

TMA4250 Project 1

Ola Rasmussen & Maiken Hatten

January 2024

Contents

1 GRFs - model characteristics	2
2 GRF - real data	7
3 Parameter estimation	11

1 GRFs - model characteristics

a)

Let X be a stationary GRF on $\mathcal{D} = [0, 50] \subset \mathbb{R}$, and assume that

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

where ρ is the (isotropic) correlation function of X . Discretize \mathcal{D} by a regular grid $\tilde{\mathcal{D}} = \{1, 2, \dots, 50\}$ and let $\mathbf{X} = (X(1), \dots, X(50))^\top$ be the discretization of X on $\tilde{\mathcal{D}}$.

The correlation function ρ must be a positive semi-definite function, because it requires the ability to produce non-negative variances. A function ρ is positive semi-definite if and only if $\forall m \in \mathbb{N}, \forall a_1, \dots, a_m \in \mathbb{R}$, and $\forall s_1, \dots, s_m \in \mathbb{R}^d$,

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \rho(s_i, s_j) \geq 0.$$

Four examples of correlation functions are plotted in Figures 1 and 2. The two Matérn correlation functions are very similar. They are highly correlated at close distances and decrease slowly. The Powered exponential correlation with power $\alpha = 1.9$ isn't very different, but it reaches zero correlation earlier than the Matérn correlation functions. It is only the Powered exponential correlation function with power $\alpha = 1$ that stands out. Locations close to each other have lower correlation with this function, but the correlation further away is larger than if power $\alpha = 1.9$ was used. One would assume that realizations using this correlation function will explore the codomain faster, because it is less influenced by close observations, while the other correlation function will make their associated GRFs explore at a more leisurely pace.

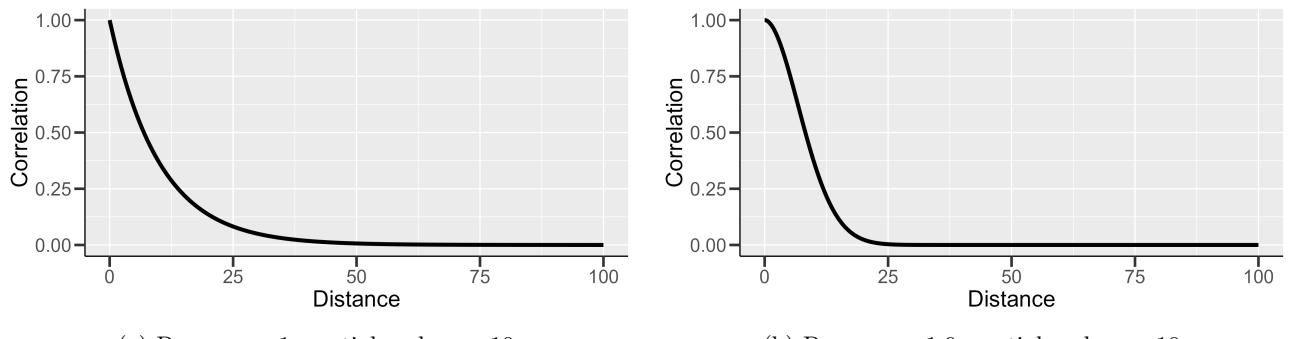


Figure 1: Powered exponential correlation function.

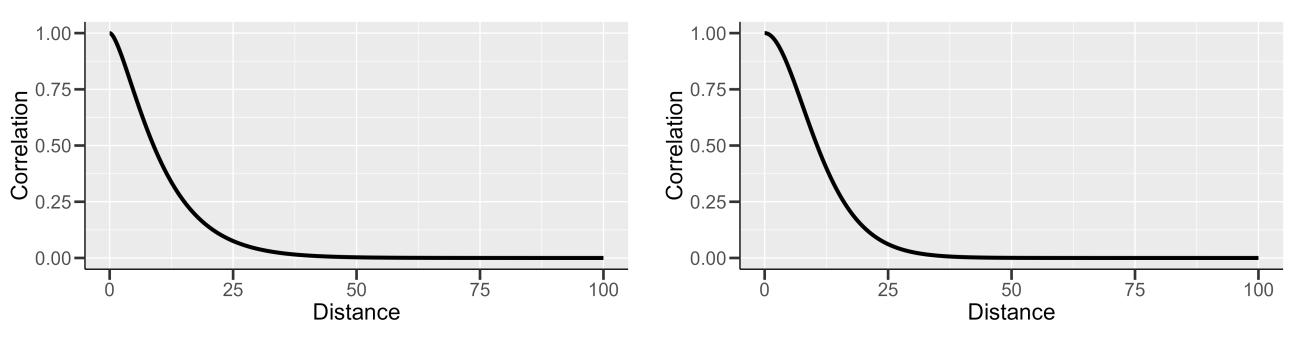


Figure 2: Matérn correlation function.

The relationship between the variogram function $2\gamma(\mathbf{h})$, marginal variance σ^2 , and the correlation function $\rho(\mathbf{h})$ can be seen in Equation (1), where \mathbf{h} is the distance between two locations. The function $\gamma(\mathbf{h})$ is called the semi-variogram function. Eight different semi-variogram functions are plotted in Figures 3 and 4. As expected, the marginal variance only scales the semi-variogram function.

$$2\gamma(\mathbf{h}) = 2\sigma^2(1 - \rho(\mathbf{h})) \quad (1)$$

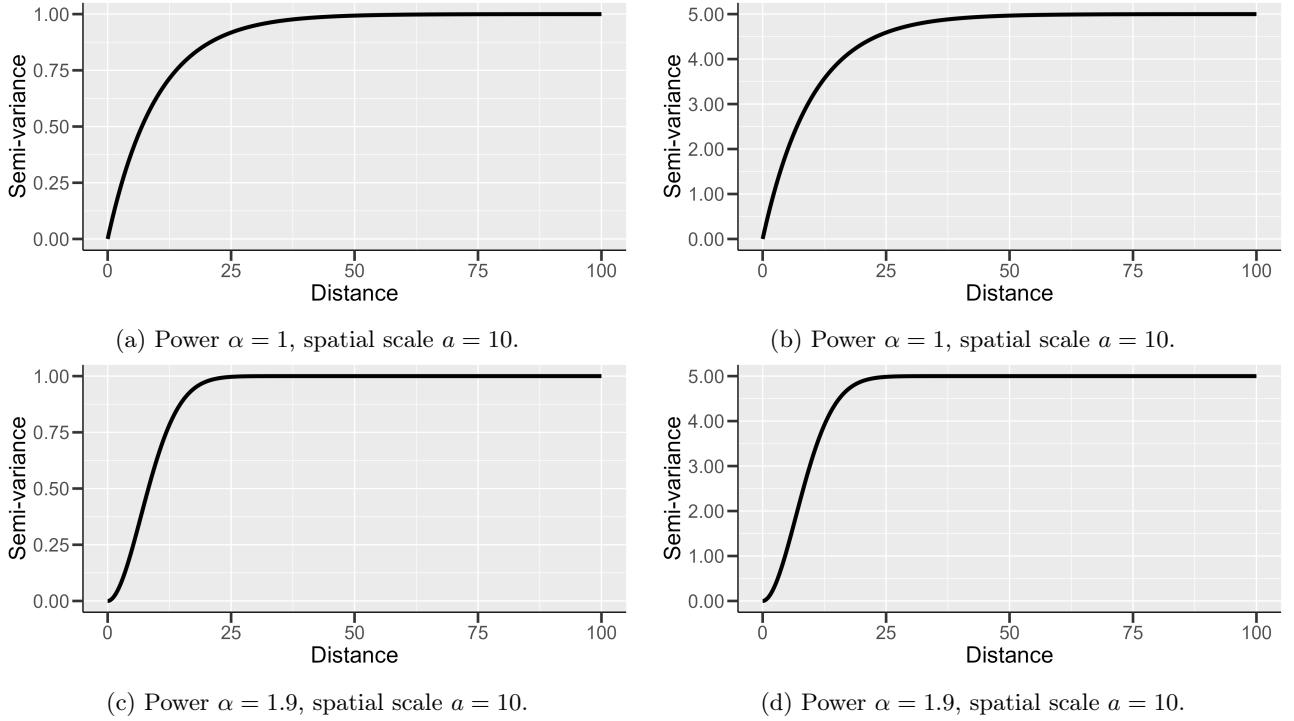


Figure 3: Powered exponential semi-variogram function. $\sigma^2 = 1$ (left) and $\sigma^2 = 5$ (right).

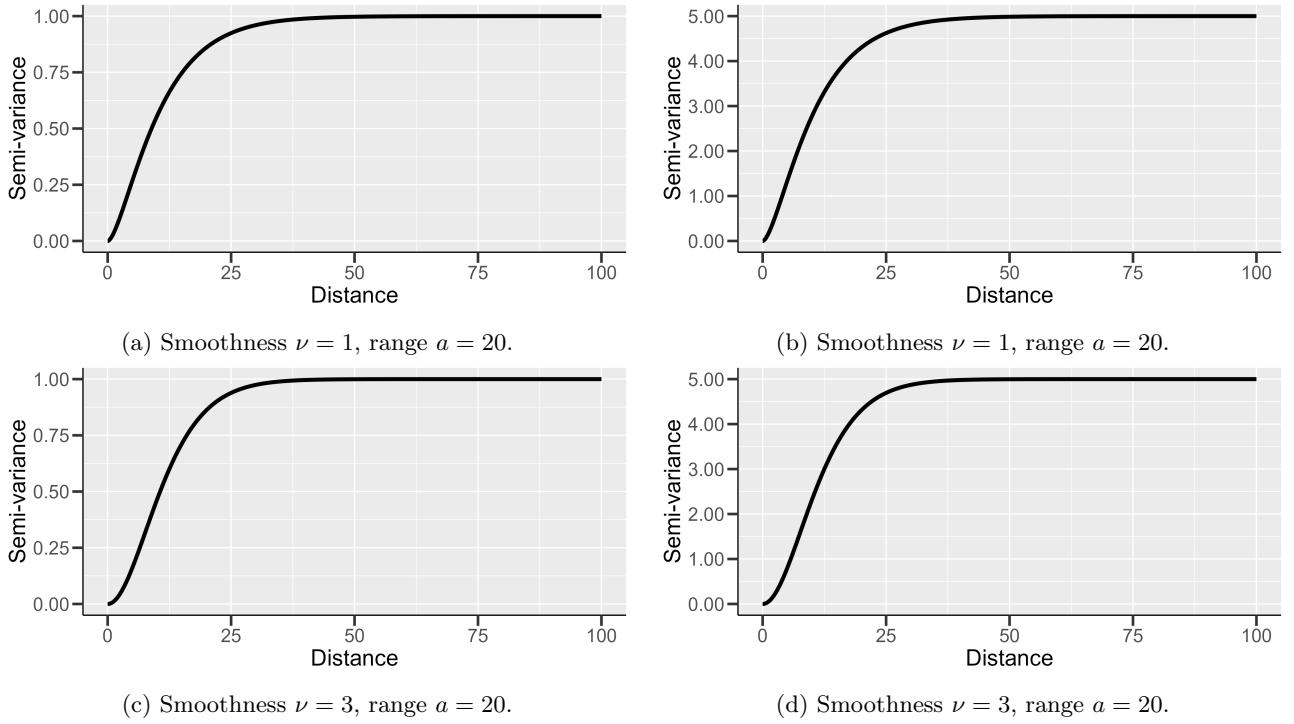


Figure 4: Matérn semi-variogram function. $\sigma^2 = 1$ (left) and $\sigma^2 = 5$ (right).

b)

Since X is a centered stationary GRF, \mathbf{X} is also a centered stationary GRF, and the components of the covariance matrix $\Sigma_{\mathbf{X}}$ are $\Sigma_{\mathbf{X},ij} = \sigma^2 \rho(\|i - j\|)$, $i, j = 1, \dots, 50$.

Figures 5 and 6 shows four realizations using the eight model combinations in Figures 3 and 4. As we assumed in Section 1a), the realizations using the Powered exponential correlation function with power $\alpha = 1$ move more rapidly. The marginal variance only affects the amplitude of the realizations, and not the general behavior.

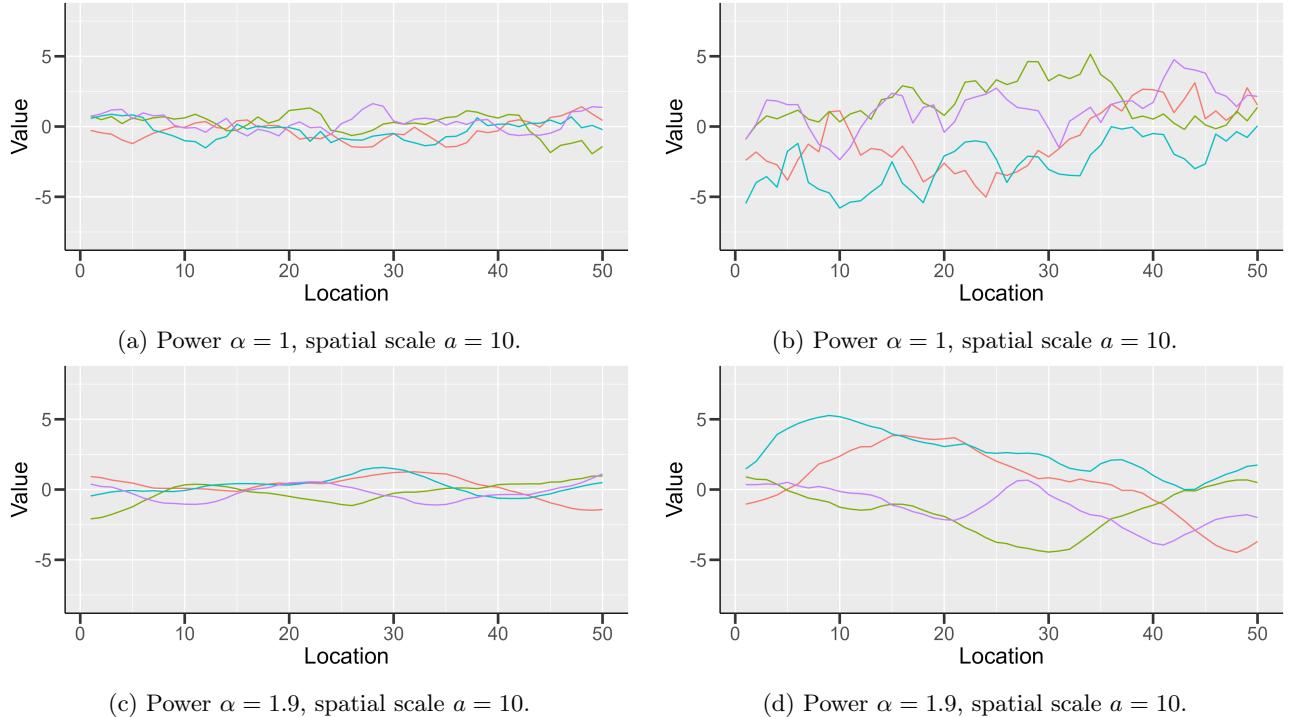


Figure 5: Powered exponential realizations. $\sigma^2 = 1$ (left) and $\sigma^2 = 5$ (right).

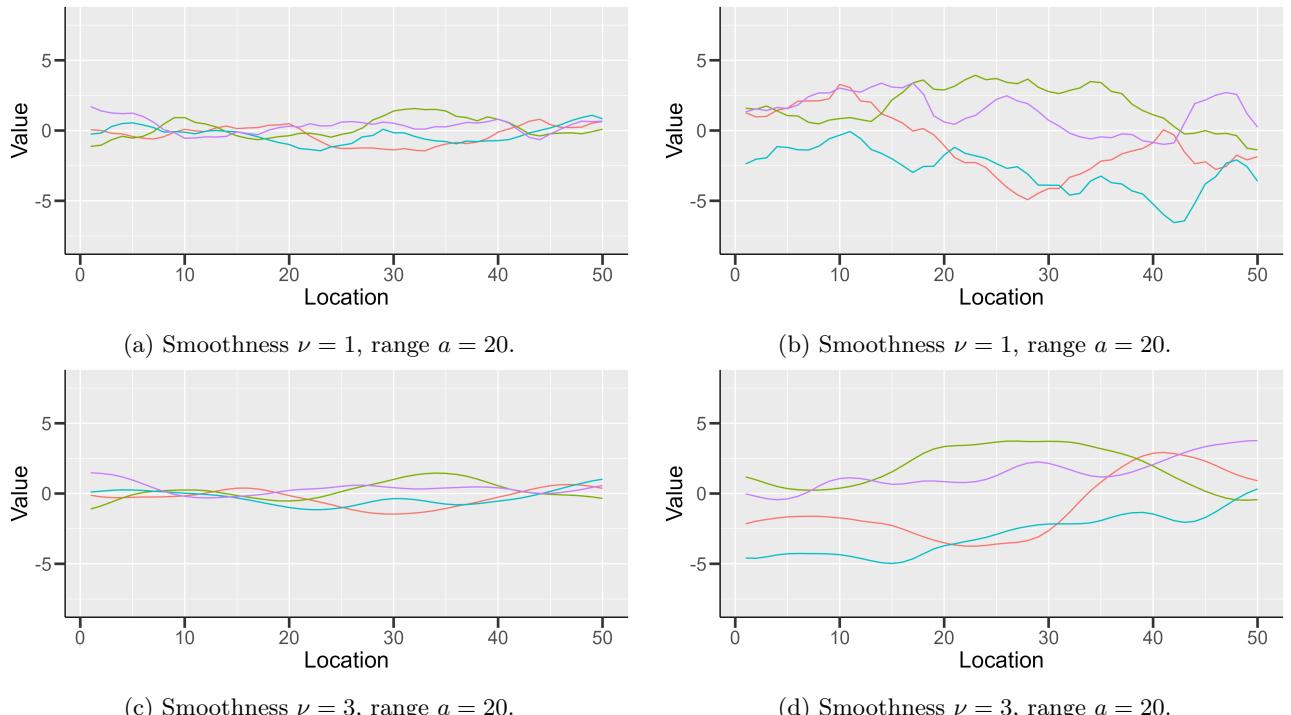


Figure 6: Matérn realizations. $\sigma^2 = 1$ (left) and $\sigma^2 = 5$ (right).

c)

Let Y_1 , Y_2 , and Y_3 be values at locations $s_1 = 10$, $s_2 = 25$, and $s_3 = 30$, respectively, that we plan to observe. The observations model is given by

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

where $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_N^2)$ and independent of X . The parameter σ_N^2 is called the nugget variance.

Since $\mathbf{Y} = (Y_1, Y_2, Y_3)^\top$ is a sum of centered trivariate GRFs, it is also a centered trivariate GRF, where the elements of the covariance matrix Σ_Y are $\Sigma_{Y,ij} = \sigma^2 \rho(\|s_i - s_j\|) + \sigma_N^2 \mathbb{1}(i = j)$, $i, j = 1, 2, 3$. $\mathbb{1}$ is the indicator function.

d)

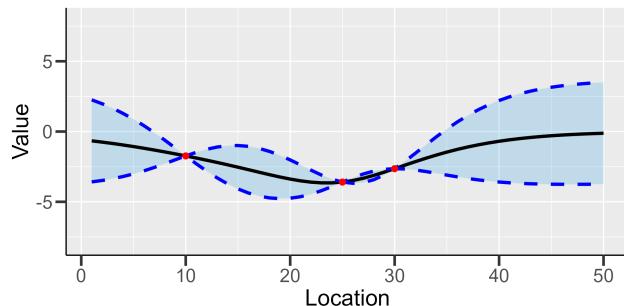
The conditional mean vector $\mu_{X|Y=y}$ and covariance matrix $\Sigma_{X|Y=y}$ are found using the formulas in Equations (2) and (3).

$$\mu_{X|Y=y} = \mu_X + \Sigma_{XY} \Sigma_Y^{-1} (y - \mu_Y), \quad (2)$$

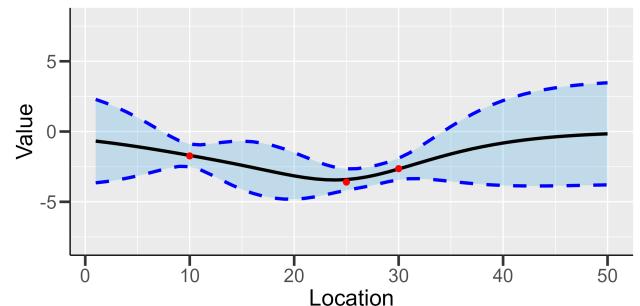
$$\Sigma_{X|Y=y} = \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX}. \quad (3)$$

The marginal means are zero, Σ_X is the covariance matrix specified in Section 1b), Σ_Y is the covariance matrix specified in Section 1c), and $\Sigma_{XY} = \Sigma_{YX}^\top$ are found by finding the covariance between the corresponding X 's and Y 's at each location. In our case, the covariance between the Y 's and X 's are the same as the covariance between the X 's at locations s_1, s_2 , and s_3 and all the other locations. I.e., $\text{Cov}(Y_i, X(s_j)) = \text{Cov}(X(s_i), X(s_j))$, $i = 1, 2, 3$, $j = 1, 2, \dots, 50$.

We will choose the red line in Figure 6(d) to condition the rest of our simulations on. The predicted mean and associated 90% prediction interval can be seen plotted in Figure 7. We see that with a nugget variance, we add uncertainty in our observations, hence the prediction interval doesn't go into our observed points. The prediction interval is largest when we are far away from any observed points, and is almost the same as the mean between point s_2 and s_3 .



(a) Nugget $\sigma_N^2 = 0$.

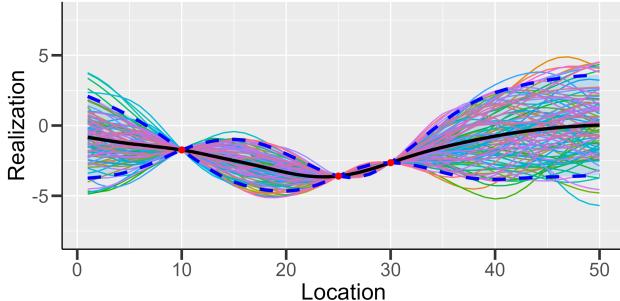


(b) Nugget $\sigma_N^2 = 0.25$.

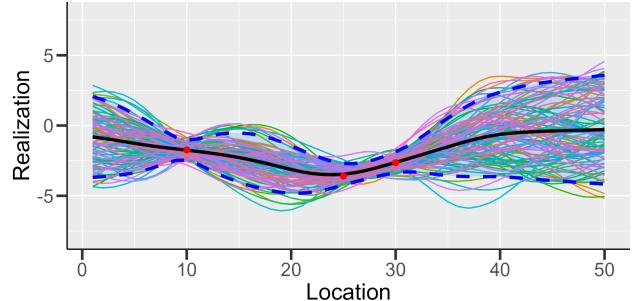
Figure 7: Our observed points (red points), mean (black line), and 90% prediction interval (shaded area).

e)

Figure 8 show 100 realizations of $\mathbf{X} \mid \mathbf{Y} = \mathbf{y}$. About 10% of our observations are outside the empirical prediction interval, which is what we want. In Figure 8(b), the realizations doesn't travel through our observed points because of the added uncertainty. The empirical prediction intervals are very similar to their analytical counterparts in Figure 7.



(a) $\sigma_N^2 = 0$.



(b) $\sigma_N^2 = 0.25$.

Figure 8: 100 realizations of $\mathbf{X} \mid \mathbf{Y} = \mathbf{y}$. Our observed points (red points), empiric mean (black line), and empiric 90% prediction interval (shaded area).

f)

Define

$$A = \sum_{s \in \mathcal{D}} \mathbb{1}(X(s) > 2)(X(s) - 2),$$

where $\mathbb{1}$ is the indicator function. This is an approxiamtion of the area under X and above the level 2. We will use the predictor \hat{A} given by

$$\hat{A} = \frac{1}{100} \sum_{i=1}^{100} \sum_{s \in \mathcal{D}} \mathbb{1}(X_i(s) > 2)(X_i(s) - 2),$$

where X_i is our i 'th realization in Figure 8(a). Our prediction \hat{A} is equal to 1.85 and the prediction variance τ^2 is equal to 18.7. We can see using Figure 8(a) that these values are reasonable because not many realizations are above a value of 2.

Another predictor \tilde{A} is given by

$$\tilde{A} = \sum_{s \in \mathcal{D}} \mathbb{1}(\hat{X}_i(s) > 2)(\hat{X}_i(s) - 2).$$

The Simple Kriging predictor $\hat{X}(s)$ is the conditional mean we calculated in Section 1d). We then get that \tilde{A} is equal to 0, which we already could have seen in Figure 7(a) because the mean (black line) is always below a value of 2.

Jensen's inequality states that for any convex function f , the function evaluated at the expectation is less than or equal to the expectation of the function. I.e., $E[f(X)] \geq f(E[X])$. Therefore since \tilde{A} is evaluated at the expectation and \hat{A} is the expectation, $\hat{A} \geq \tilde{A}$.

g)

This problem is very similar to other problems we have had in other courses, so there wasn't anything that was particularly difficult. Is was a bit hard to keep track of the different parameterizations used in the R libraries and the lectures, but that was about it. The only new stuff was the added uncertainty of our observations. We had never done that before, so it was fun figuring that out.

2 GRF - real data

a)

The data extracted from `topo.dat` can be seen in Figure 9. It contains 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 $\mathbf{X} = (X(s_1), \dots, X(s_{52}))^\top$. The different elevations are grouped together, where the lower elevations are at the top of the plot and the higher elevations are at the bottom. So a stationary GRF can be a suitable model for this data since the correlation function is isotropic (a function of distance), and hence stationary.

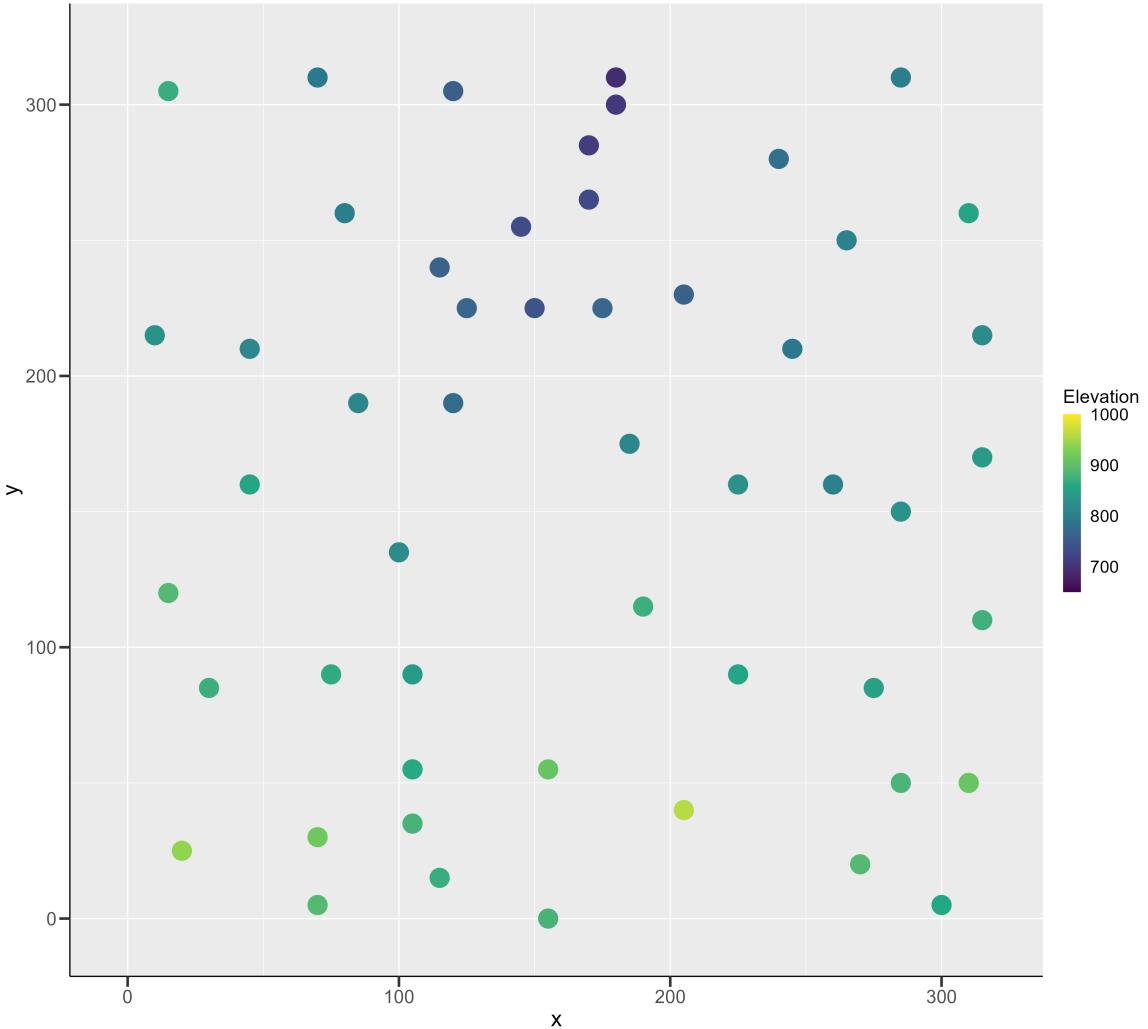


Figure 9: Scatterplot of `topo.dat`.

b)

Assume the GRF X is modelled by

$$\begin{aligned} E[X(\mathbf{s})] &= \mathbf{g}(\mathbf{s})^\top \boldsymbol{\beta}, \quad \mathbf{s} \in \mathcal{D}, \\ \text{Var}[X(\mathbf{s})] &= \sigma^2, \quad \mathbf{s} \in \mathcal{D}, \\ \text{Corr}[X(\mathbf{s}), X(\mathbf{s}')] &= \rho(\|\mathbf{s} - \mathbf{s}'\|), \quad \mathbf{s}, \mathbf{s}' \in \mathcal{D}, \end{aligned}$$

where $\mathbf{g}(\mathbf{s}) = (1, g_2(\mathbf{s}), \dots, g_{n_g}(\mathbf{s}))$ is a n_g -vector of known explanatory spatial variable for $\mathbf{s} \in \mathcal{D}$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{n_g})^\top$ is a n_g -vector of unknown parameters. Moreover, let the marginal variance be $\sigma^2 = 2500$ and the correlation function be $\rho(h) = \exp\left\{- (0.01h)^{1.5}\right\}$, $h \in [0, \infty)$.

We want to find the best linear unbiased predictor (BLUP) \hat{X}_0 for $X_0 = X(\mathbf{s}_0)$ at an unobserved location $\mathbf{s}_0 \in \mathcal{D}$. The predictor must be based on \mathbf{X} and satisfy the following:

1. $\hat{X}_0 = \mathbf{a}^\top \mathbf{X}$ for $\mathbf{a} \in \mathbb{R}^{52}$.
2. $E[\hat{X}_0] = E[X_0]$.
3. $\text{MSE}(\mathbf{s}_0) = E\left[\left(X_0 - \hat{X}_0\right)^2\right]$ is minimized.

We denote $\Sigma = \text{Cov}[\mathbf{X}]$, $\sigma_0^2 = \text{Var}[X_0]$, and $\mathbf{c} = \text{Cov}[\mathbf{X}, X_0]$, $\mathbf{c} \in \mathbb{R}^{52}$. The minimization problem to be solved is therefore

$$\text{MSE}(\mathbf{s}_0) = \sigma_0^2 - 2\mathbf{a}^\top \mathbf{c} + \mathbf{a}^\top \Sigma \mathbf{a}.$$

We consider the quantity

$$\phi(\mathbf{a}, \nu) = \sigma_0^2 - 2\mathbf{a}^\top \mathbf{c} + \mathbf{a}^\top \Sigma \mathbf{a} - 2\nu(\mathbf{a}^\top \mathbf{Z} - z_0^\top),$$

where ν is the Lagrange multiplier, \mathbf{Z} is the design matrix, and \mathbf{z}_0 is given. ϕ is minimized where the partial derivatives w.r.t. \mathbf{a} and ν are equal to zero. We then obtain

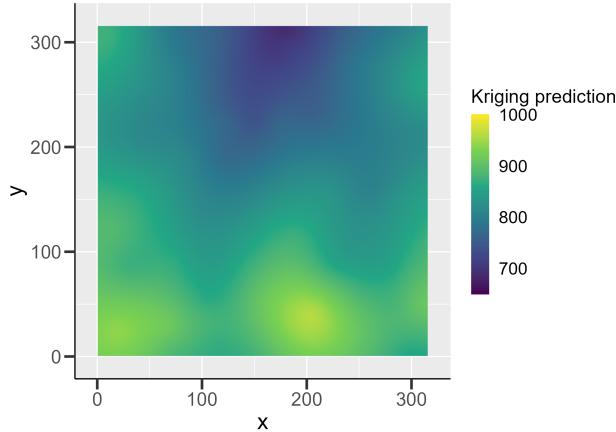
$$\begin{aligned} \nu &= (\mathbf{Z}^\top \Sigma^{-1} \mathbf{Z})^{-1} (\mathbf{z}_0 - \mathbf{Z}^\top \Sigma^{-1} \mathbf{c}) \\ \mathbf{a} &= \Sigma^{-1} \mathbf{c} + \Sigma^{-1} \mathbf{Z} (\mathbf{Z}^\top \Sigma^{-1} \mathbf{Z})^{-1} (\mathbf{z}_0 - \mathbf{Z}^\top \Sigma^{-1} \mathbf{c}) \end{aligned}$$

By substituting these into $\text{MSE}(\mathbf{s}_0)$ we obtain the BLUP \hat{X}_0 of X_0 and the prediction variance $\tau^2(\mathbf{s}_0)$

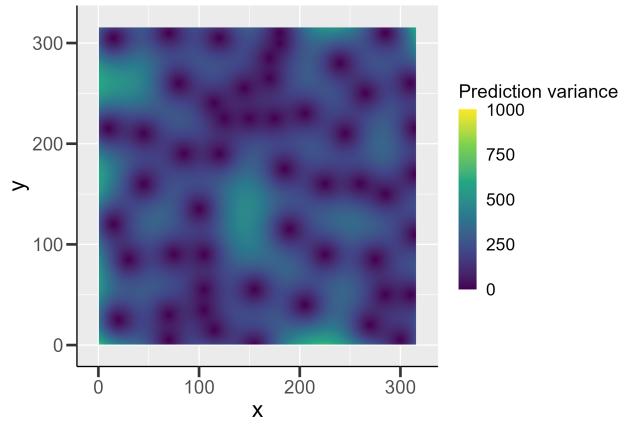
$$\begin{aligned} \hat{X}_0 &= \left(\mathbf{c}^\top \Sigma^{-1} + (\mathbf{z}_0 - \mathbf{Z}^\top \Sigma^{-1} \mathbf{c})^\top (\mathbf{Z}^\top \Sigma^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \Sigma^{-1} \right) \mathbf{X}, \\ \tau^2(\mathbf{s}_0) &= \sigma_0^2 - \mathbf{c}^\top \mathbf{Z}^{-1} \mathbf{c} + (\mathbf{z}_0 - \mathbf{Z}^\top \Sigma^{-1} \mathbf{c})^\top (\mathbf{Z}^\top \Sigma^{-1} \mathbf{Z})^{-1} (\mathbf{z}_0 - \mathbf{Z}^\top \Sigma^{-1} \mathbf{c}). \end{aligned}$$

c)

Using the regular grid $\tilde{\mathcal{D}} = \{1, 2, \dots, 315\}^2$ of \mathcal{D} , we have plotted in Figure 10 the Kriging predictions $\hat{X}(s)$, $s \in \tilde{\mathcal{D}}$, and prediction variances $\sigma_{\hat{X}}^2(s)$, $s \in \tilde{\mathcal{D}}$, of an ordinary Kriging model. The predictions are reasonable when compared to our observed elevations in Figure 9. The variances are all zero at our observed points, but very large the further we move from them.

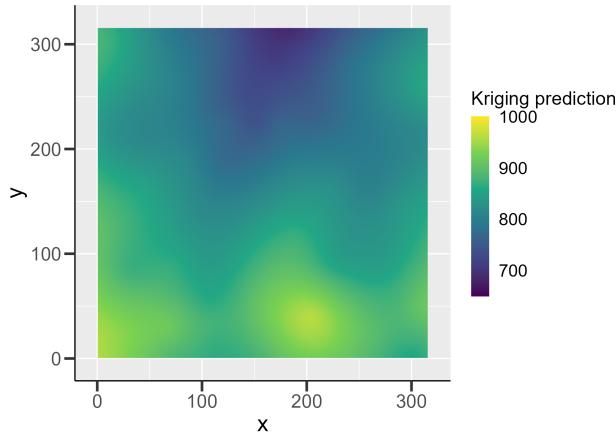


(a) Kriging predictions.

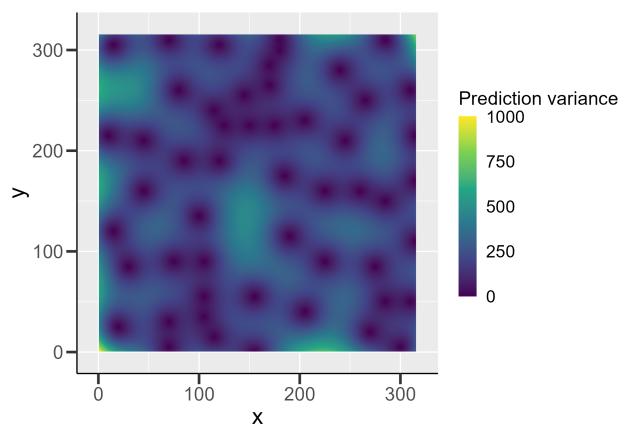


(b) Kriging prediction variances.

Figure 10: Ordinary Kriging model.



(a) Kriging predictions.



(b) Kriging prediction variances.

Figure 11: Universal Kriging model with second order mean function.

d)

We let the reference variable $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2$ be denoted $\mathbf{s} = (s_1, s_2)$, set $n_g = 6$, and define \mathbf{g} to be all polynomials $s_1^k s_2^l$ for $(k, l) \in \{(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (0, 2)\}$. The resulting 6-dimensional vector $\mathbf{g}(\mathbf{s})$ is $\mathbf{g}(\mathbf{s}) = (1, s_1, s_2, s_1 s_2, s_1^2, s_2^2)^\top$, and the expected value of $X(\mathbf{s})$ is

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

The universal Kriging predictions and variances using this mean function is plotted in Figure 11. There is no discernible difference between the ordinary and universal Kriging predictions.

e)

Using the ordinary Kriging predictor and associated prediction variance from Section 2c), we can calculate the probability that the elevation at $\mathbf{s}_0 = (100, 100)^\top$ is larger than 850. This is done using the prediction $\hat{X}_0 = E[X(\mathbf{s}_0)] = 838.6781$, and the prediction variance $\sigma_{\hat{X}}^2(\mathbf{s}_0) = 100.8866$ as seen below:

$$P(\hat{X}_0 > 850) = P\left(\frac{\hat{X}_0 - 850}{\sigma_{\hat{X}}(\mathbf{s}_0)} > z\right) = P\left(\frac{838.6781 - 850}{10.04423} > z\right) \approx 87\%$$

The elevation h needed for a probability of 90% that X_0 is below it is given by

$$P\left(\frac{\hat{X}_0 - h}{\sigma_{\hat{X}}(\mathbf{s}_0)} < z\right) = 0.9 \quad \Rightarrow \quad \frac{\hat{X}_0 - h}{\sigma_{\hat{X}}(\mathbf{s}_0)} = 1.281552 \quad \Rightarrow \quad h = 825.8059.$$

f)

In this problem, we spent a good chunk of time designing the plots. We wanted to make them nice and contain only the necessary information, and we think we succeeded on that part. It was fun working with location data, but it was a bit difficult to make the algorithms fast with how many distances we had to compute.

3 Parameter estimation

a)

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

$$\mathbb{E}[X(\mathbf{s})] = \mu = 0, \quad \mathbf{s} \in \mathcal{D},$$

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

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

We let $\tilde{\mathcal{D}} = \{1, 2, \dots, 30\}^2$ be a regular grid of \mathcal{D} , and set the marginal variance to $\sigma^2 = 2$ and the spatial scale to $a = 3$. One realization of this model is plotted in Figure 12.

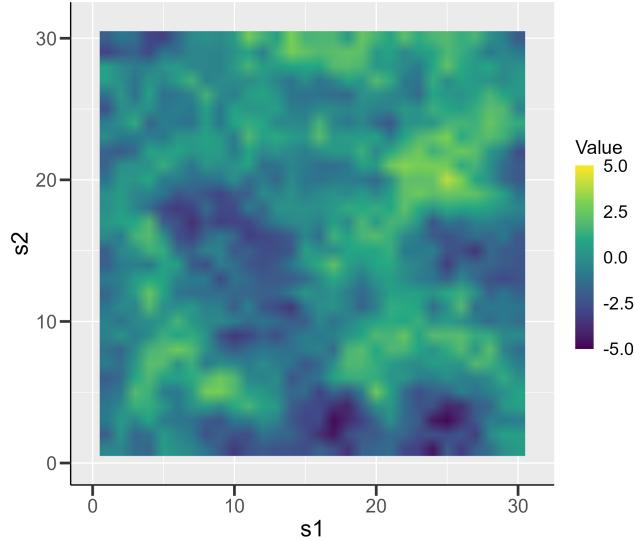


Figure 12: One realization of X .

b)

The empirical semi-variogram based on the realization in Figure 12 is, together with the true semi-variogram function, plotted in Figure 13. We observe from Figure 13 that the empirical semi-variogram aligns more closely with the theoretical semi-variogram at shorter distances, yet they begin to diverge more as the distance increases. This divergence can be due to how there are fewer number of observations for greater distances, especially given that the empirical semi-variogram is derived from just one realization.

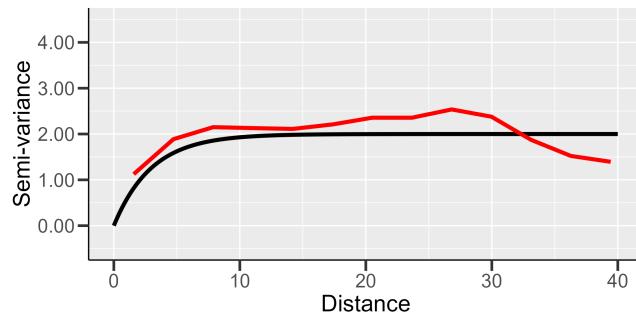


Figure 13: Empirical semi-variogram function (red line), true semi-variogram function (black line).

c)

2 new realizations of this model are plotted in Figure 14. Their respective empirical semi-variance functions are plotted in Figure 15. Figure 15 further show that at small distances, the empirical semi-variogram follows the theoretical semi-variogram closely but starts to deviate as the distance increases. These observations are also the result of having less data at large distances, as the number of data pairs used for estimating decreases.

Moreover, Figure 12 and Figure 14 show that each realization of the GRF will be different due to the random nature of the process, and we also observe that the empirical semi-variograms in Figure 15 contain some degree of randomness. This reflects how the empirical semi-variogram found from one single realization will reflect the sampling variability of that particular realization.

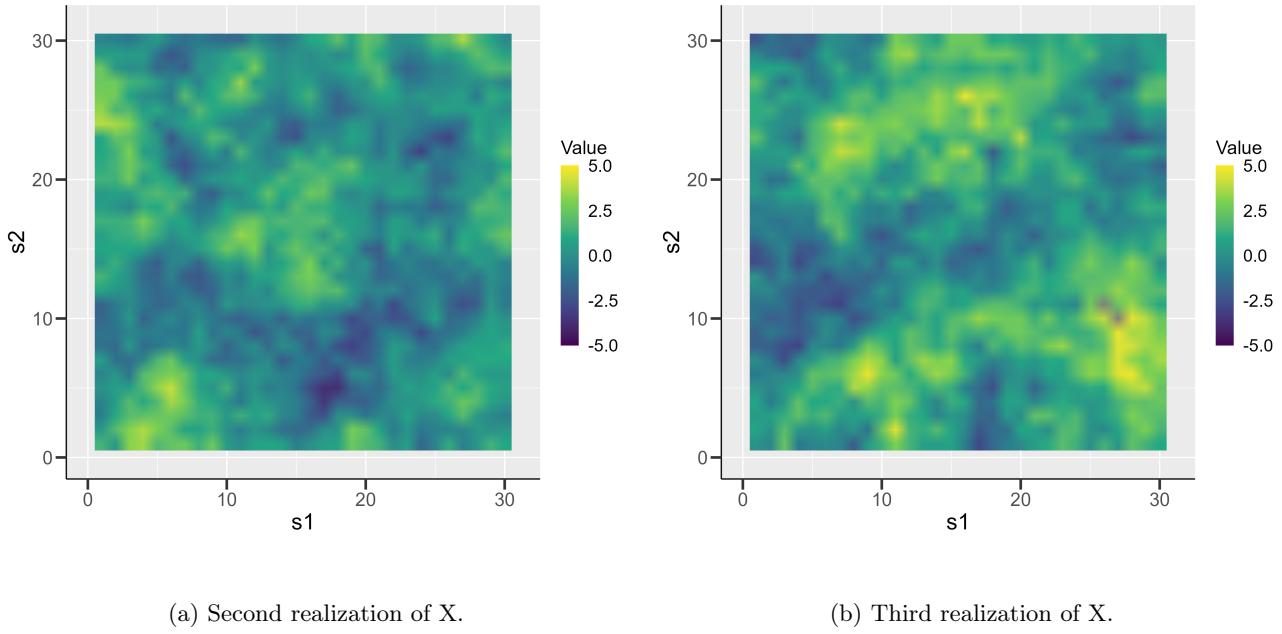


Figure 14: Two new realizations of X.

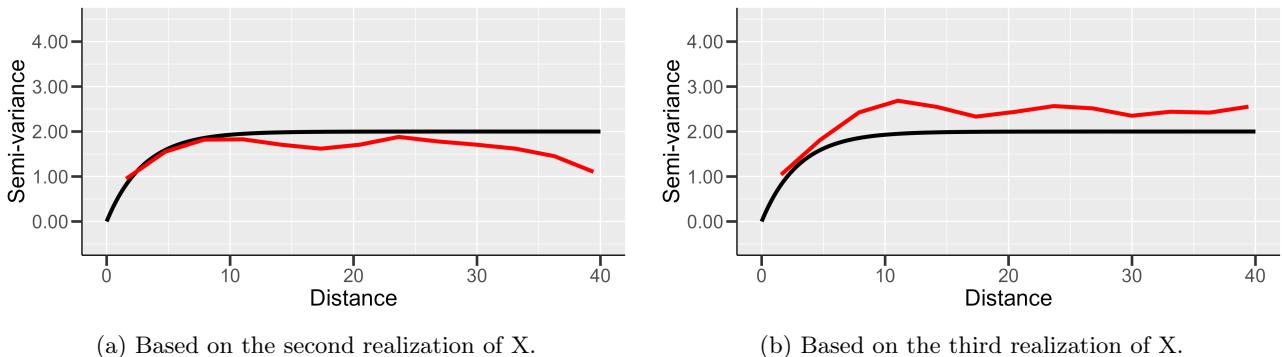


Figure 15: Empirical semi-variogram function (red line), true semi-variogram function (black line).

d)

Based on 36 random locations of the realization in Figure 12, we have computed the empirical semi-variogram function and plotted it in Figure 16. The empirical semi-variogram in Figure 16 fluctuates considerably, and this is due to the fewer data points leading to greater sampling variability.

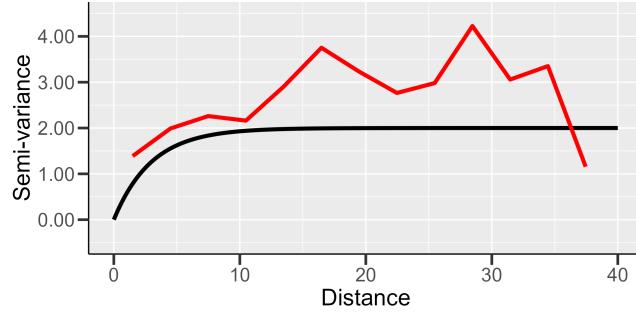


Figure 16: Empirical semi-variogram function (red line) based on 36 random locations, true semi-variogram function (black line).

We now consider the model parameters, σ^2 and a , to be unknown. To estimate them, we use a maximum likelihood criterion based on an exact observation of X at both all locations in $\tilde{\mathcal{D}}$, and at the 36 locations. We see this in Figure 17, where we have used the same 36 locations as in Figure 16. Both estimates of the semi-variogram functions, either using all locations or 36 locations, appear to be significantly smoother with less fluctuation, as well as more accurate, than the empirical semi-variograms. In theory, when performing MLE, estimating using more observations should yield a better estimate of the true model parameters, so the blue line in Figure 17 should be more alike the black line in Figure 17 than the green line. However, if the 36 random locations are evenly spread out over the domain, they can still produce a good estimate.

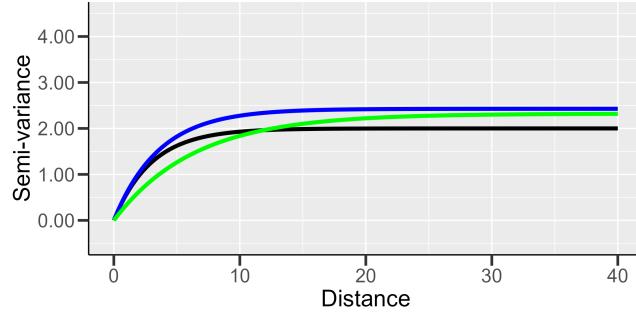
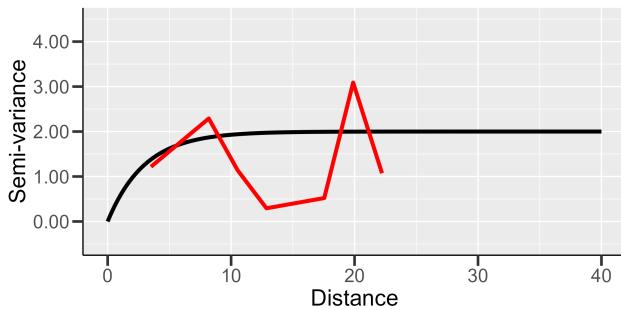


Figure 17: Estimated semi-variogram of MLE based on all locations (blue line), estimated semi-variogram of MLE based on 36 random locations (green line), true semi-variogram function (black line).

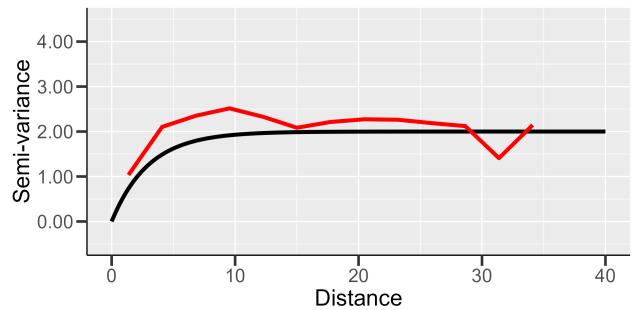
e)

We now perform the same procedure as in d) but with 9 and 100 locations selected uniformly at random. The empirical semi-variogram functions generated for these locations are shown in Figure 18, and their corresponding estimated semi-variogram functions are shown in Figure 19.

While both empirical semi-variograms in Figure 18 show considerable variability, Figure 18a shows greater fluctuation, which is a consequence of having an even more limited number of data points. Meanwhile, this is contrasted by the empirical semi-variogram based on 100 random locations in Figure 18b. Although it is not entirely smooth, it is noticeably more stable, and it stays closer to the true semi-variogram function. This highlights how by increasing the number of data points sampled, we can obtain a more reliable semi-variogram function.



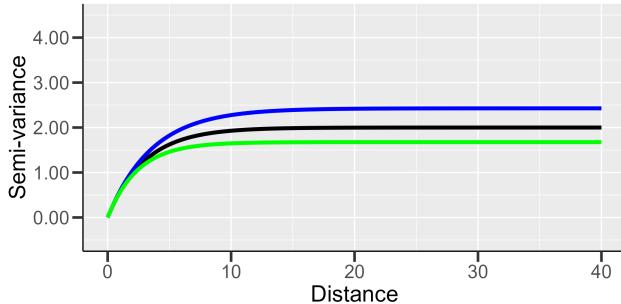
(a) Based on 9 random locations.



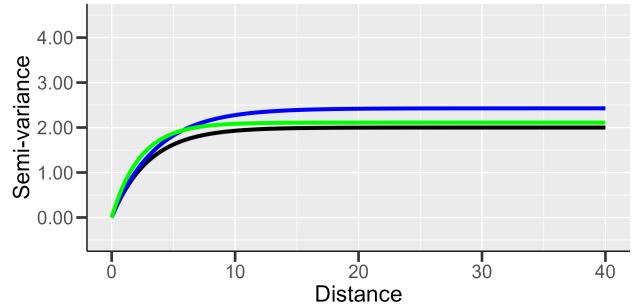
(b) Based on 100 random locations.

Figure 18: Empirical semi-variogram function (red line) based on n random locations, true semi-variogram function (black line).

Figure 19 shows the semi-variograms estimated using MLE compared to the true semi-variogram functions for 9 locations and selected and 100 locations selected. We see from Figure 19a that the estimated semi-variogram function is farther from the true semi-variogram function than the estimated semi-variogram sampled from 100 locations in Figure 19b. As a result, while it is possible to produce a reasonable estimate of the semi-variogram with a fewer number of data points, the accuracy of this estimate does improve as more data is considered.



(a) Based on 9 random locations.



(b) Based on 100 random locations.

Figure 19: Estimated semi-variogram of MLE based on all locations (blue line), estimated semi-variogram of MLE based on n random locations (green line), true semi-variogram function (black line).

f)

In this section, we began by finding the empirical semi-variograms of realizations of GRFs with exponential correlation functions at all locations on the grid. We then compared empirical semi-variograms based on varying numbers of random locations. We assumed that the semi-variogram based on 9 random locations would fit poorest to the true semi-variogram function, and that the semi-variogram based on 100 locations would fit the best, and this was confirmed by the plots produced. Selecting different amounts of random locations could potentially be expanded upon, as it could be interesting to see if choosing more than 100 locations further improves the accuracy of the results. We could explore whether there is a point at which adding more locations does not significantly refine the semi-variogram anymore.

Regarding parameter estimation, we used the `likfit()` function in R to estimate the model parameters, and the results produced were quite accurate to the true parameter values. We further observed that estimated semi-variograms perform smoother and better than empirical semi-variograms. `likfit()` was a somewhat time demanding step while generating the plots, and it would have been interesting to see whether alternative methods could have been used that achieve similar results while taking less time to compute.