TMA4250 Spatial Statistics: Project 1

Christian Moen, Jim Totland

February 2022

Problem 1: GRFs – model characteristics

We consider the stationary Gaussian random field (GRF) X on $\mathcal{D} = [1, 50] \subset \mathbb{R}$ and assume that

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

where ρ is the correlation function of X. We discretize $\mathcal{D}$ by a regular grid $\tilde{\mathcal{D}} = \{1, 2, ..., 50\}$ and let $\boldsymbol{X} = (X(1), \ldots, X(50))^T$ be the discretization of X on $\tilde{\mathcal{D}}$.

We will consider two types of correlation functions: the powered exponential with power $\alpha \in \{1, 1.9\}$ and spatial scale $a = 10$, as well as Matérn with smoothness parameter $\nu \in \{1, 3\}$. The parametrizations used can be found in the lecture slides of this course. Furthermore, we will let the marginal variance $\sigma^2 \in \{1, 5\}$.

a)

The positive semi-definite (p.s.d.) property of the correlation function can be stated as follows. $\forall m \in \mathbb{Z}_+$, $\forall a_1, \ldots, a_m \in \mathbb{R}$ and $\forall \boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathcal{D}$, we have

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \rho(\boldsymbol{s}_i, \boldsymbol{s}_j) \geq 0.$$

We must require that the correlation function is p.s.d., since

$$\begin{aligned} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \rho(\boldsymbol{s}_i, \boldsymbol{s}_j) &= \sigma^{-2} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j c(\boldsymbol{s}_i, \boldsymbol{s}_j) \\ &= \sigma^{-2} \mathrm{Var}\left[\sum_{i=1}^{m} a_i X(\boldsymbol{s_i}) \right], \end{aligned}$$

where c is the covariance function. The variance must be greater than or equal to zero, which means that the correlation function must be p.s.d.

The four possible correlation functions are plotted in figure 1. Since we consider a second-order stationary random field, X, we know that it is quadratic mean continuous (q.m.) on $\mathcal{D}$ if and only if the stationary covariance function $C(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = 0$. This is certainly the case for the covariance functions considered here. Moreover, if C is $2k$ times differentiable at $\boldsymbol{h} = 0$, then X is k times q.m. differentiable. The differentiability of the realizations resulting from the different parameter combinations will become clearer in b).

Next, we consider the semi-variogram, defined as

$$\gamma(\boldsymbol{h}) = \frac{1}{2} \mathrm{Var}[X(\boldsymbol{h}) - X(0)] = C(\boldsymbol{0}) - C(\boldsymbol{h}).$$

The function 2γ is called the variogram. The eight possible semi-variograms are displayed in figure 2.

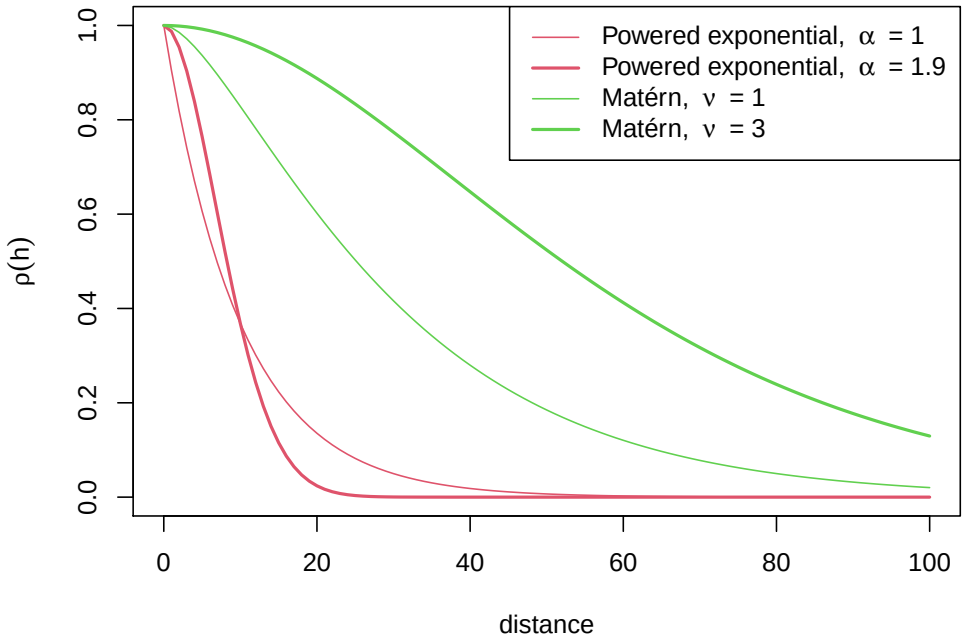


Figure 1: The correlation functions of the two covariance models plotted as a function of distance, h, for the different parameter combinations.

b)

By the definition of a GRF, $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_X)$. The parameters are calculated from the mean- and covariance function of the GRF, such that $\boldsymbol{\mu} = \mathbf{0}$ and $(\Sigma_X)_{ij} = \sigma^2\rho(\|i-j\|)$. We simulate four realizations of X on $\tilde{\mathcal{D}}$ for the two covariance models and their respective parameter combinations. The result is displayed in figure 3.

Unsurprisingly, we see that higher marginal variance leads to more variability in the realizations. When the power parameter of the powered exponential $\alpha = 1$, the realizations seem to be non-differential, i.e. the covariance function is not differentiable at $\boldsymbol{h} = 0$. When $\alpha = 1.9$, the realizations are fairly smooth. Considering the Matérn covariance model, a higher smoothness parameter, ν, leads to smoother realizations, as one would expect.

c)

We plan to observe Y_1, Y_2 and Y_3 at locations $s_1 = 10$, $s_2 = 25$ and $s_3 = 30$, respectively. The observation model is given by

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

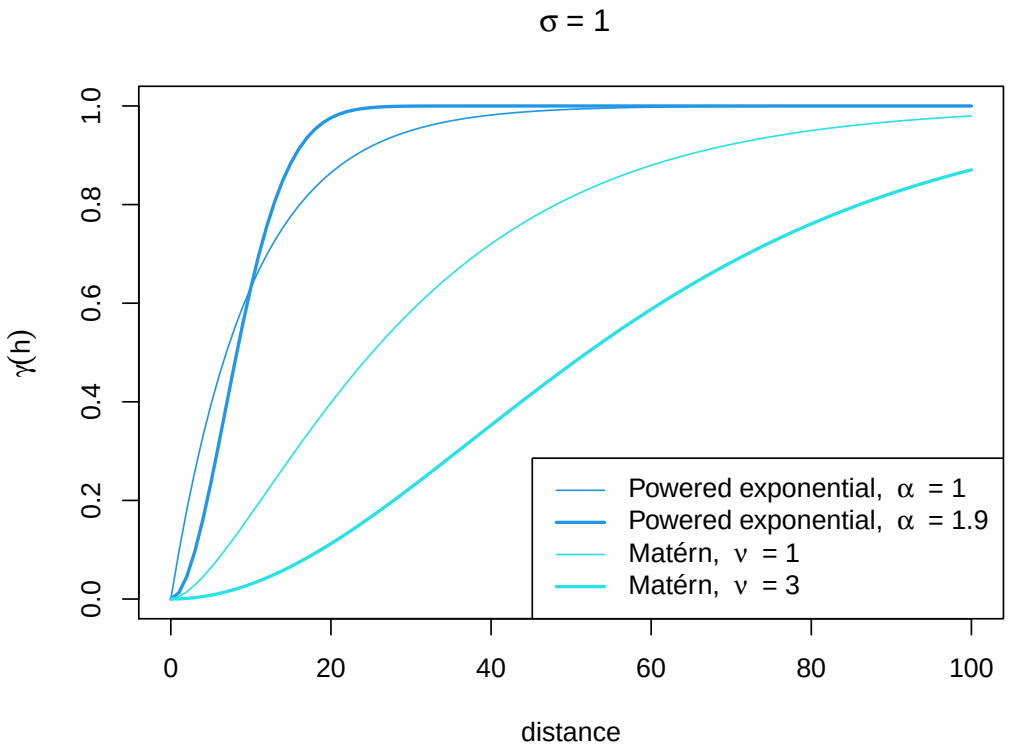
where $\varepsilon_1, \varepsilon_2, \varepsilon_3 \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_N^2)$ and independent of X. We let $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^T$. Let $\tilde{\boldsymbol{X}} = (X(s_1), X(s_2), X(s_3))^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$. Since $\boldsymbol{Y}$ is a linear combination of multivariate normal distributions, it is multivariate normally distributed itself, with mean

$$\mathrm{E}[\boldsymbol{Y}] = \mathrm{E}[\tilde{\boldsymbol{X}} + \boldsymbol{\varepsilon}] = \mathbf{0},$$

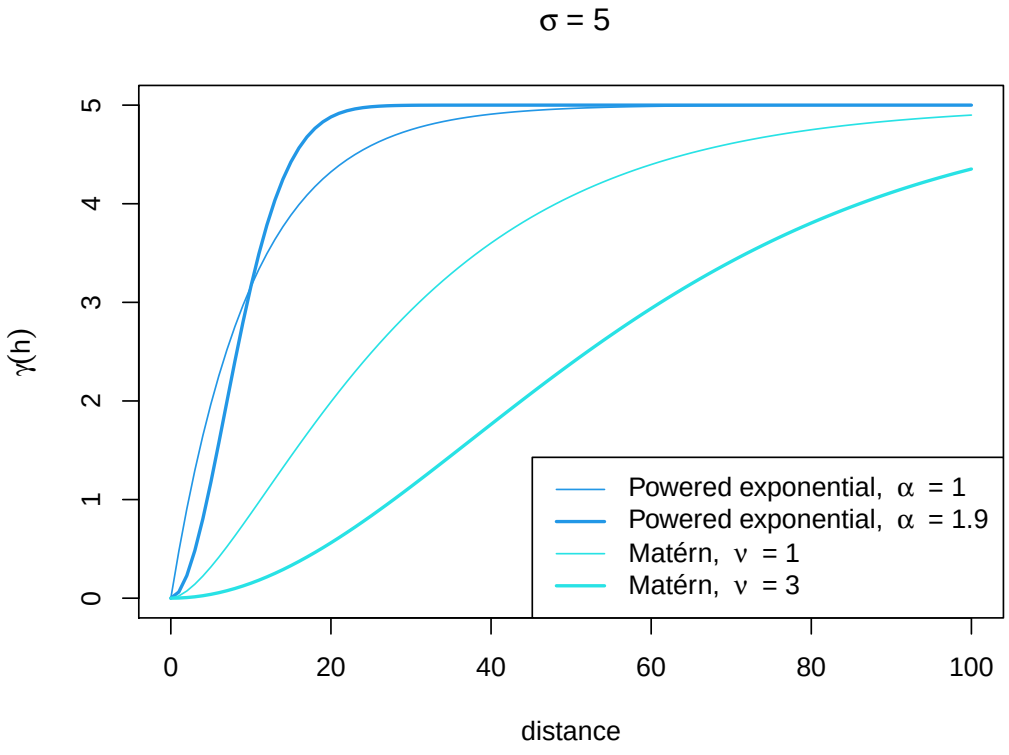
and covariance matrix

$$\begin{aligned}\Sigma_Y &:= \mathrm{Cov}(\boldsymbol{Y}) = \mathrm{Cov}(\tilde{\boldsymbol{X}} + \boldsymbol{\varepsilon}) \\ &= \mathrm{Cov}(\tilde{\boldsymbol{X}}) + \sigma_N^2 I_3 = H\Sigma_X H^T + \sigma_N^2 I_3,\end{aligned}$$

where H is a 3×50 matrix of zeros with the exceptions $H_{1,10} = H_{2,25} = H_{3,30} = 1$.

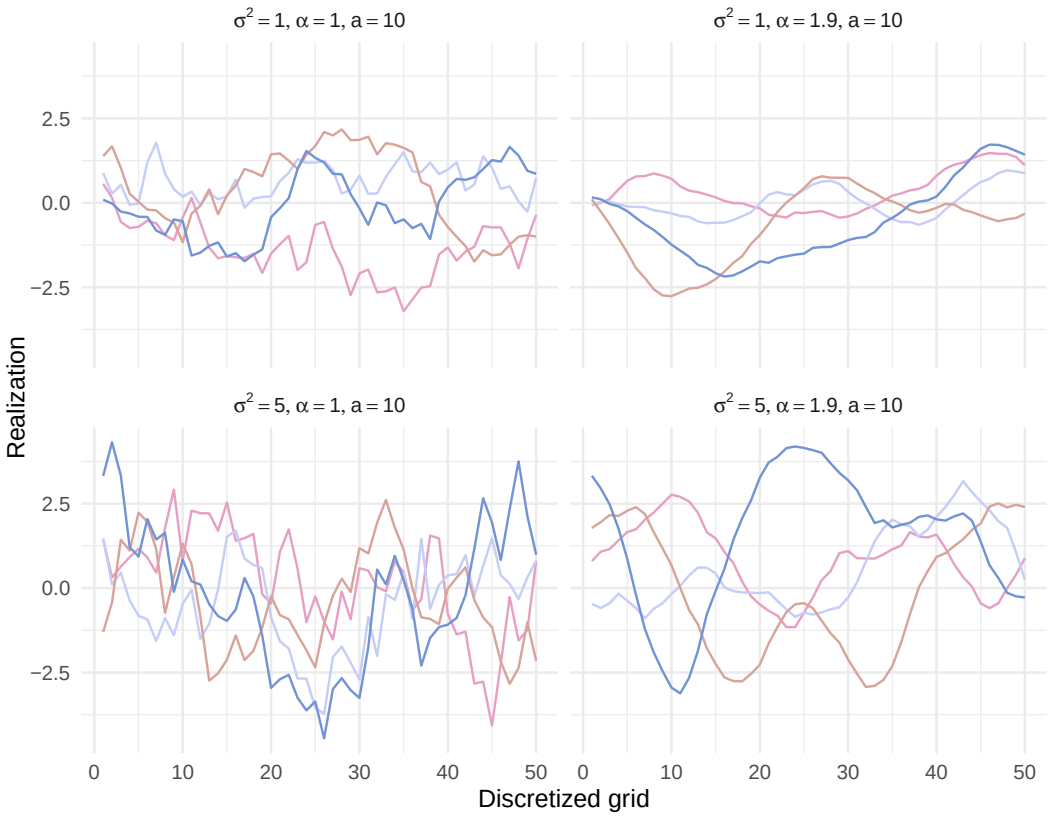


(a)

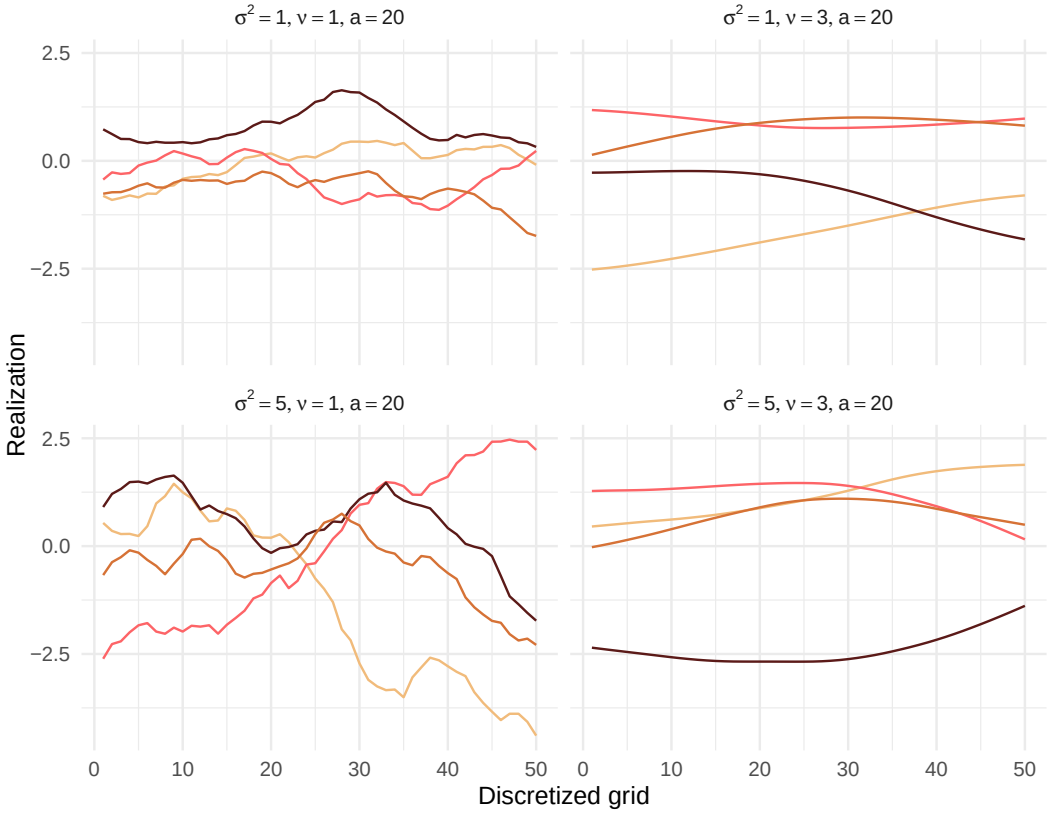


(b)

Figure 2: The semi-variograms plotted for each of the combinations of parameters and covariance models. (a) $\sigma^2 = 1$, (b) $\sigma^2 = 5$.



(a) Powered exponential



(b) Matérn

Figure 3: Four realizations for the covariance models (a) powered exponential and (b) Matérn for all the parameter combinations. The corresponding parameters are shown above each plot.

d)

We now consider $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$. To find its distribution and corresponding parameters, we construct

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim \mathcal{N}_{50+3}(\boldsymbol{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

Then, we know that $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_{X|Y}, \Sigma_{X|Y})$, where

$$\begin{aligned} \Sigma_{X|Y} &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}, \\ \boldsymbol{\mu}_{X|Y} &= \mathrm{E}[\boldsymbol{X}] + \Sigma_{XY}\Sigma_Y^{-1}\left(\boldsymbol{y} - \mathrm{E}[\boldsymbol{Y}]\right) = \Sigma_{XY}\Sigma_Y^{-1}\boldsymbol{y}. \end{aligned}$$

To predict $\boldsymbol{X}$, we will use the best linear unbiased prediction (BLUP), which is given by

$$\hat{\boldsymbol{X}} = \boldsymbol{\mu}_{X|Y},$$

and we will use the pointwise prediction variances, $\mathrm{diag}\left(\Sigma_{X|Y}\right)$, to construct confidence intervals.

Finally, we consider the Matérn covariance model with $\sigma^2 = 5$ and $\nu = 1$ and pick one realization from the simulation in b) to use as the observation $\boldsymbol{Y} = \boldsymbol{y}$. Let also $\sigma_N^2 \in \{0, 0.25\}$. The prediction $\hat{\boldsymbol{X}}$ is plotted with 90% confidence intervals in figure 4 for both values of σ_N^2. We observe that when $\sigma_N^2 = 0$ the confidence intervals have zero width at the observation points, i.e. there is no uncertainty there. Furthermore, we observe that the uncertainty grows as we get farther away from the observations, and that the width of the confidence interval is greatest in the right end point of the grid, which is the region with fewest observations close by.

e)

Now, we simulate from $\mathcal{N}(\boldsymbol{\mu}_{X|Y}, \Sigma_{X|Y})$ to create an empirical prediction with empirical prediction intervals. We simulate 100 realizations for both $\sigma_N^2 = 0$ and $\sigma_N^2 = 0.25$. The result is displayed in figure 5. When $\sigma_N^2 = 0$, all realizations pass through the observations, because there is no uncertainty in the observation model, exactly as we saw in d). When $\sigma_N^2 = 0.25$, the realizations do not have to pass through the observations, though the confidence intervals are narrowest at the observation points, since the uncertainty is minimal there.

f)

Let

$$A = \sum_{s \in \tilde{\mathcal{D}}} \mathbb{I}(X(s) > 2)(X(s) - 2)$$

where $\mathbb{I}$ is the indicator function. This is an approximation of the area under X and above level 2. We use the 100 realizations of $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$ with $\sigma_N^2 = 0$ generated in e) to create a prediction $\hat{A}$. This yields $\hat{A} = 1.801$ and a prediction variance $S_{\hat{A}}^2 = 19.113$.

Continuing with the assumption $\sigma_N^2 = 0$, we consider the alternative predictor

$$\tilde{A} = \sum_{s \in \tilde{\mathcal{D}}} \mathbb{I}(\hat{X}(s) > 2)(\hat{X}(s) - 2),$$

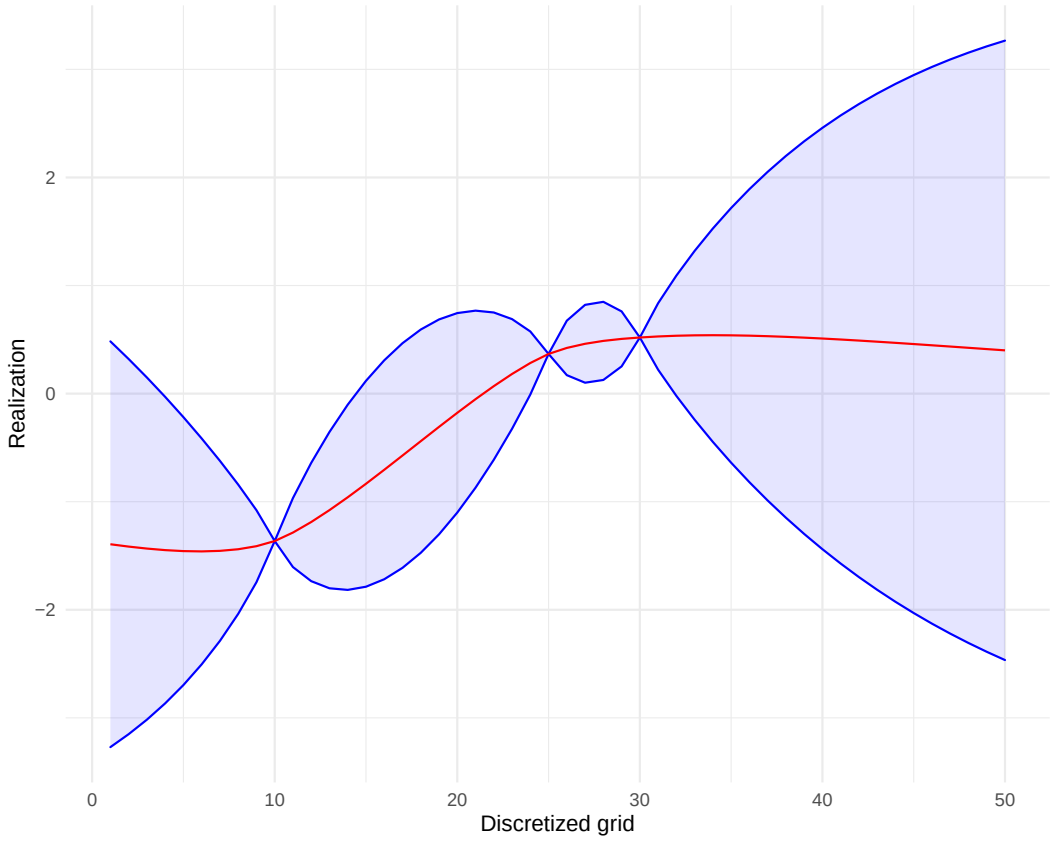
where $\hat{X}(s)$ is the simple Kriging predictor at location s. This corresponds to using the BLUP calculated in d), which result in $\tilde{A} = 0$. Comparing to figure 4, this is exactly what we would expect.

Jensen's inequality states that if the function ϕ is convex and Z is a random variable, then

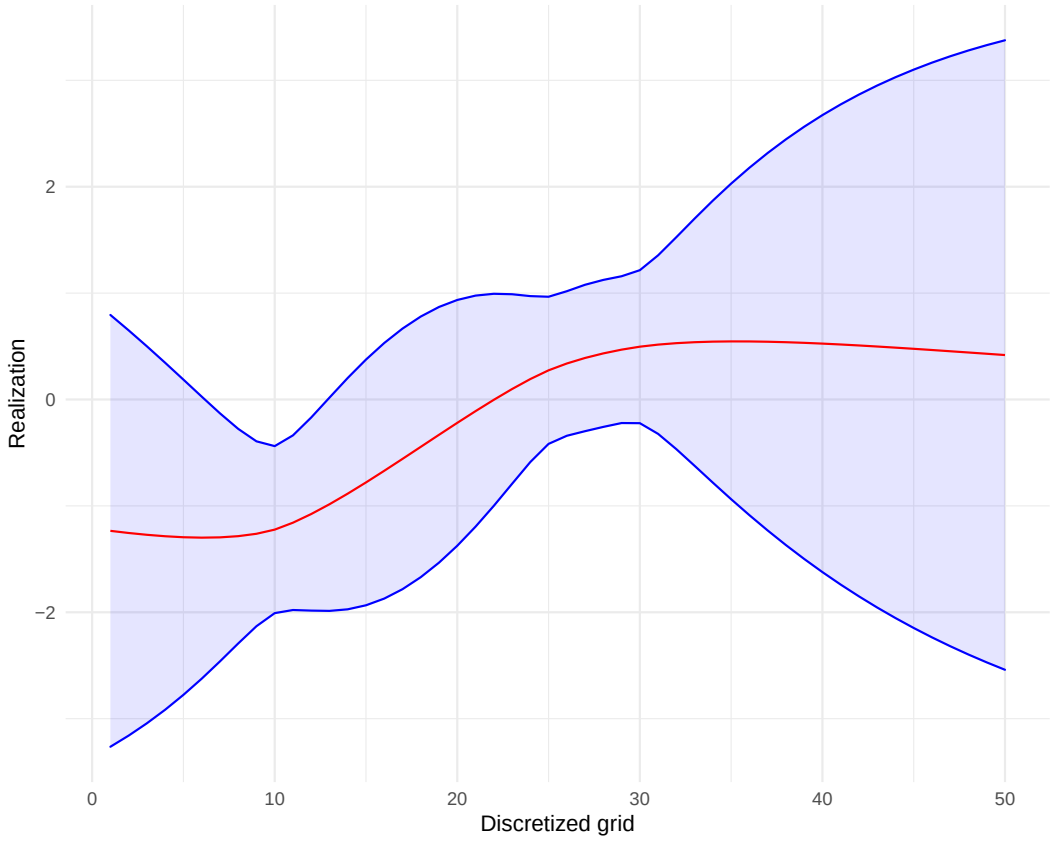
$$\phi(\mathrm{E}[Z]) \leq \mathrm{E}[\phi(Z)].$$

The function

$$f(z) = \mathbb{I}(z)(z - 2)$$

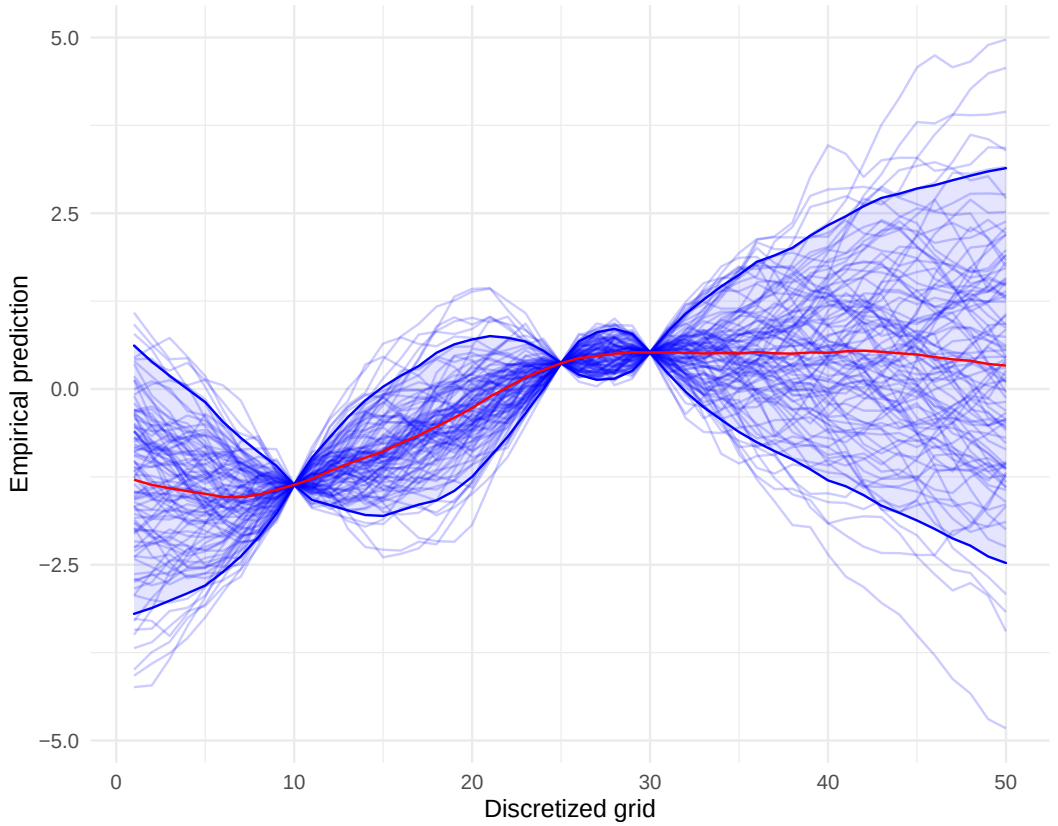


(a)

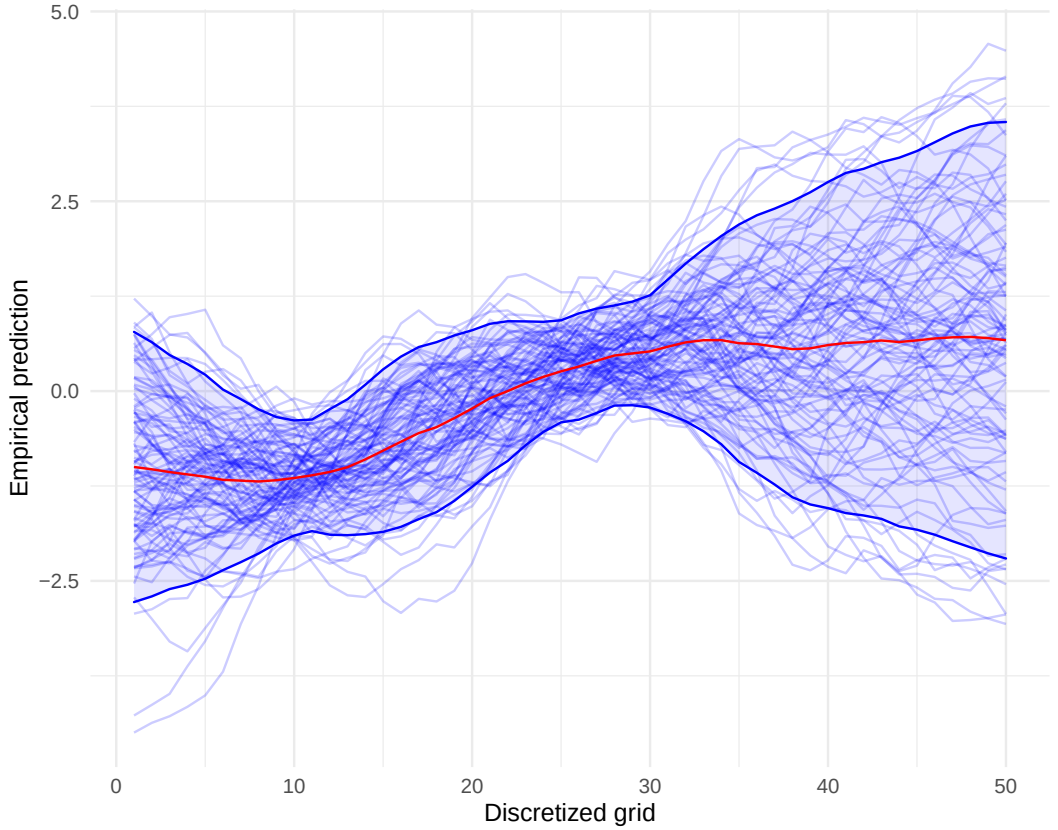


(b)

Figure 4: The BLUP of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ for (a) $\sigma_N^2 = 0$ and (b) $\sigma_N^2 = 0.25$.



(a)



(b)

Figure 5: The mean of 100 simulations of the posterior $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$ is plotted in red and the empirical 90% confidence intervals are plotted in dark blue for (a) $\sigma_N^2 = 0$ and (b) $\sigma_N^2 = 0.25$. All the realizations which have been used are plotted in a lighter shaded blue.

is easily verified to be convex, and since a sum of convex function is itself a convex, we conclude that $A(x)$ is a convex function. By using the 100 realizations of the posterior, we essentially have that $\hat{A} \approx \mathrm{E}[A(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y})]$. The simple Kriging predictor is equivalent to the expected value of the posterior, which means that $\tilde{A} = A(E[\boldsymbol{X}|\boldsymbol{Y}]) = \boldsymbol{y})$. Consequently, by Jensens's inequality one would expect that

$$\hat{A} \geq \tilde{A}.$$

We also note that the prediction variance of $\hat{A}$ is quite large, and it is thus unlikely that it will be exactly zero, which supports the result that $\hat{A} \geq \tilde{A}$.

g)

By considering a GRF in 1 dimension, we have seen how the choice of covariance function and its corresponding parameters are crucial for the realizations, especially with regard to smoothness and variability. We have also studied a simple observation model, which shows how Gaussian errors in the observations affect the generated realizations of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ and the predictions and confidence intervals, both empirically and using the BLUP.

Problem 2: GRF - real data

We consider a dataset of terrain elevations available in the file `topo.dat`. The dataset contains $n = 52$ observations located in the domain $\mathcal{D} = [0, 315]^2 \subset \mathbb{R}^2$. Let X be a GRF on $\mathcal{D}$ and let the vector of exact observations be $\boldsymbol{X} = (X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_{52}))^T$.

a)

To get an initial understanding of the data we visualize it by a contour, surface and scatter plot shown in figure 6. The contour plot 6a and surface plot 6b are generated using functions `image.plot` and `plotly` respectively, on the discrete grid $\tilde{\mathcal{D}} = \{1, 2, ..., 315\}^2$. The values for the functions are found by interpolation using the function `interp` on the observations $\boldsymbol{X}$. The scatter plot is generated by using the `scatterplot3d` to capture the linear trend of the data.

The contour and surface plots in figure 6 show a somewhat crude and unnatural terrain. There seem to be a decreasing trend in elevation in positive y direction which also is suggested by the linear plane in plot 6c. This suggest non-constant mean function of the GRF, hence, a stationary GRF model may not be suitable for these observations.

b)

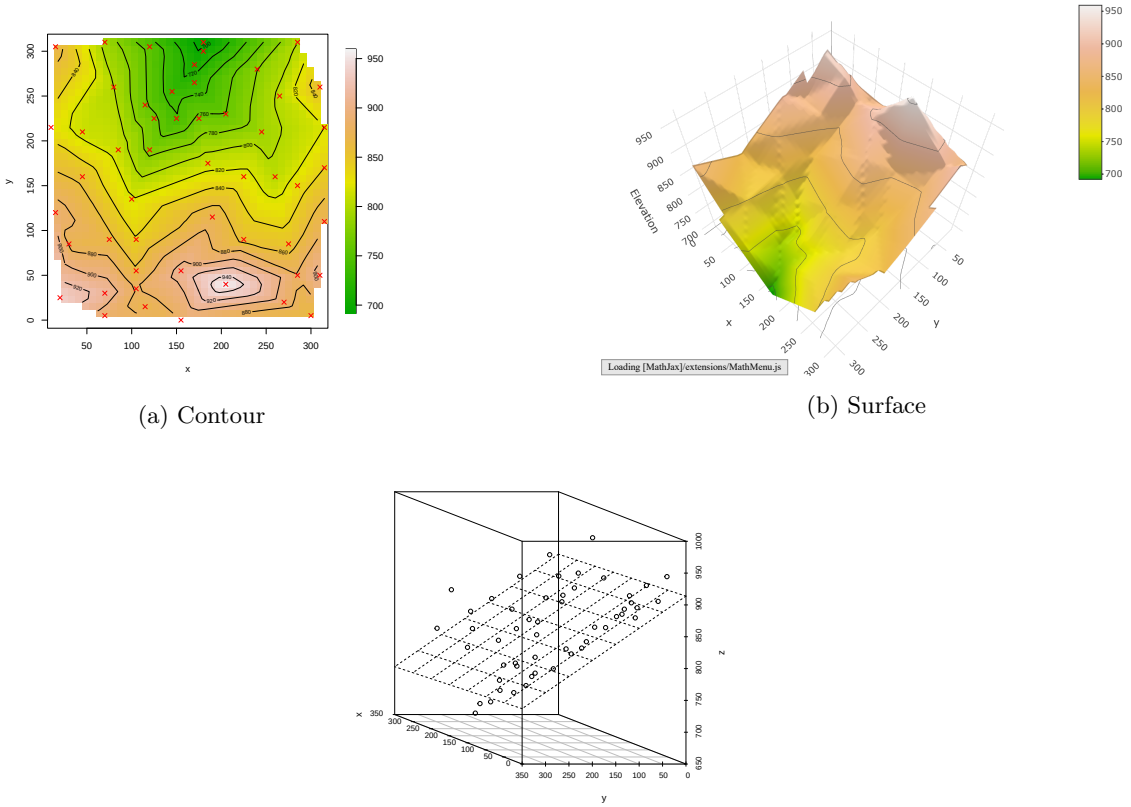
Previously we have seen the use of simple Kriging on fabricated data. Now we will consider universal Kriging on the dataset described in section 2a. Assume that the GRF is modeled by

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

where $\boldsymbol{g}(\boldsymbol{s}) = (1, g_2(\boldsymbol{s}), ..., g_{n_g}(\boldsymbol{s}))^T$ is a known vector of explanatory spatial variables for $\boldsymbol{s} \in \mathcal{D}$, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_{n_g})^T$ is the vector of unknown parameters. Let the marginal variance $\sigma^2 = 50^2$ and the exponentially powered isotropic correlation function $\rho(h) = \exp -(0.01h)^{1.5}$, where $h = ||\boldsymbol{s} - \boldsymbol{s}'|| \in [0, \infty)$ for $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{D}$. The goal is to find the best linear unbiased predictor of $X_0 = X(\boldsymbol{s}_0)$ at the unobserved location $\boldsymbol{s}_0$. More specifically, we want to find the $\boldsymbol{a} \in \mathbb{R}^n$ in the predictor $\hat{X}_0 = \boldsymbol{a}^T \boldsymbol{X}$ such that $\mathrm{E}[\hat{X}_0] = \mathrm{E}[X_0]$ and $\mathrm{MSE} = \mathrm{E}[(X_0 - \hat{X}_0)^2]$ is minimized. If we define

$$\begin{aligned}
\boldsymbol{z}_i &= (g_1(\boldsymbol{s}_i), ..., g_{n_g}(\boldsymbol{s}_i))^T, \quad i = 0, 1, \ldots, n, \\
Z &= [\boldsymbol{z}_1^T, ..., \boldsymbol{z}_n^T]^T,
\end{aligned}$$

we can write the unbiased-constraint as



(a) Contour

(b) Surface

(c) Linear plane

Figure 6: Interpolated contour and surface plot, (a) and (b), from observations $\boldsymbol{X(s)}$ and a linear plane fitted to said observations (c).

$$\mathrm{E}[\hat{X}_0] = \mathrm{E}[\boldsymbol{a}^T\boldsymbol{X}] = \boldsymbol{a}^T\mathrm{E}[\boldsymbol{X}] = \boldsymbol{a}^T Z\boldsymbol{\beta} = E[X_0] = \boldsymbol{z}_0^T\boldsymbol{\beta}$$
$$\implies \boldsymbol{a}^T Z = \boldsymbol{z}_0^T. \tag{1}$$

As for the MSE, we can write this as

$$\begin{aligned}
\mathrm{MSE} &= \mathrm{E}[(X_0 - \hat{X}_0)^2] \\
&= \mathrm{Var}[X_0 - \hat{X}_0] + \underbrace{E[X_0 - \hat{X}_0]^2}_{\text{Unbiased} \implies 0} \\
&= \mathrm{Cov}(X_0 - \hat{X}_0, X_0 - \hat{X}_0) \\
&= \mathrm{Cov}(X_0, X_0) + \mathrm{Cov}(\hat{X}_0, \hat{X}_0) - 2\mathrm{Cov}(\hat{X}_0, X_0) \\
&\overset{\hat{X}_0 = \boldsymbol{a}^T\boldsymbol{X}}{=} \sigma_0^2 + \boldsymbol{a}^T\Sigma\boldsymbol{a} - 2\boldsymbol{a}^T\boldsymbol{c},
\end{aligned} \tag{2}$$

where $\sigma_0^2 \coloneqq \mathrm{Var}[X_0]$, $\Sigma \coloneqq \mathrm{Var}[\boldsymbol{X}]$ and $\boldsymbol{c} \coloneqq \mathrm{Cov}(\boldsymbol{X}, X_0)$. Thus, we have the minimization problem

$$\begin{aligned}
&\min_{\boldsymbol{a}\in\mathbb{R}^n} \ \sigma_0^2 + \boldsymbol{a}^T\Sigma a - 2\boldsymbol{a}^T\boldsymbol{c}, \\
&\qquad \text{subject to} \quad \boldsymbol{a}^T Z = \boldsymbol{z}_0^T.
\end{aligned}$$

Solving the minimization problem yield the universal Kriging predictor

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (Z^T\Sigma^{-1}Z)^{-1}Z^T\Sigma^{-1}\boldsymbol{X} =: H\boldsymbol{X} \\
\hat{X}_0 &= \boldsymbol{z}_0^T\hat{\boldsymbol{\beta}} + \underbrace{\boldsymbol{c}\Sigma^{-1}(\boldsymbol{X} - Z\hat{\boldsymbol{\beta}})}_{\text{Simple Kriging}}
\end{aligned} \tag{3}$$

Next, we find an expression for the prediction variance, given by

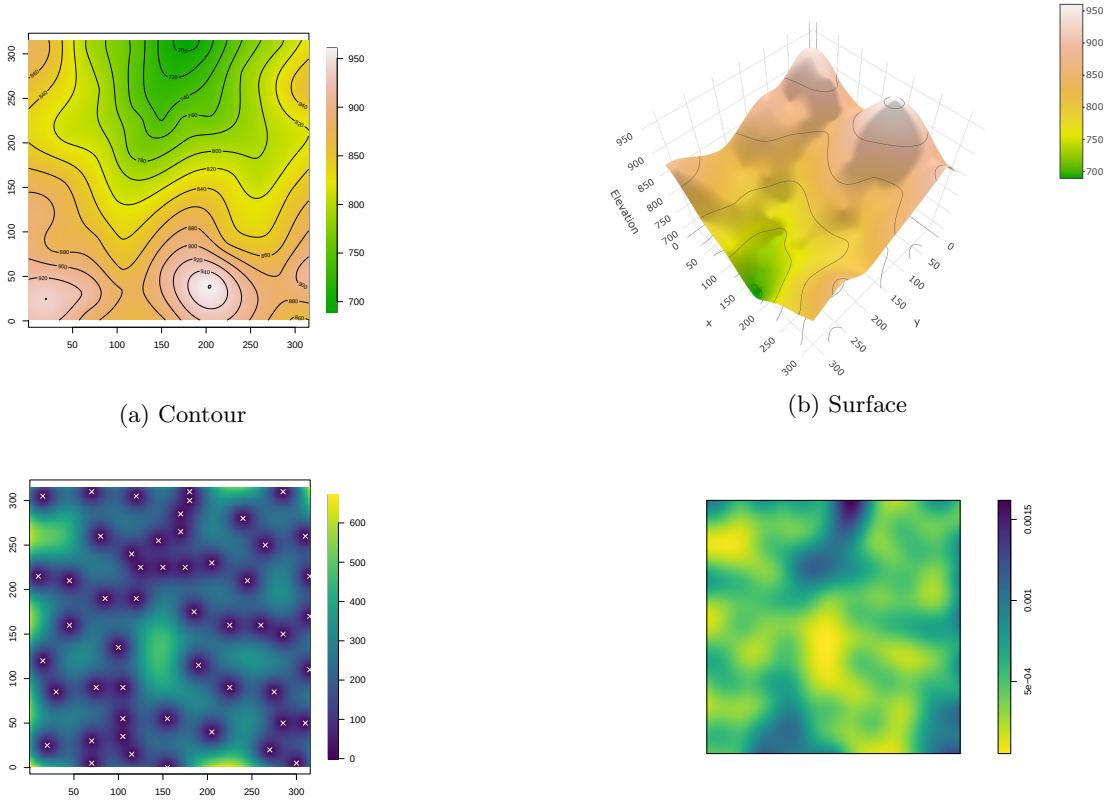
$$\begin{aligned}
\mathrm{Var}[\hat{X}_0 - X_0] &= \mathrm{Var}[\boldsymbol{z}_0^T\hat{\boldsymbol{\beta}} + \boldsymbol{c}\Sigma^{-1}(\boldsymbol{X} - Z\hat{\boldsymbol{\beta}}) - X_0] \\
&= \mathrm{Var}[\underbrace{(\boldsymbol{z}_0^T H + \boldsymbol{c}\Sigma^{-1} - \boldsymbol{c}\Sigma^{-1}ZH)}_{:=B}\boldsymbol{X} - X_0] \\
&= \mathrm{Var}[X_0] + \mathrm{Var}[B\boldsymbol{X}] - 2\mathrm{Cov}(B\boldsymbol{X}, X_0) \\
&= \sigma_0^2 + B\Sigma B^T - 2B\boldsymbol{c} =: \sigma^2_{\hat{X}}(\boldsymbol{s}_0).
\end{aligned}$$

Adding complexity to the parametrization of the expectation function g, e.g. by adding more free parameters, would allow for more flexibility in the fitted model and hence reduce its bias. However, according to the bias-variance trade-off, this may increase the variance of the parameters fitted, $\hat{\boldsymbol{\beta}}$, and consequently we would think that the marginal variance of the exact observation, σ^2, should be higher.

c)

Consider the special case of universal Kriging with $\mathrm{E}[X(\boldsymbol{s})] = \beta_1, \ \boldsymbol{s} \in \mathcal{D}$. We use the regular grid $\tilde{\mathcal{D}} = \{1, 2, \ldots, 315\}^2$ and calculate the Kriging predictor $\hat{X}(\boldsymbol{s})$ and associated prediction variance $\sigma^2_{\hat{X}}(\boldsymbol{s})$, $\boldsymbol{s} \in \tilde{\mathcal{D}}$, by using the function `krige.conv` in R. Inserting this into the model explained in section 2b yielded the results displayed in figure 7.

In Figure 7 both the contour plot and surface plot of the ordinary Kriging predictor show similarities with the interpolated results in figure 6. We notice that both these plots are smoother than their corresponding in figure 6. Plot 7c show low Kriging variance close to the observations and higher where there are less observations. Also we see that the plot bears resemblance to the observation density plot 7d, which makes sense due to the Brownian motion structure of the variance.



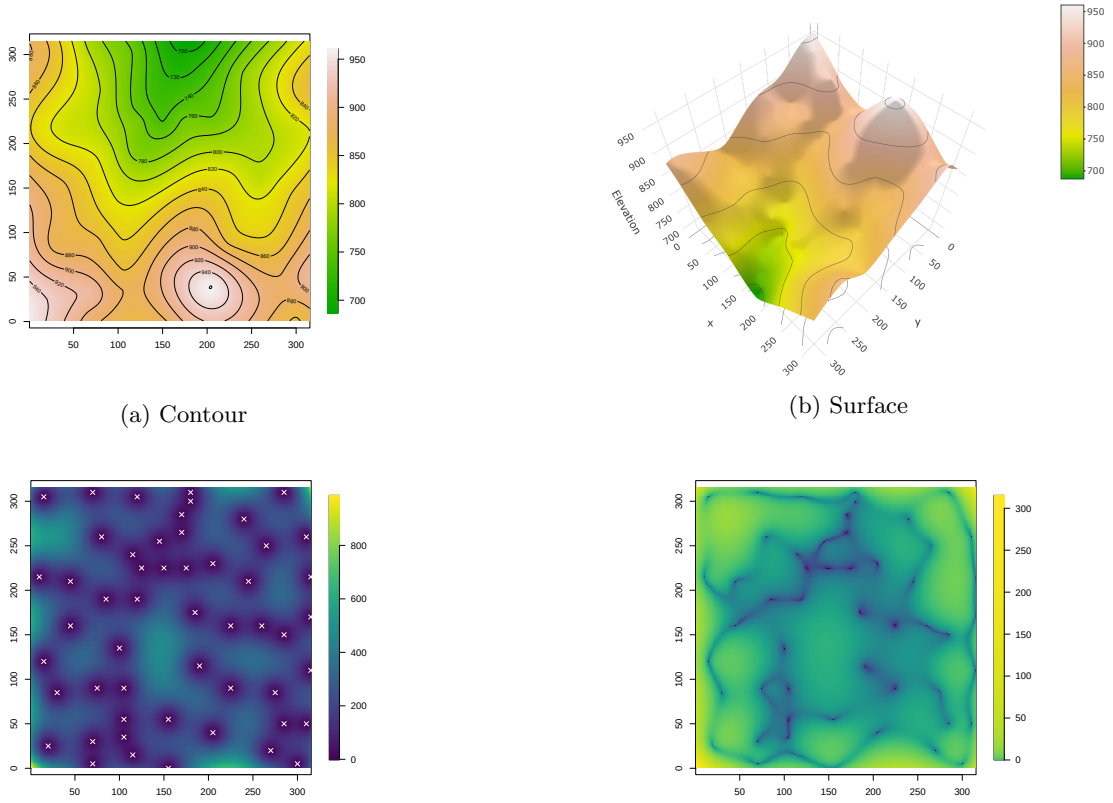
(a) Contour

(b) Surface

(c) Variance

(d) Observation density

Figure 7: The ordinary Kriging predictor displayed by contour plot (a) and surface plot (b), and it's associated prediction variance displayed by a two-dimensional colorgraded plot (c). The white crosses in (c) are the locations of the observations. The plot in the bottom right (d) show the density of our observations.



(a) Contour

(b) Surface

(c) Variance

(d) Variance difference

Figure 8: Same structure as in figure 7, but the plots are produced by a universal Kriging model. The bottom right plot (d) show the universal Kriging predicted variance subtracted the ordinary Kriging predicted variance.

d)

We denote the site $\boldsymbol{s} = (s_1, s_2)$, and set $n_g = 6$ to be the total number of known functions in $\boldsymbol{g}$. Furthermore, we specify each function in $\boldsymbol{g}$ to be the polynomials $s_1^k s_2^l$ for $(k, l) \in \{(0,0), (1,0), (0,1), (1,1), (2,0), (0,2)\}$. Thus, the 6-dimensional vector is

$$\boldsymbol{g}(\boldsymbol{s}) = [1 \quad s_1 \quad s_2 \quad s_1 s_2 \quad s_1^2 \quad s_2^2]$$

The expected value of $X(\boldsymbol{s})$ is then

$$\mathrm{E}[X(\boldsymbol{s})] = \boldsymbol{g}(\boldsymbol{s})^T \boldsymbol{\beta} = \beta_1 + \beta_2 s_1 + \beta_3 s_2 \beta_4 s_1 s_2 + \beta_5 s_1^2 + \beta_6 s_2^2.$$

The results of this model are displayed in figure 8.

The contour and surface plots, 8a and 8b, in figure 8 are smoother than their corresponding plots in figure 6. Also we notice that they are very similar to their corresponding plots in figure 7. We notice some larger differences close to the boundary $y = 0$ in figure 9, where the universal Kriging predictor is higher than the ordinary Kriging predictor. Also, close to the location $\boldsymbol{s} = (315, 250)$ we see that the universal Kriging predictor is lower than the ordinary Kriging predictor. This may be because the universal Kriging predictor follows the trend of neighboring observations more closely, which coincide with our prior beliefs from section 2b regarding variance-bias trade-off. We notice that the scale of the color gradient in plot 8c is larger than that of plot 7c, which is further investigated in the variance difference plot 8d. Here we notice that the universal Kriging variance is higher than the ordinary Kriging variance for all locations s, which also coinciding with our prior beliefs.

e)

Now we investigate the unobserved location $\boldsymbol{s}_0 = (100, 100)^T$ where we get the Kriging predictor $\hat{X}((100, 100)^T) \approx 839$ with associated variance $\sigma^2_{\hat{X}}(\boldsymbol{s}_0) \approx 101$ by the attributes `predict` and `krige.var` of the `krige.conv` object in

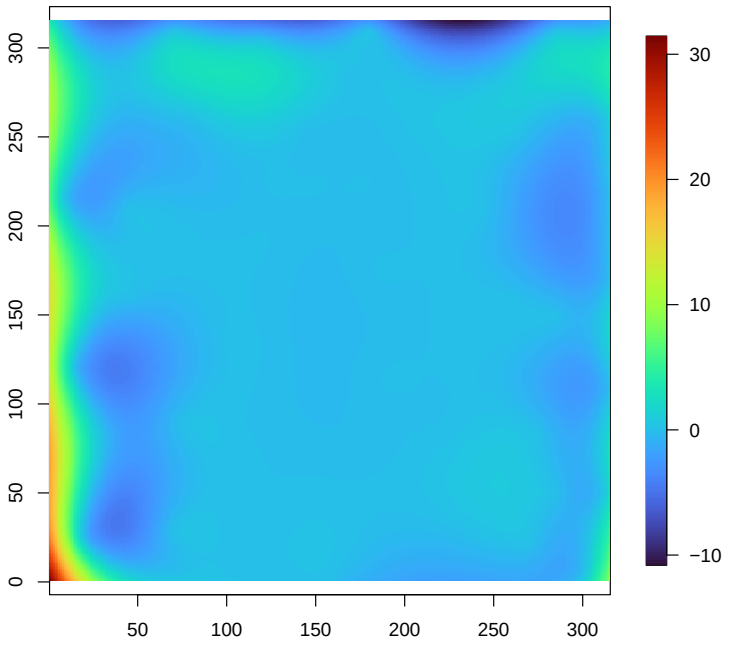


Figure 9: The universal Kriging predictor subtracted the ordinary Kriging predictor.

R, respectively. Since we assumed a stationary GRF, our prediction $\hat{X}_0$ is the conditional expectation $\mathrm{E}(X_0|\boldsymbol{X})$. Furthermore, the distribution of $\hat{X}_0$ is Gaussian and the error $X_0 - \hat{X}_0 \sim \mathcal{N}(0, \sigma^2_{\hat{X}}(\boldsymbol{s}_0))$. Thus, the probability for the elevation to be higher than 850 m at this location is

$$\begin{aligned} P(X_0 > 850|\boldsymbol{X}) &= 1 - \Phi\left(\frac{850 - \mathrm{E}[X_0]}{\sqrt{\mathrm{Var}X_0}}\right) \\ &= 1 - \Phi\left(\frac{850 - \hat{X}_0}{\sqrt{\sigma^2_{\hat{\boldsymbol{X}}}}}\right) \\ &\approx 0.130, \end{aligned}$$

where Φ is the standard normal cumulative distribution. Furthermore we investigate the critical elevation for which there is a 0.90 probability that the true elevation is lower, that is, we want to solve $0.9 = P(X_0 < x|\boldsymbol{X}) = \Phi((x - \hat{X}_0)/\sqrt{\sigma^2_{\hat{\boldsymbol{X}}}})$ for x. We get

$$\begin{aligned} x &= \hat{X}_0 + \Phi^{-1}(0.9)\sqrt{\sigma^2_{\hat{\boldsymbol{X}}}} \\ &\approx 852\ m \end{aligned}$$

f)

We have considered ordinary and universal Kriging models which have been tested on a real data set of $n = 52$ observations of terrain elevation. First we presented the data using interpolation, then we applied ordinary Kriging and discovered a smoother terrain prediction. Lastly we applied universal Kriging with a second order polynomial prior and discovered very similar results as with ordinary Kriging. Some differences were found close to the boundaries of the Kriging predictors, where universal continued the trend of nearby observations captured by its more complex prior expectation, and the ordinary Kriging predictor had a tendency to move back to the constant prior expectation. Also we discovered that the universal Kriging variance was higher than those of the ordinary Kriging for all locations except at the observed locations.

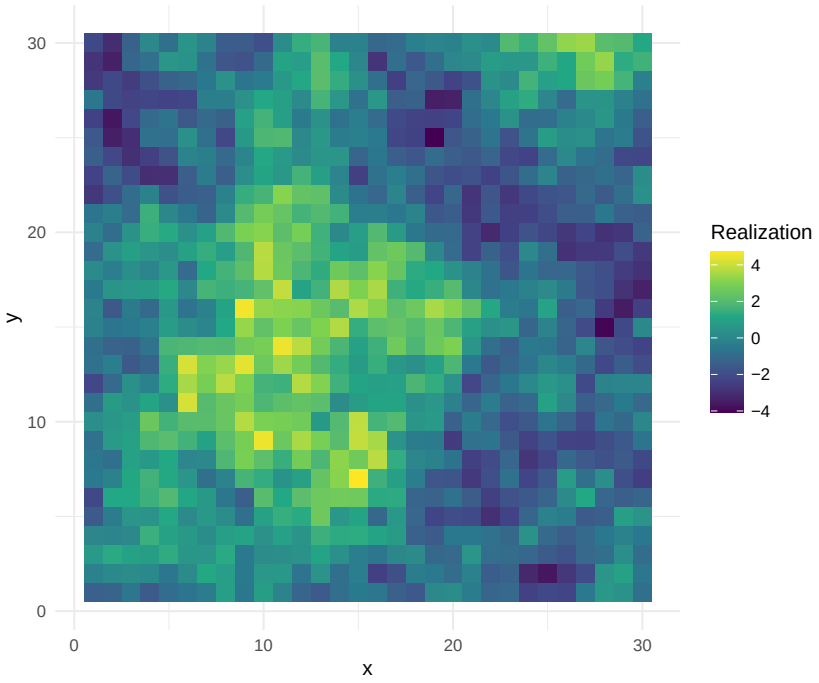


Figure 10: One realization of the GRF described in problem 3 on the discretized grid $\tilde{\mathcal{D}}$.

Problem 3: Parameter estimation

We consider the stationary GRF $\{X(\boldsymbol{s});\ \boldsymbol{s} \in \mathcal{D} = [1, 30]^2 \subset \mathbb{R}\}$ with

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

We let $\tilde{\mathcal{D}} = \{1, 2, \ldots, 30\}^2$ be a regular grid of $\mathcal{D}$, and set the marginal variance $\sigma^2 = 2$ and the spatial scale $a = 3$.

a)

We generate one realization of X on $\tilde{\mathcal{D}}$. The result is displayed in figure 10.

b)

We take the realization from a) to be an exact observation and use this to compute an empirical semi-variogram using the function `variog` from the R-package `GeoR`. The result is plotted jointly with the true semi-variogram in figure 11. We note that the empirical semi-variogram fits the true semi-variogram quite poorly. One could change the number of bins used in the `variog` function, but the approximation still remains a poor fit.

c)

We repeat the procedure in a) and b) three times. The results are plotted in figure 12. It is evident that the empirical variograms are highly variable and depend largely on the realization they are based on.

d)

We consider the realization from a) and sample 36 locations from $\tilde{\mathcal{D}}$ uniformly at random and compute an empirical semi-variogram with the same procedure as in c). The result is displayed in figure 13. We observe that the empirical

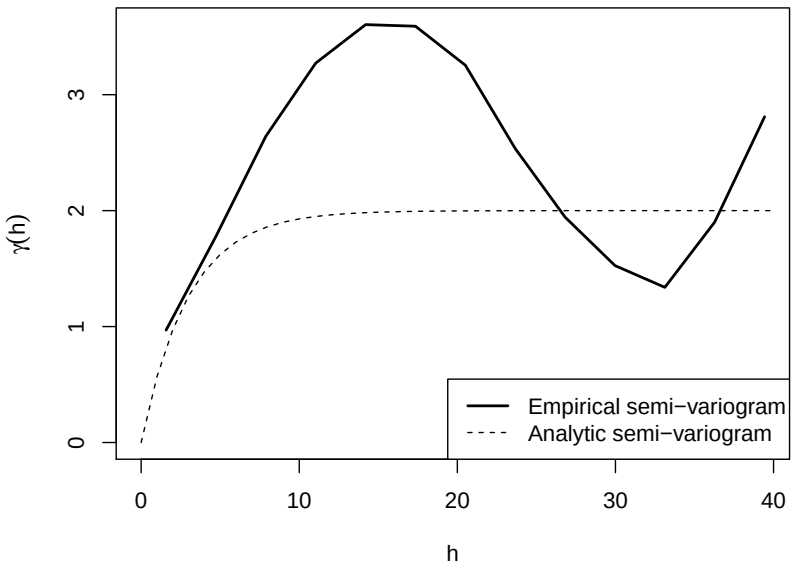


Figure 11: The empirical semi-variogram (solid line) plotted with the true semi-variogram (dashed line) for the GRF presented in problem 3.

semi-variogram is very jagged/noisy, which is not surprising, since we use a mere 36 points to estimate the covariance function for all possible distances between points on the grid.

Next, we consider the parameters σ^2 and a to be unknown, but still assume an exponential covariance model. Then, we use a maximum-likelihood criterion to estimate the parameters with the function `likfit` in the R-package `GeoR`. To estimate the parameters, we first use all the grid points on $\tilde{\mathcal{D}}$, before we only use 36 locations selected uniformly at random. The results are displayed in figure 14. We observe that that the estimated semi-variogram based on only 36 grid points fits the true semi-variogram better than the estimate based on all grid points.

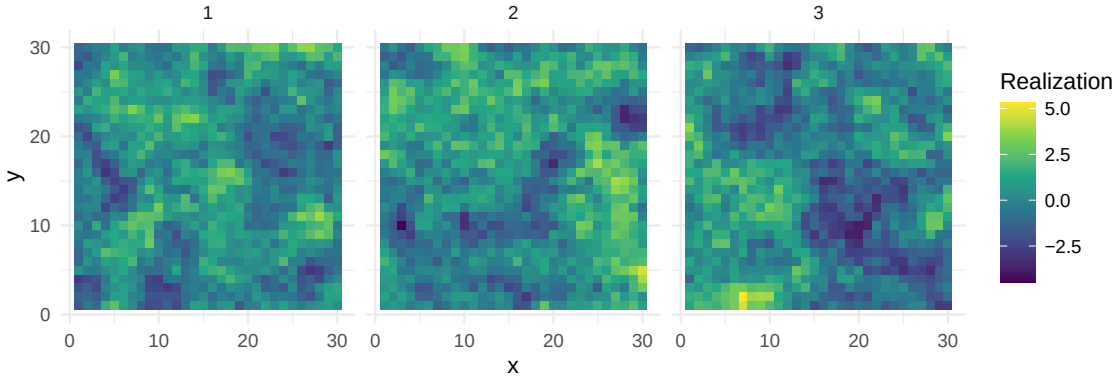
To compare the Empirical approach with the maximum likelihood criterion, we observe that the latter gives a comparatively smoother and more reasonable estimate of the semi-variogram. However, the maximum likelihood approach depends on the assumption of an exponential covariance function, which (in real applications) might be very unreasonable. In that regard, the empirical approach is perhaps more flexible and less biased, though in this particular case, where the covariance function is known, the maximum likelihood approach is clearly superior.

e)

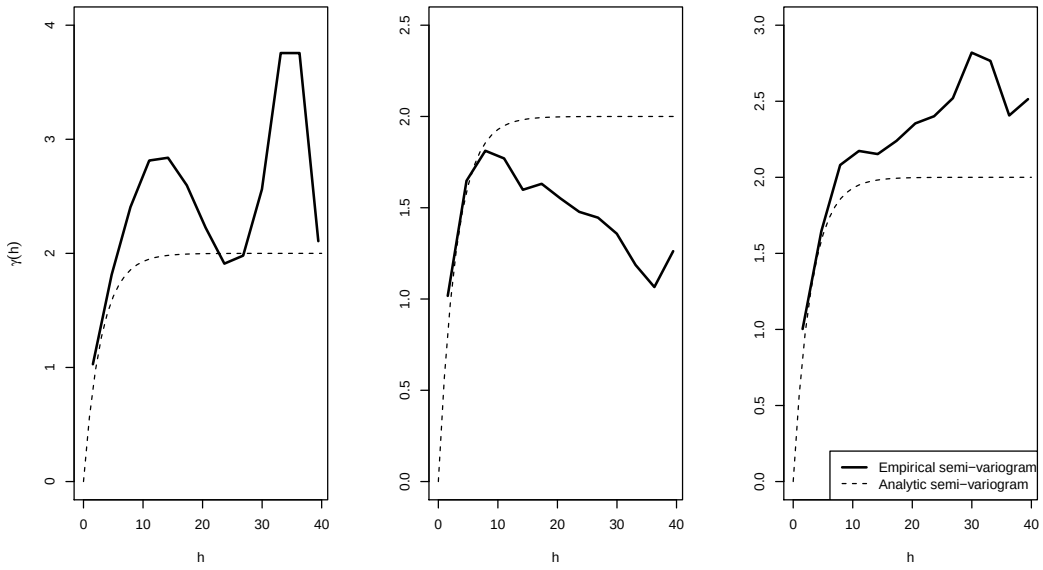
We repeat the procedure in d) for 9, 64 and 100 locations. The resulting estimated semi-variograms are displayed in figure 15. We see that all the empirical semi-variograms are very non-smooth and fit the true semi-variogram poorly. The ML semi-variograms fit well for short distances, $h < 5$, maybe except for the one based on only 9 locations. There seems to be no clear evidence that more grid points leads to better estimates, but this should be assessed through a much larger number of simulations.

f)

In this section we have studied how empirical semi-variograms as well as semi-variograms based on a maximum likelihood estimation fit to the true semi-variogram. It is evident that the empirical semi-variograms are highly variable, and surprisingly the one based on only 36 locations seems to be a better fit than the one based on all the grid points. Estimating the parameters of the covariance function by a maximum likelihood criterion yields more sensible result, but do require that you choose a covariance function. These also seem to be quite variable, and again, surprisingly, the one based on only 36 grid points gives the most accurate parameter estimates. This is probably just a testament to their variability, and more extensive testing would probably show that more grid



(a)



(b)

Figure 12: (a) Three realizations of the GRF in problem 3 on the grid $\tilde{\mathcal{D}}$ and (b) their corresponding empirical semi-variograms plotted with the true semi-variogram.

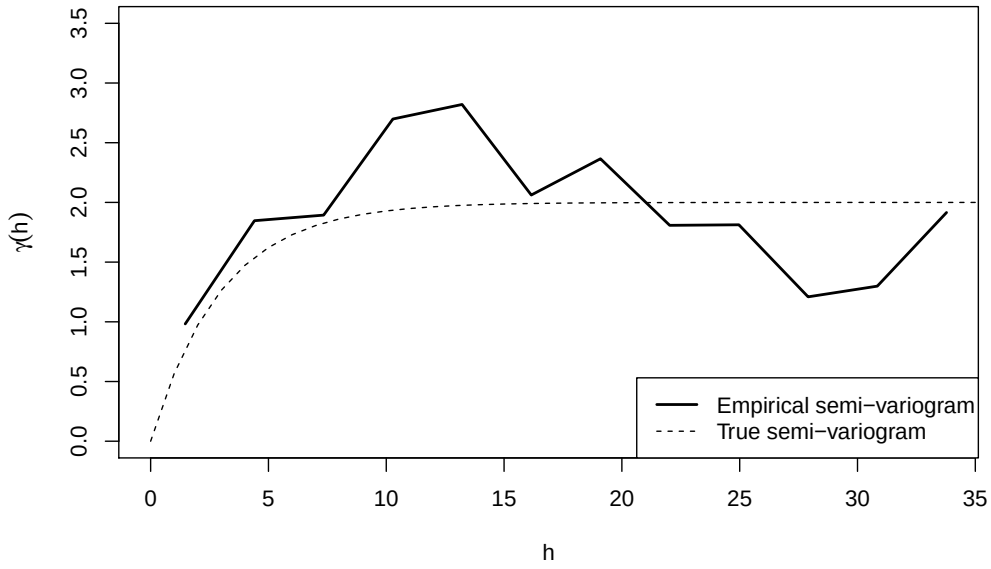


Figure 13: The empirical semi-variogram (solid line) based on 36 exact observations plotted with the true semi-variogram (dashed line) for the GRF presented in problem 3.

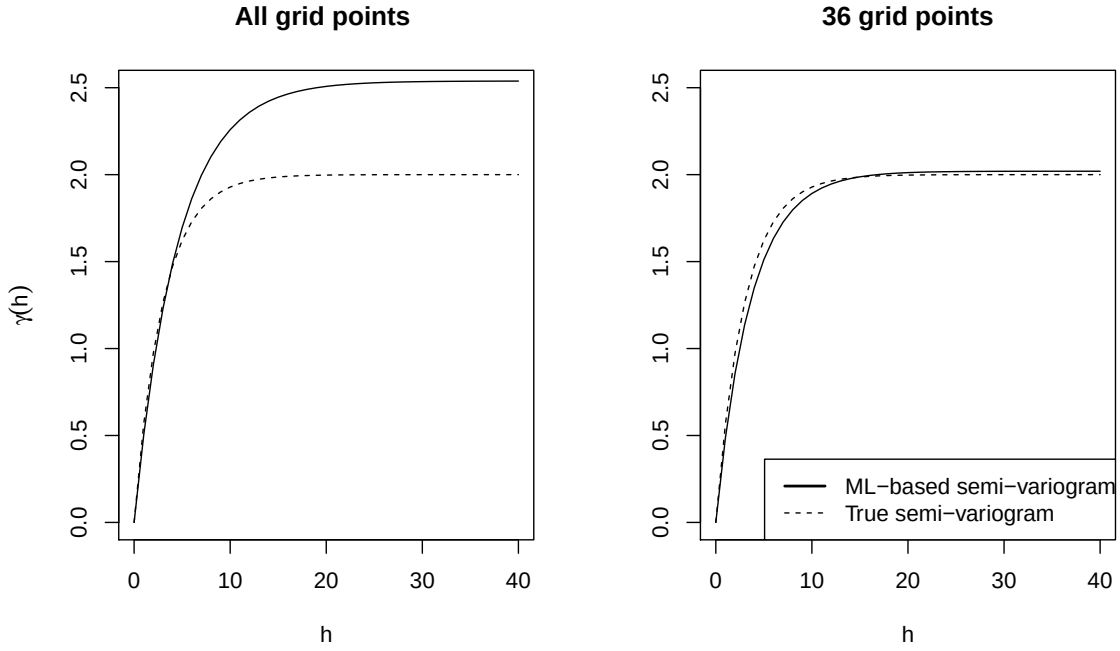
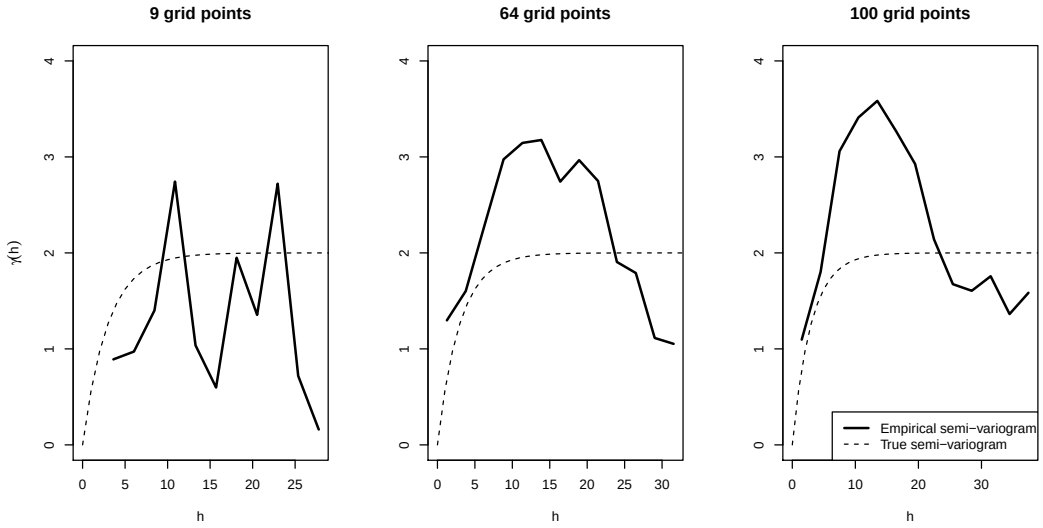
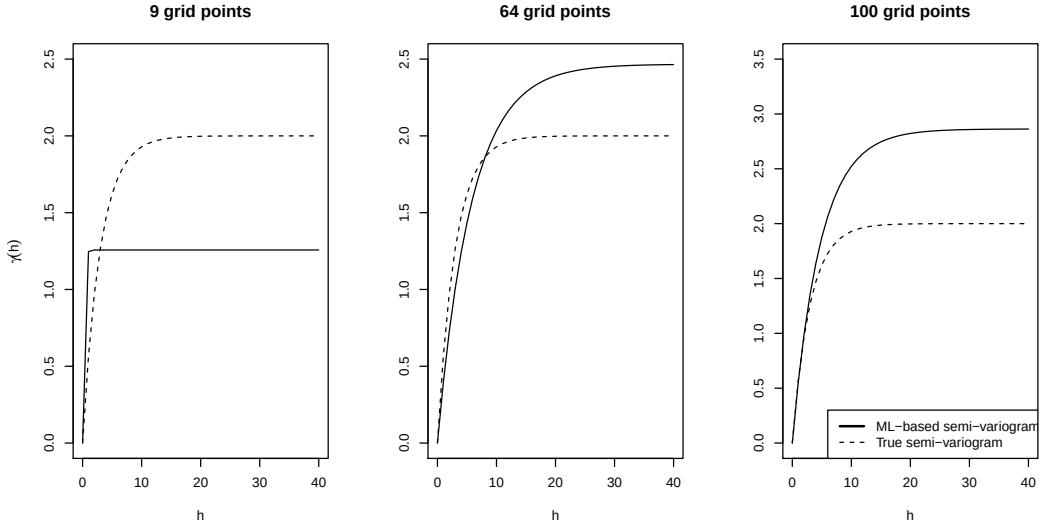


Figure 14: Maximum likelihood semi-variograms (solid lines) based on exact observations on all grid points (left) and on 36 exact observations (right) plotted with the true semi-variogram (dashed line) for the GRF presented in problem 3.



(a)



(b)

Figure 15: (a) Empirical semi-variograms and (b) ML semi-variograms, both computed from 9, 64 and 100 grid points.

points, on average, yields better estimates.