction interval becomes

$$\hat{\mathbf{X}} \pm 1.645\sqrt{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$$

In Figure 6, the best linear unbiased predictor for X is shown, along with 90% prediction intervals, for the two cases of $\sigma_N^2 \in \{0, 0.25\}$.



(a) $\sigma_N^2 = 0$ (b) $\sigma_N^2 = 0.25$

Figure 6: Best linear unbiased predictor with prediction intervals

In Figure 6a, we see that when there is no nugget variance, the best linear unbiased prediction fits to the data points exactly. The prediction intervals also get smaller as the prediction approaches a data point, and they become larger as the prediction moves away from a data point. In Figure 6b, a nugget variance $\sigma = 0.25$ is introduced, so that the prediction no longer intersects the data points. As a consequence, the prediction intervals become much wider - however, they still narrow slightly when the prediction approaches the data points.

Since X should be two times differentiable due to the parameter $\nu = 3$, we expect the prediction intervals to be more narrow than in the case of e.g. $\nu = 1$, since the functions produced would then be allowed to fluctuate more. In fact, we already saw this tendency in Figure 5a and b.

Empirical estimates

We simulate 100 realizations of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$, and calculate the sample mean and variance, which we can view as empirical estimates of Equation (3). We denote by $\bar{X}$ the sample mean, that is,

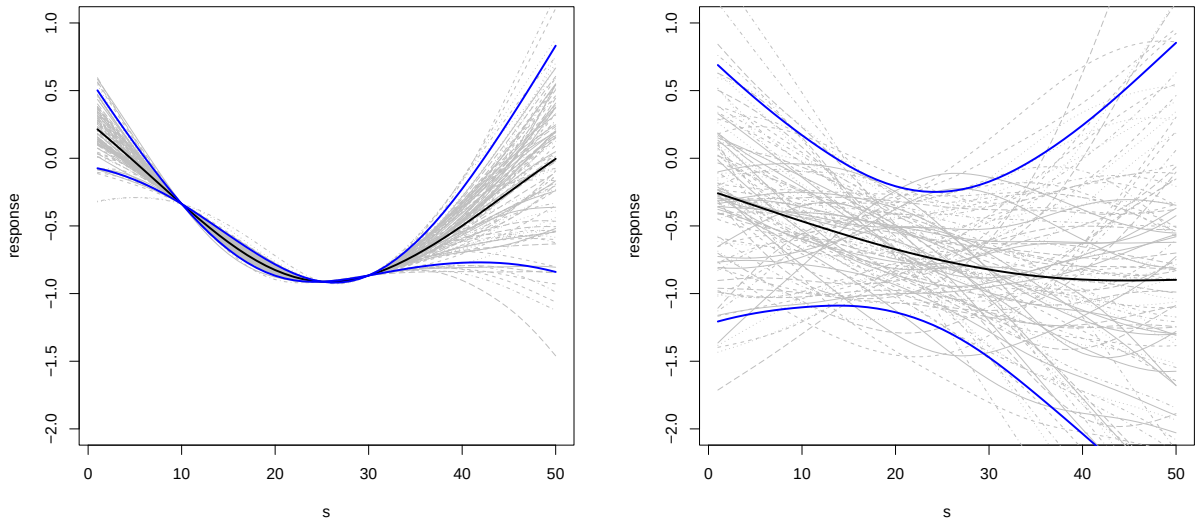
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$$

where $\mathbf{X}_i$ are the different realizations. Since we use the sample mean and variance, we use the test observer

$$T = \frac{X - \bar{X}}{\sqrt{\mathrm{Var}(\bar{X})(1 - \frac{1}{n})}}$$

which is t-distributed with 99 degrees of freedom. The 90% prediction interval is then

$$\bar{X} \pm 1.660\sqrt{\mathrm{Var}(\bar{X})(1 - \frac{1}{n})}$$



(a) $\sigma_N^2 = 0$ (b) $\sigma_N^2 = 0.25$

Figure 7: 100 realizations with estimated prediction and estimated prediction intervals

The empirically obtained predictor and prediction intervals seem to coincide well with the analytical results in Figure 6. Since $\mathbf{X}_i \sim N(\hat{X}, \text{Cov}(\hat{X} - \mathbf{X}))$, for $i = 1, ..., n$, we know that

$$\lim_{n\to\infty} \bar{X} = \hat{X}$$
$$\lim_{n\to\infty} T = Z \sim N(0, 1)$$

i.e. when $n \to \infty$, we obtain the analytical prediction and prediction variance.

Approximating the area under X

We define an approximation to the area under X and above the line 2 as

$$A = \sum_{s=1}^{50} \mathbb{I}(X(s) > 2)(X(s) - 2)$$

where $\mathbb{I}$ is the indicator function. One way to estimate A would be to use the 100 realizations above, and calculate the mean. Denote this estimate by $\hat{A}$, which takes the form

$$\hat{A} = \frac{1}{100} \sum_{i=1}^{100} \sum_{s=1}^{50} \mathbb{I}(\mathbf{X}_i(s) > 2)(\mathbf{X}_i(s) - 2) = \frac{1}{100} \sum_{i=1}^{100} A_i$$

where $\mathbf{X}_i$ is the vector containing one realization. We find that $\hat{A} = 0$.

We can compare this estimator to another predictor, which takes the form

$$\tilde{A} = \sum_{s=1}^{50} \mathbb{I}(\hat{\mathbf{X}}(s) > 2)(\hat{\mathbf{X}}(s) - 2)$$

where $\hat{\mathbf{X}}$ is the Simple Kriging predictor in Equation (3).

We find that $\tilde{A} = 0$. Both estimates are 0, which is not surprising, as none of the realizations in Figure 6a seem to go above the line 2. However, we can compare the two predictors of A by way of Jensen's inequality, which states that

$$\mathrm{E}(\phi(X)) \geq \phi(\mathrm{E}(X))$$

where ϕ is a convex function. We use that $A(X)$ is a convex function. Then,

$$\mathrm{E}(A(X)) \approx \frac{1}{n}\sum_{i=1}^{n} A_i(X) = \hat{A} \geq A(\mathrm{E}(X)) = A(\hat{X}) = \tilde{A}$$

so one expects that $\hat{A} \geq \tilde{A}$.

Summary of experiences

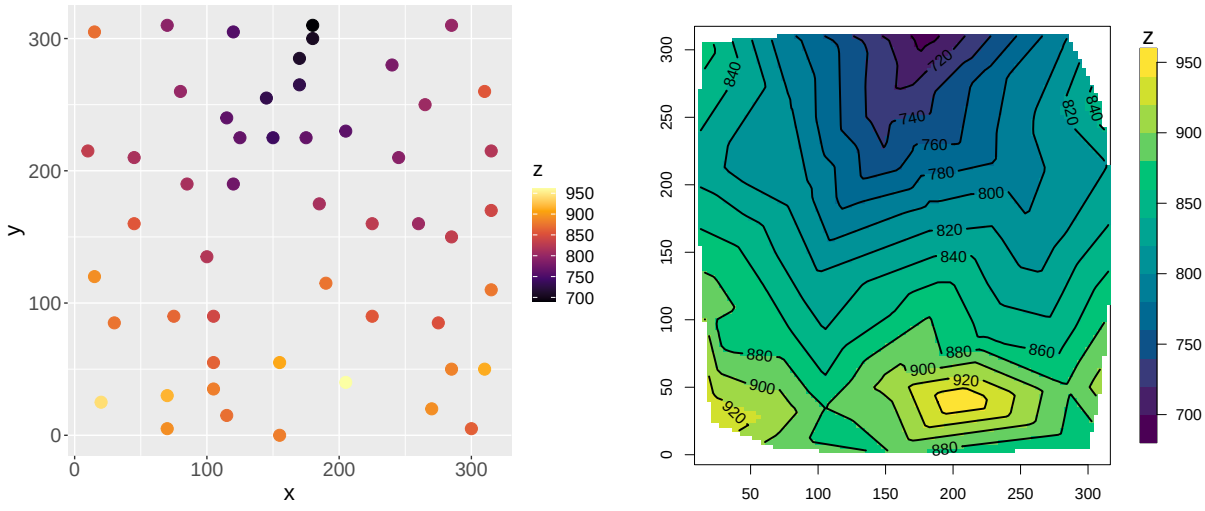
We present here some of the experiences we have made during part 1 of the project.

- Notation can be very difficult to navigate, but it is essential to be consistent and use it properly, to avoid as much confusion as possible.
- The model parameters seem to control the realizations to a very large degree - the smoothness of the covariance function especially.

Part 2 - real data

In this part, we consider a data set consisting of observations of terrain elevation in 52 locations in $\mathbb{R}^2$. We begin with some data visualization.

Data visualization



(a) Scatterplot of data points

(b) Bivariate interpolation

Figure 8: Correlation functions

In Figure 8a, we plot the data points on a grid $[0, 315]^2$. In Figure 8b, a linear interpolation of the data points is shown, along with the estimated equidistance curves. We notice in b) that the elevation seems to be increasing towards lower values on the y-axis, which might indicate non-stationarity (non-constant mean function).

Universal kriging predictor

Let $\mathcal{D} = [0, 315]^2 \subset \mathbb{R}^2$. Consider the GRF X on $\mathcal{D}$, modelled by

$$\begin{aligned} \mathrm{E}(X(\mathbf{s})) &= \mathbf{g}(\mathbf{s})^T\boldsymbol{\beta}, \quad \mathbf{s} \in \mathcal{D} \\ \mathrm{Var}(X(\mathbf{s})) &= \sigma^2, \quad \mathbf{s} \in \mathcal{D} \\ \mathrm{Corr}(X(\mathbf{s}), X(\mathbf{s}')) &= \rho(\|\mathbf{s} - \mathbf{s}'\|), \quad \mathbf{s}, \mathbf{s}' \in \mathcal{D} \end{aligned}$$

where $\mathbf{g}(\mathbf{s}) = (1, g_2(\mathbf{s}), \ldots, g_{n_g}(\mathbf{s}))^T$ is a vector of known spatial variables, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n_g})^T$ is a vector of unknown parameters.

Let $\mathbf{s}_0 \in \mathcal{D}$ be an arbitrary unobserved location, and let $\mathbf{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))^T$ be the vector of observed locations. The best linear unbiased predictor $\hat{X}_0$ of $X(\mathbf{s}_0) = X_0$ has to satisfy three properties, namely

$$\begin{aligned} &1)\ \hat{X}_0 = \mathbf{a}^T\mathbf{X}, \quad \mathbf{a}^T \in \mathbb{R}^n \\ &2)\ \mathrm{E}(\hat{X}_0) = \mathrm{E}(X_0) \\ &3)\ \mathrm{MSE} = \mathrm{E}((X_0 - \hat{X}_0)^2) \text{ is minimized.} \end{aligned}$$

From 2), we get that

$$\begin{aligned} &\mathrm{E}(\hat{X}_0) = \mathrm{E}(X_0) \\ \iff &\mathbf{a}^T\left(\mathbf{g}(\mathbf{s}_1)^T\boldsymbol{\beta}, \ldots, \mathbf{g}(\mathbf{s}_n)^T\boldsymbol{\beta}\right) = \mathbf{g}(\mathbf{s}_0)^T\boldsymbol{\beta} \\ \iff &\mathbf{a}^T\mathrm{G} = \mathbf{g}(s_0)^T \end{aligned}$$

where G is the $(n \times n_g)$ matrix $\mathrm{G} = (\mathbf{g}(\mathbf{s}_1)^T, \ldots, \mathbf{g}(\mathbf{s}_n)^T)^T$. From 3), we find the objective function to minimise:

$$\begin{aligned} &\mathrm{E}((X_0 - \hat{X}_0)^2) = \mathrm{E}(X_0^2 - 2X_0\hat{X}_0 + \hat{X}_0^2) \\ &= \mathrm{Var}(X_0) - 2\mathbf{a}^T\mathrm{Cov}(X_0, \mathbf{X}) + \mathbf{a}^T\mathrm{Var}(\mathbf{X})\mathbf{a} \\ &= \sigma^2 - 2\mathbf{a}^T\mathbf{c} + \mathbf{a}^T\Sigma\mathbf{a} \end{aligned}$$

where $\mathbf{c} = \mathrm{Cov}(X_0, \mathbf{X})$ and $\Sigma = \mathrm{Var}(\mathbf{X})$. The optimisation problem becomes

$$\begin{aligned} \min_{\mathbf{a} \in \mathbb{R}^n} \ &(\sigma^2 - 2\mathbf{a}^T\mathbf{c} + \mathbf{a}^T\Sigma\mathbf{a}) \\ \text{s.t. } &\mathbf{a}^T\mathrm{G} = \mathbf{g}(\mathbf{s}_0) \end{aligned}$$

which can be solved by method of Lagrange multipliers. The solution to the optimization problem is

$$\begin{aligned} \hat{X}_0 &= \mathbf{g}(\mathbf{s}_0)^T\hat{\boldsymbol{\beta}} + \mathbf{c}^T\Sigma^{-1}(\mathbf{X} - \mathrm{G}\boldsymbol{\beta}), \text{ where} \\ \hat{\boldsymbol{\beta}} &= (\mathrm{G}^T\Sigma^{-1}\mathrm{G})^{-1}\mathrm{G}^T\Sigma^{-1}\mathbf{X} \end{aligned}$$

and the predictor variance becomes

$$\mathrm{Var}(X_0 - \hat{X}_0) = \sigma^2 - \mathbf{c}^T\Sigma^{-1}\mathbf{c} + (\mathbf{g}(\mathbf{s}_0) - \mathrm{G}\Sigma^{-1}\mathbf{c})^T(\mathrm{G}^T\Sigma^{-1}G)^{-1}(\mathbf{g}(\mathbf{s}_0) - \mathrm{G}\Sigma^{-1}\mathbf{c})$$

The marginal variance σ^2 does not change with a different parameterization of the expectation function. However, the prediction variance does, as it depends on the matrix G.

Ordinary Kriging

We consider the case of ordinary Kriging, namely

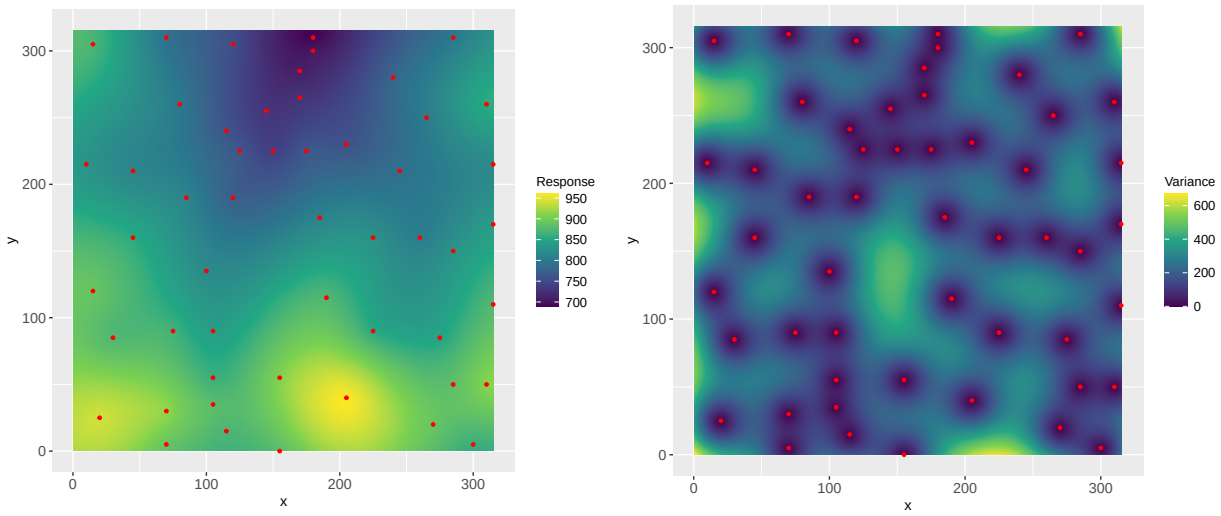
$$\mathrm{E}(X(\mathbf{s})) = \beta_1, \ \mathbf{s} \in \mathcal{D}$$

As the covariance function, we will use

$$\rho(h) = \sigma^2 \exp(-(0.01h)^{1.5}), \ h \in [0, \infty)$$

i.e. the powered exponential with $a = 100$ and $\alpha = 1.5$. We set the marginal variance to be $\sigma^2 = 2500$. We use the package `geoR` with the function `krige.conv` to calculate the predictions and prediction variance, for both Ordinary and Universal Kriging. In Figure 9, the Ordinary Kriging predictor is shown (left), along with the prediction variance (right).

```
## krige.conv: model with constant mean
## krige.conv: Kriging performed using global neighbourhood
```



(a) Prediction

(b) Prediction variance

Figure 9: Ordinary kriging. Data points are red

We see that the prediction variance approaches zero as the prediction approaches the known data points. The variance increases in areas where there are few data points nearby.

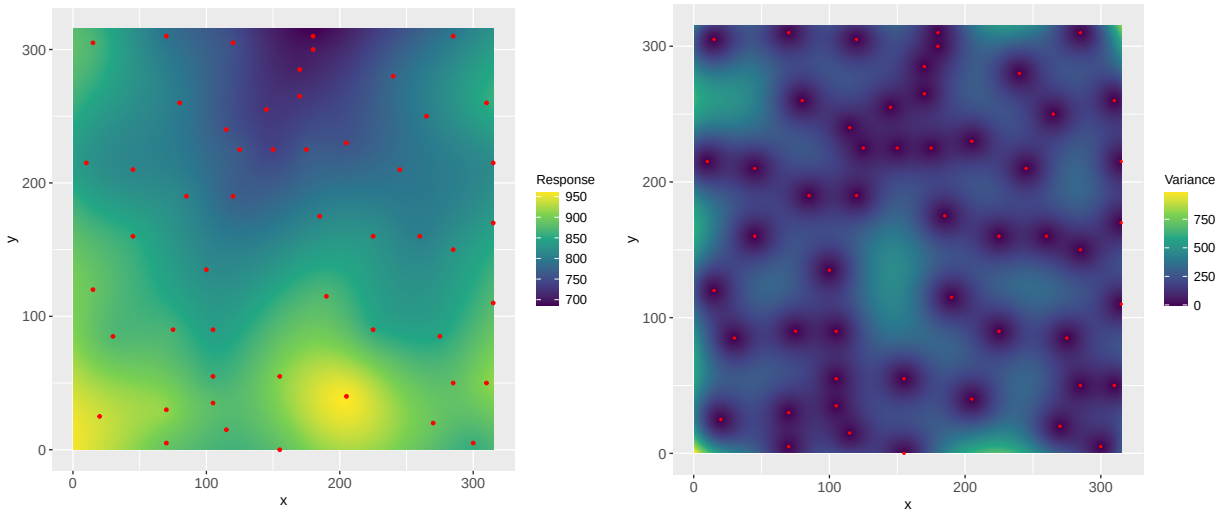
Universal Kriging

Denote $\mathbf{s} = (s_1, s_2) \in \mathcal{D}$, and define $\mathbf{g}(\mathbf{s})$ to contain all polynomials $s_1^k s_2^l$ for $(k, l) \in \{(0,0), (1,0), (0,1), (1,1), (2,0), (0,2)\}$, such that

$$\mathbf{g}(\mathbf{s}) = (1, s_1, s_2, s_1 s_2, s_1^2, s_2^2)$$
$$\mathrm{E}(X(\mathbf{s})) = \mathbf{g}(\mathbf{s})^T \boldsymbol{\beta} = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1 s_2 + \beta_4 s_1^2 + \beta_5 s_2^2$$

In Figure 10, the Universal Kriging predictor is shown with the prediction variance, for the mean structure specified above.

```
## krige.conv: model with mean given by a 2nd order polynomial on the coordinates
## krige.conv: Kriging performed using global neighbourhood
```



(a) Prediction

(b) Prediction variance

Figure 10: Universal kriging. Data points are red

The Universal Kriging predictor with the polynomial mean structure seems to produce results very similar to that of Ordinary Kriging. At first glance in Figure 10b, the variance might seem to be lower, but this is because the color scale range is higher, i.e. the variance range in the Universal Kriging predictor is bigger.

The estimated coefficients $\boldsymbol{\beta}$ are:

```
##         beta0         beta1         beta2         beta3         beta4
##  9.416681e+02 -1.109082e+00 -5.181555e-02  3.005938e-03 -8.672447e-04
##         beta5
##  1.326022e-04
```

The first coefficient β_0 is much larger than the rest of the coefficients, which indicates that the mean structure seems to be mostly constant and not so much dependent on the spatial structure. In that sense, Universal Kriging gives a prediction which is similar to that of Ordinary Kriging.

Elevation in the location $\mathbf{s}_0 = (100, 100)$

We consider the location $\mathbf{s}_0 = (100, 100)$. Using the ordinary Kriging predictor (which is Gaussian, since X is a GRF) found previously, with its associated prediction variance, we can calculate the probability of the

elevation being higher than 850 meter at this location.

$$\begin{aligned} \mathrm{P}(\hat{X}(\mathbf{s}_0) > 850) &= \mathrm{P}\left(Z > \frac{850 - \beta_1}{\sqrt{\mathrm{Var}(\hat{X}(\mathbf{s}_0))}} \right) \\ &= 1 - \mathrm{P}\left(Z \leq \frac{850 - \beta_1}{\sqrt{\mathrm{Var}(\hat{X}(\mathbf{s}_0))}} \right) \end{aligned}$$

where Z is standard normal. Using the `pnorm` function in R, we find

$$\mathrm{P}(\hat{X}(\mathbf{s}_0) > 850) = 0.3949761$$

In addition, we can find the elevation for which the probability is 90% that the true elevation lies below 850 meters. Denote this elevation by Y. We use the inverse cumulative distribution function, by solving

$$\begin{aligned} \mathrm{P}(\hat{X}(\mathbf{s}_0) < Y) = 0.9 &\Leftrightarrow F_{\hat{X}}(Y) = 0.9 \\ &\Leftrightarrow Y = F_{\hat{X}}^{-1}(0.9) \end{aligned}$$

Using the `qnorm` function in R, we find

$$Y = 868.1109, \ [\text{meters}]$$

Summary of experiences

- In Figure 8b, we saw some indications that the mean structure might not be constant, and so Ordinary Kriging might not be the best choice. However, Universal Kriging gave coefficient estimates which indicated that the mean was mostly constant. The takeaway here is that interpreting 2D data by eye can be difficult and must be done with care.

Problem 3: Parameter estimation

We consider the stationary GRF $\{X(s); s \in \mathcal{D} = [0, 30]^2 \subset \mathbb{R}\}$ with

$$\mathrm{E}[X(s)] = 0, \quad s \in \mathcal{D} \tag{4}$$

$$\mathrm{Var}[X(s)] = \sigma^2 = 2^2, \quad s \in \mathcal{D} \tag{5}$$

$$\mathrm{Corr}[X(s), X(s')] = \exp(-\|s - s'\|/a), \quad s, s' \in \mathcal{D}. \tag{6}$$

Define the grid $\tilde{\mathcal{D}} = \{1, 2, \ldots, 30\}^2$. A realization can be generated on $\tilde{\mathcal{D}}$ by generating a vector of iid standard normal variables and multiplying with the Cholesky factor of the covariance matrix.

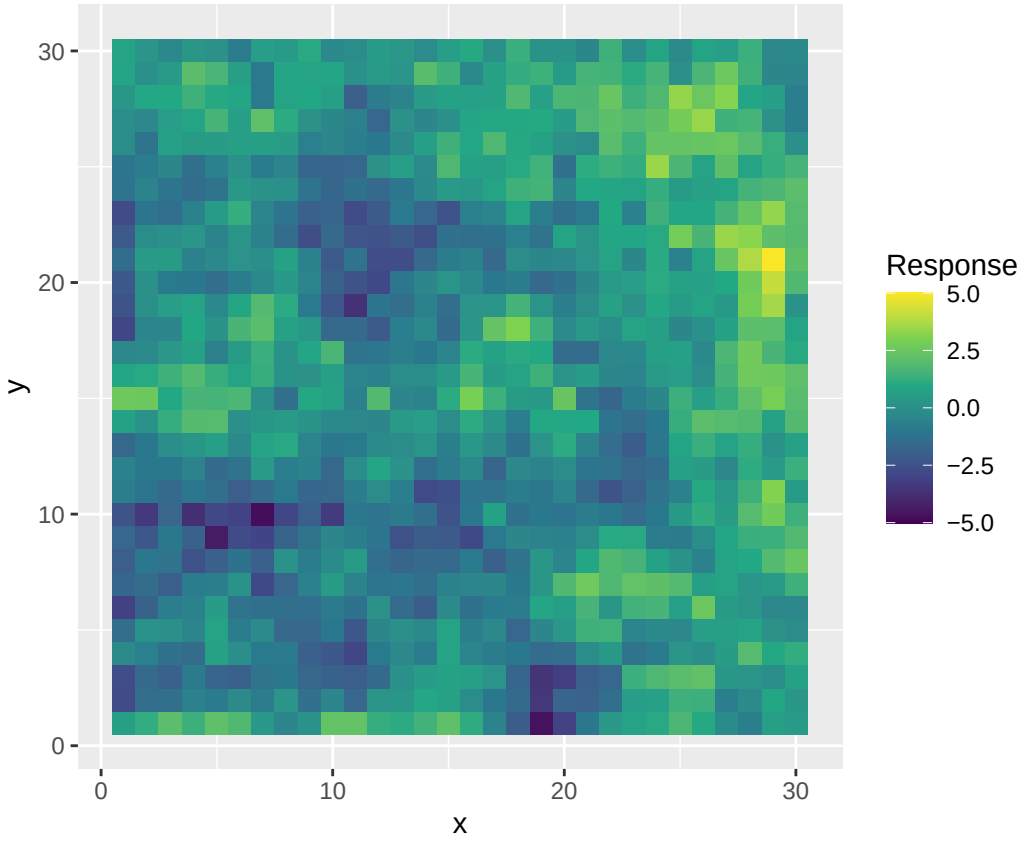


Figure 11: A realization of the random field defined in 3a)

Using the realization computed in 3a) as an exact observation, we can plot the empirical variogram and compare it to the exact variogram.

```
## variog: computing omnidirectional variogram
```

Figure 12: Exact variogram plotted against the computed empirical variogram of a single realization

We see that the empirical variogram deviates a little from the true variogram, especially for long ranges. The variogram range is 10% of the side lengths of the domain, which could lead to simulations with a slightly different correlation structure than we specified. If we decreased our variogram range (and refined our simulation grid correspondingly), this would become less of a problem.

We repeat this the previous process for 3 different realizations.

Figure 13: 3 realizations with their empirical semivariograms

We again see that the correlation structure of our simulations is different than what we expect.

We draw 36 locations uniformly from $\tilde{\mathcal{D}}$.

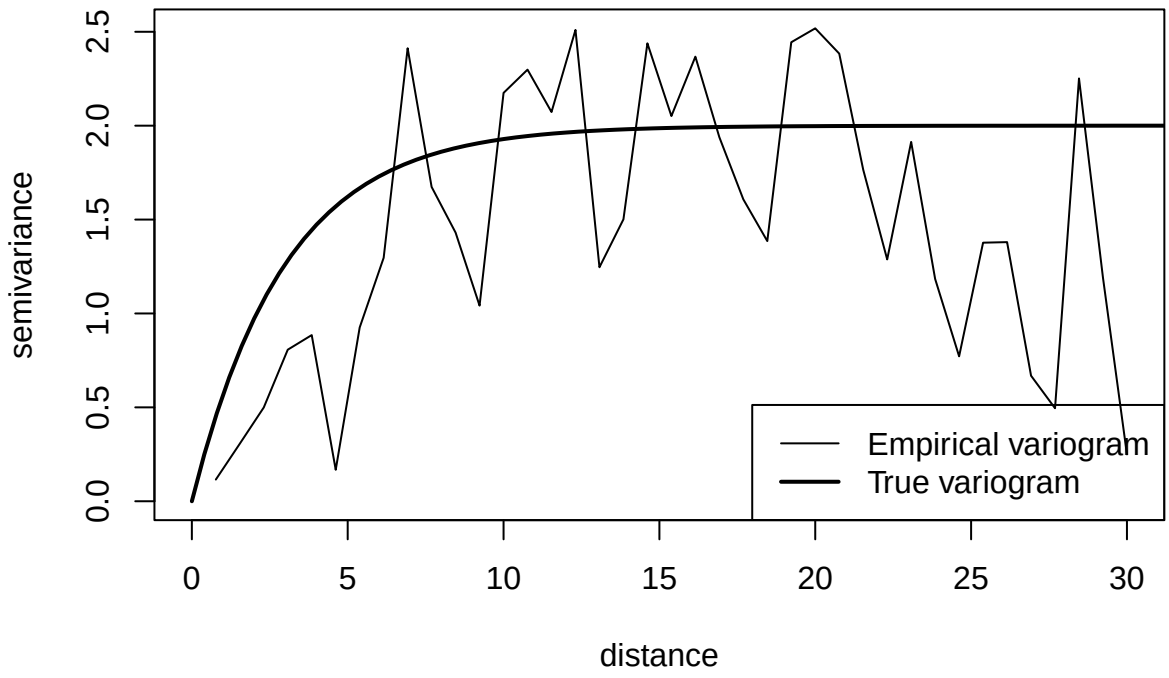


Figure 14: Semivariogram with 36 observations with their empirical semivariograms

We see that the variance of the semivariogram estimator is much higher when we have little data. In a) to c), we had a dataset containing 900 observations of the GRF, now we only have 36. This means the uncertainty of the semivariogram estimator is much higher for each distance bin than it was when we had a full observation on $\tilde{\mathcal{D}}$.

We now compare the empirical semivariograms computed with the full observation and the 36 sample locations by plotting them in the same plot, and we also try to estimate the parameters σ^2 and a of the covariance function.

The estimated σ^2 and a with 36 observations were

```
ml.small$cov.pars
```

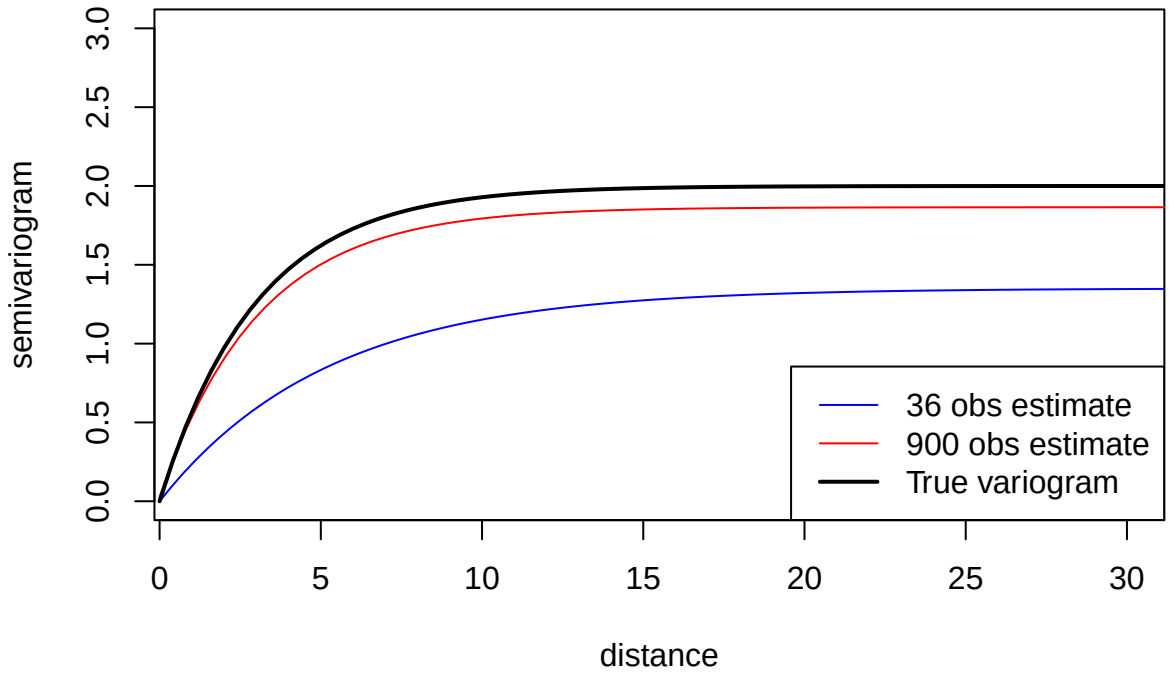
```
## [1] 1.350457 5.211823
```

The estimated σ^2 and a with 900 observations were

```
ml.big$cov.pars
```

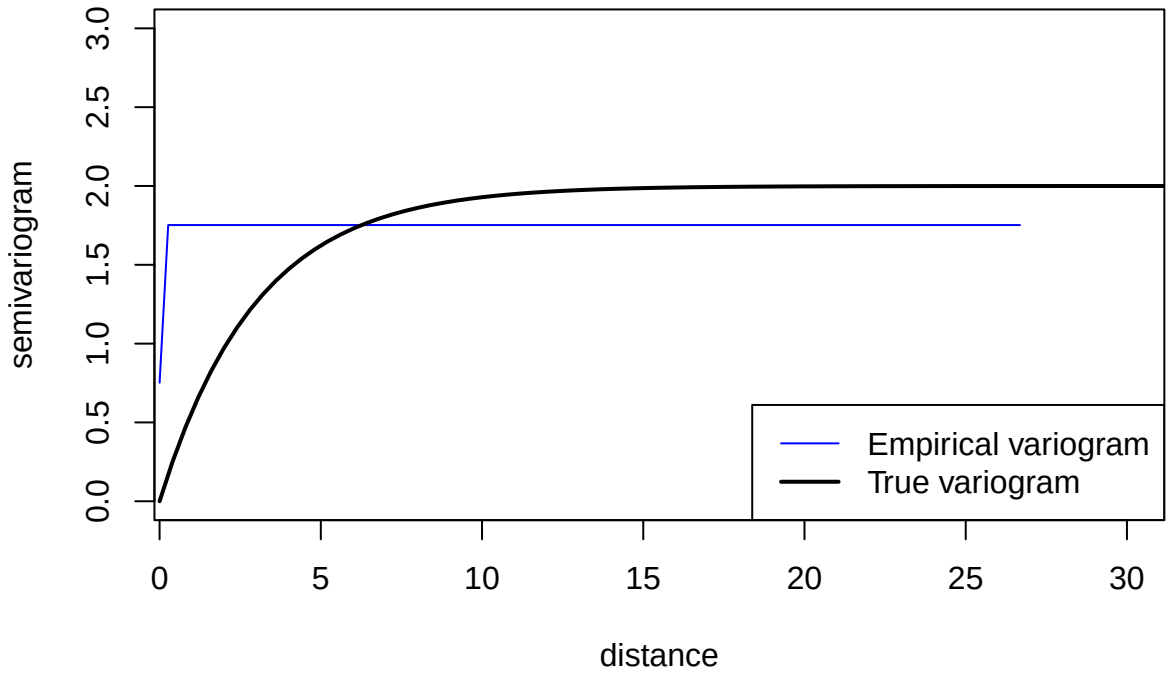
```
## [1] 1.838253 3.079753
```

With 900 observations, the ML estimate of σ^2 and a are obviously much more accurate. We can also visualize the variograms these estimated parameters describe, and see that the estimate based on the full grid observation is closer to the true variogram than the 36 observation one.



We repeat this process for 9 and 100 locations randomly chosen from the discretized domain.

We first handle the case with 9 sampling locations.



We immediately see that something went wrong in the maximum likelihood estimation. Probably, we simply did not have enough data to converge.

The computed parameters were

```
res$sigma2
```

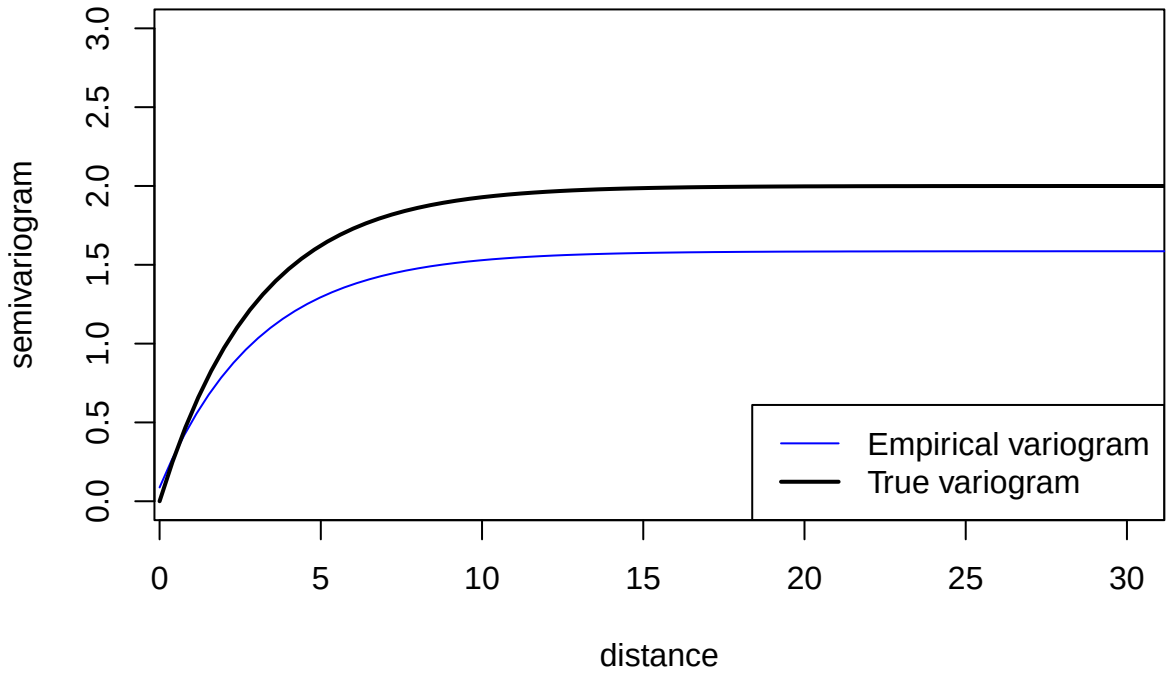
```
## [1] 0
```

```
res$a
```

```
## [1] 0
```

Again, we see that the algorithm did not converge.

The case with 100 locations is handled similarly.



Now we see that the ML algorithm converges, and that it seems like it produces reasonable estimates to the true model parameters.

The computed parameter estimates were

```
res$sigma2
```

```
## [1] 1.49849
```

```
res$a
```

```
## [1] 3.053859
```

This confirms that the model parameters were estimated properly

In summary, we can say that empirical variograms can provide an indication of the true model parameters, but they can also be highly unreliable. This is true especially of cases with little data, but the estimates can also be relatively poor even when we have access to many observations.

When using a model based maximum likelihood approach to variogram estimation, we also need to be cautious of the convergence of the maximum likelihood algorithm. This becomes less of an issue when the amount of data increases.