GODKJENT

Range [A-B]

Comments: 111
- Very good Report ...
- Minor improvements in text
- Coordinate figures in Prob 3 ?
- Write in present time !!

VERY GOOD ..

TMA4250 - Spatial Statistics, Problem set 1

Mads Adrian Simonsen and Karina Lilleborge

February 2021

Problem 1: Gaussian RF - model characteristics

We consider a continuous spatial variable $\{r(x) : x \in \mathtt{D} : [1, 50] \subset \mathbb{R}^1\}$ which is stationary with the following parameters

$$\begin{aligned} \mathrm{E}[r(x)] &= \mu_r = 0 \\ \mathrm{Var}[r(x)] &= \sigma_r^2 \in \{1, 5\} \\ \mathrm{Corr}[r(x), r(x')] &= \rho_r(\tau; \xi_r, \nu_r), \quad \tau = |x - x'|, \, \xi_r = 10 \end{aligned}$$

We discretize $\mathtt{D}$ into $\mathtt{L} = \{1, 2, ..., 50\}$, and the discretized random field becomes $\{r(x) : x \in \mathtt{L}\}$.

First, we let the spatial correlation function, $\rho_r(\tau)$, be powered exponential with parameter $\nu_r \in \{1, 1.9\}$, that is

$$\rho_r(\tau) = e^{-(\tau/\xi_r)^{\nu_r}}. \tag{1}$$

Secondly, we would like to look at a Matérn spatial correlation function with parameter $\nu_r \in \{1, 3\}$, which gives

$$\rho_r(\tau) = \frac{2^{1-\nu_r}}{\Gamma(\nu_r)} (\tau/\xi_r)^{\nu_r} K_{\nu_r}(\tau/\xi_r). \tag{2}$$

The variogram function, $\gamma_r(\tau)$, is defined by

$$\gamma_r(\tau) = \frac{1}{2} \mathrm{E}[(r(\mathbf{x}') - r(\mathbf{x}))^2] = \sigma_r^2 (1 - \rho_r(\tau)). \tag{3}$$

a)

Let $\rho_r : \mathbb{R}^q \to \mathbb{R}$ be the spatial correlation function, with $q \in \mathbb{N}$. If for any $n \geq 2$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho_r(\mathbf{x}_i - \mathbf{x}_j) \geq 0,$$

for all configurations $\{\mathbf{x}_i \in \mathbb{R}^q : i = 1, \ldots, n\}$,

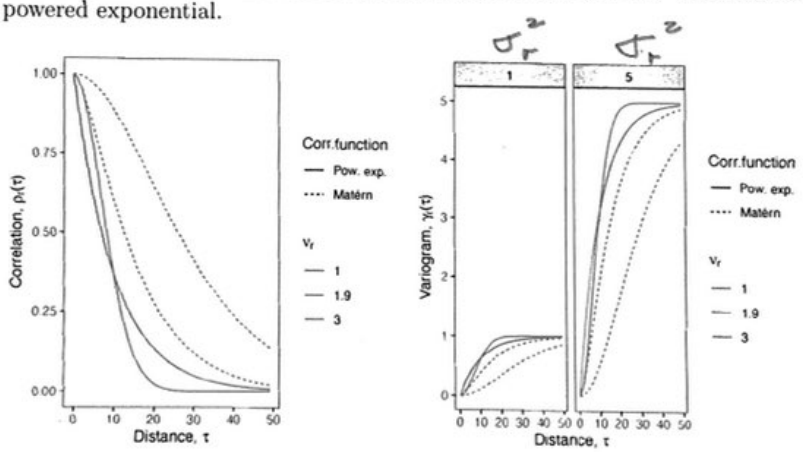
for all weights $\boldsymbol{\alpha} = [\alpha_1 \quad \ldots \quad \alpha_n]^\mathsf{T} \in \mathbb{R}^n$,

then $\rho_r(\boldsymbol{\tau})$ is a non-negative definite function for $\boldsymbol{\tau} \in \mathbb{R}^q$. The reason we require the spatial correlation function to be positive definite is to ensure that all covariance matrices for all configurations and all dimensions are positive definite matrices, such that for any random vector $\mathbf{X} \in \mathbb{R}^n$ and arbitrary vector $\boldsymbol{\gamma} \in \mathbb{R}^n$ we have

$$\boldsymbol{\gamma}^\top \mathrm{Var}[\mathbf{X}]\boldsymbol{\gamma} = \mathrm{Var}[\boldsymbol{\gamma}^\top \mathbf{X}] \geq 0.$$

We observe the two different spatial correlation functions with the two different choices for ν_r. This is displayed in Figure 1a. We note that the powered exponential has a shorter range for the correlation than the Matérn spatial correlation. We observe that the choice of ν_r for the powered exponential affects the correlation differently for shorter and longer distances. On a shorter distance, $\tau \in (0, 10)$, the larger choice for ν_r makes a stronger correlation, but for $\tau > 10$ we notice that $\nu_r = 1$ gives a larger correlation than for $\nu_r = 1.9$. For the Matérn spatial correlation function we note that a larger value for ν_r makes the spatial correlation larger for all distances.

Figure 1b shows the variogram function for both $\sigma_r^2 = 1$ and 5. As we would expect from the definition of this variogram function in (3), the value of σ_r^2 only scales the variogram function. Since this is just the inverted correlation function, we observe that the Matérn correlation has a slower rate than the powered exponential.



(a) Correlation.

(b) Variogram.

Figure 1: Correlation and variogram grid plots for both powered exponential (solid line) and Matérn (dashed line) as the spatial correlation function. The powered exponential is shown with parameter $\nu_r \in \{1, 1.9\}$, and the Matérn with $\nu_r \in \{1, 3\}$. The variogram is divided for $\sigma_r^2 = 1$ and $\sigma_r^2 = 5$.

b)

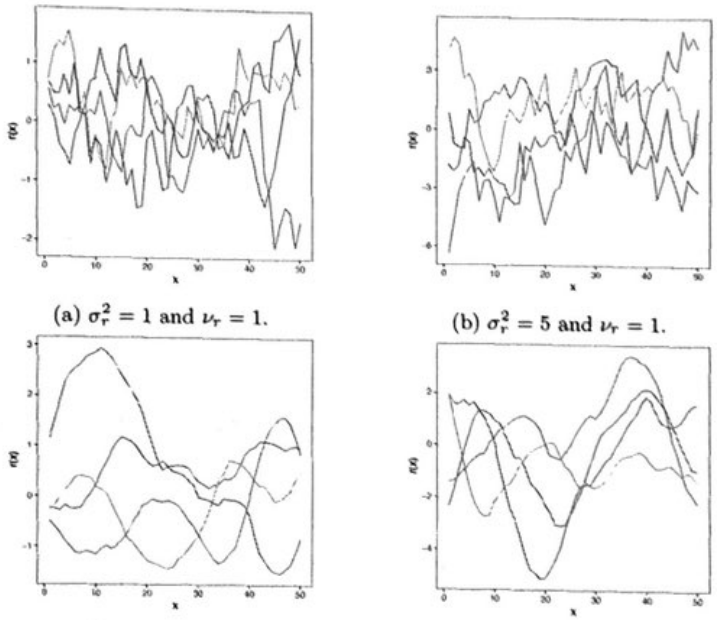
The prior Gaussian RF is defined on the discretized representation

$$\mathbf{r} = \begin{bmatrix} r_1 & \dots & r_n \end{bmatrix}^\mathsf{T} = \begin{bmatrix} r(x_1) & \dots & r(x_n) \end{bmatrix}^\mathsf{T} \in \mathbb{R}^n,$$

with $n = 50$ as

$$\mathbf{r} \sim p(\mathbf{r}) = \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \boldsymbol{\Sigma}_r^\rho) = \phi_n(\mathbf{r}; \mathbf{0}, \sigma_r^2 \boldsymbol{\Sigma}_r^\rho), \tag{4}$$

where the ith row and jth column of $\boldsymbol{\Sigma}_r^\rho$ is given by $\rho_r(|x_i - x_j|)$.



(a) $\sigma_r^2 = 1$ and $\nu_r = 1$.

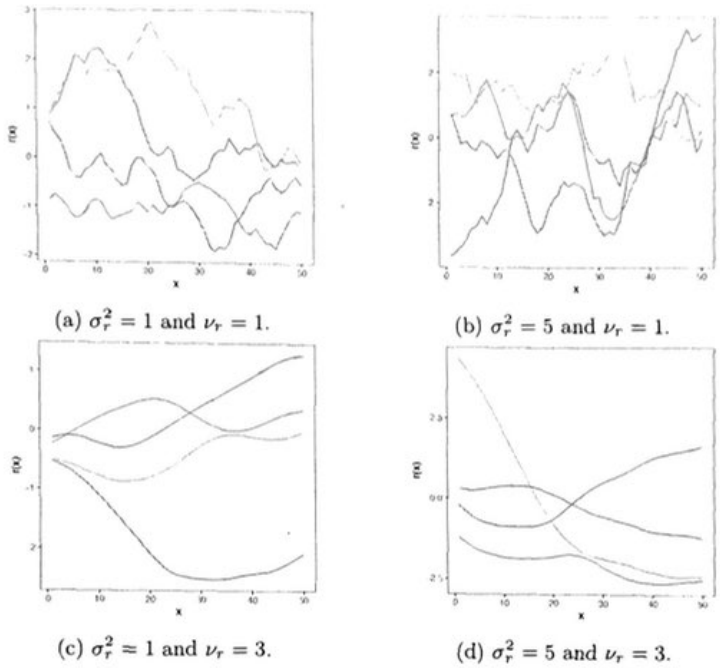
(b) $\sigma_r^2 = 5$ and $\nu_r = 1$.

(c) $\sigma_r^2 = 1$ and $\nu_r = 1.9$.

(d) $\sigma_r^2 = 5$ and $\nu_r = 1.9$.

Figure 2: Four realizations of the Gaussian random field, $r(x)$ from each of the four configurations for ν_r and σ_r^2 when the correlation function is powered exponential.

We sample from the distribution in (4), and vary σ_r^2, the spatial correlation function and ν_r. Figure 2 shows four realizations for each set of parameters, where the spatial correlation function is powered exponential and $\sigma_r^2 \in \{1, 5\}$ and $\nu_r \in \{1, 1.9\}$. The first two subfigures, 2a and 2b, visualizes the effect of increasing σ_r^2. They show that the values in each grid point have a larger range for larger spatial variance. We also note that the choice of $\nu_r = 1.9$ makes the spatial correlation highly noticeable, as all the points are staying relatively close to their closest neighbours.



(a) $\sigma_r^2 = 1$ and $\nu_r = 1$.

(b) $\sigma_r^2 = 5$ and $\nu_r = 1$.

(c) $\sigma_r^2 = 1$ and $\nu_r = 3$.

(d) $\sigma_r^2 = 5$ and $\nu_r = 3$.

Figure 3: Four realizations from each of the four configurations for ρ_r and σ_r^2 when the correlation function is Matérn.

Similarly, Figure 3 shows four realizations for each model parameter set, that is $\sigma_r^2 \in \{1, 5\}$ and $\nu_r \in \{1, 3\}$. In the realizations for the two first model parameter sets, shown in Subfigures 3a and 3b, we see some correlation between the points, even when $\nu_r = 1$. The range for the realizations is also larger for the ones in Subfigure 3b than for those in Subfigure 3a. This is due to the increase in σ_r^2. When we increase ν_r we see a clearer spatial correlation between grid points. This is shown in Subfigures 3c and 3d.

c)

Let the spatial variable be observed as $\{d(x) : x \in 10, 25, 30 \subset \mathtt{L}\}$, according to the acquisition model,

$$d(x) = r(x) + \epsilon(x), \quad x \in \{10, 25, 30\},$$

where the measurement errors $\epsilon(x) \sim \mathrm{N}(0, \sigma_\epsilon^2)$ are independent. We also assume that $\epsilon(x)$ and $r(x')$ are independent for all x and x'. The Gauss-linear likelihood model relative to the discretized spatial variable $\mathbf{r} \in \mathbb{R}^n$, with $n = 50$ and the

observed data $\mathbf{d} \in \mathbb{R}^m$, with $m = 3$ is then given by

$$[\mathbf{d} \mid \mathbf{r}] = \mathbf{H}_0\mathbf{r} + \boldsymbol{\epsilon} \sim p(\mathbf{d} \mid \mathbf{r}) = \phi_m(\mathbf{d}; \mathbf{H}_0\mathbf{r}, \sigma_\epsilon^2\mathbf{I}_m). \tag{5}$$

The point observation matrix $\mathbf{H}_0 \in \mathbb{R}^{m\times n}$ indicates where the data was observed. $\mathbf{H}_0$ consists of three ones, located at column 10, 25 and 30 in row one, two and three respectively, while the rest of the entries are zeros.

The likelihood $p(\mathbf{d} \mid \mathbf{r})$ in (5) is a function of the parameter $\mathbf{r}$ given the data $\mathbf{d}$ which is constant, and is often written as $\mathcal{L}(\mathbf{r}; \mathbf{d})$. Therefore (5) is not a pdf, and its integral with respect to the parameter $\mathbf{r}$ does not necessarily integrate to 1.

We select one of the realization with powered exponential correlation function with parameter $\nu = 1.9$ and $\sigma_r^2 = 5$ and get the following values at the three locations

$$r(10) = 0.2898, \quad r(25) = 0.0087 \quad \text{and} \quad r(30) = 1.0487. \tag{6}$$

d)

The posterior model is given by

$$[\mathbf{r} \mid \mathbf{d}] \sim p(\mathbf{r} \mid \mathbf{d}) = \phi_n\left(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d}\right), \tag{7}$$

where

$$\boldsymbol{\mu}_{r|d} = \boldsymbol{\Sigma}_r^\rho \mathbf{H}_0^\mathsf{T}\left(\mathbf{H}_0\boldsymbol{\Sigma}_r^\rho\mathbf{H}_0^\mathsf{T} + \frac{\sigma_\epsilon^2}{\sigma_r^2}\mathbf{I}_m\right)^{-1}\mathbf{d}, \tag{8}$$

and

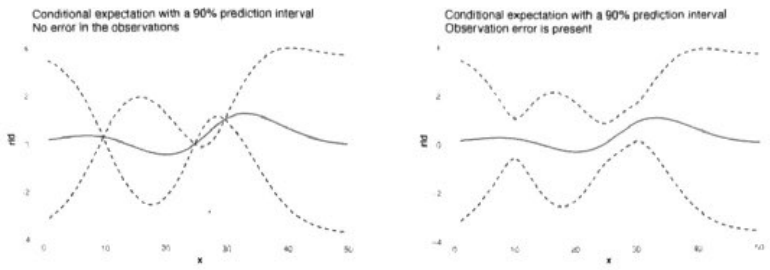
$$\boldsymbol{\Sigma}_{r|d} = \sigma_r^2\boldsymbol{\Sigma}_r^\rho\left(\mathbf{I}_n - \mathbf{H}_0^\mathsf{T}\left(\mathbf{H}_0\boldsymbol{\Sigma}_r^\rho\mathbf{H}_0^\mathsf{T} + \frac{\sigma_\epsilon^2}{\sigma_r^2}\mathbf{I}_m\right)^{-1}\mathbf{H}_0\boldsymbol{\Sigma}_r^\rho\right). \tag{9}$$

Using the observations given in (6), we compute the conditional mean and variance from (8) and (9), with $\sigma_\epsilon^2 = 0$, and also with observation error, $\sigma_\epsilon^2 = 0.25$. The result is shown in Subfigures 4a and 4b.

We observe that in Subfigure 4a the 90% prediction is zero in all the observed points, due to $\sigma_\epsilon^2 = 0$. The uncertainty increases quickly outside these points.

In Subfigure 4b there are some uncertainty related to all points in the grid, but it clearly decreases the closer we get to the observation points.

For both predictions in Figure 4, except in the observation points, we see that the conditional mean and prediction interval are almost identical. Looking at the equations (8) and (9) we note that the term $\sigma_\epsilon^2/\sigma_r^2$, which gives $0.25/5 = 0.05$ and $0/5 = 0$, for the realization with and without observation error respectively, which is a small difference.



(a) No observation error. (b) With observation error $\sigma_\epsilon^2 = 0.25$.

Figure 4: Conditional on observations in $x \in \{10, 25, 30\}$, we do prediction on the whole domain. Red line shows the mean, and the blue dashed line is the upper and lower 0.9-prediction interval.

e)

We do 100 realizations from the posterior distribution. Based on these realization we estimate the mean and covariance matrix empirically, and compute an empirical 90% prediction interval. The result is shown in Subfigures 5a and 5b. Comparing the two, we see that the choice of σ_ϵ^2 is affecting the length of the prediction interval in the observations points. For no observation error, the interval closes in on the exact observed points, and all 100 realizations are going through these values.

Similarly, for Subfigure 5b we see the estimated prediction interval narrowing for the values where we have observations. But, since this model has uncertainty related to these observations, there are realizations that are not going through the observed points. In both plots we observe that the different realizations are differing from the mean more when being further from the observed points, which is what we would expect as the correlation here is weaker to the observed points, and the uncertainty is mostly based on the random field, $\sigma_r^2 = 5$.

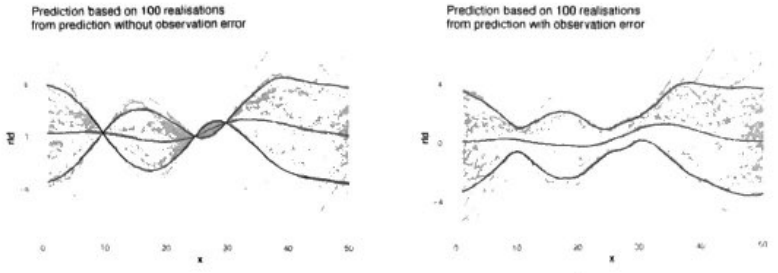
If we compare between Subfigures 4a and 5a we notice that from 100 realizations, we get a good estimate of the prediction and prediction interval. As we would expect for a 90% prediction interval, some of the realizations are outside of the estimated (and therefore also the analytical) prediction interval for some grid points.

f)

Let us consider the non-linear function on $\{r(x); x \in \mathtt{D}\}$;

$$A_r = \sum_{x \in \mathtt{L}} I(r(x) > 2)\,(r(x) - 2)\,, \tag{10}$$

which approximates the area under the spatial variable and above level 2. We sample 100 realizations from the posterior model with $\sigma_\epsilon^2 = 0$ to compute an



(a) Without observation error. (b) With observation error.

Figure 5: One-hundred realizations (in light blue) based on the posterior mean (in red) and covariance matrix. A 90% prediction interval for these predictions is shown in dark blue.

estimate $\hat{A}_r \approx 7.02$. From Subfigure 5a we would expect this approximation to be positive, as we see that many realizations go above $r(x) = 2$.

As an alternative approach to estimate A_r, we use the predicted spatial variable $\hat{\mathbf{r}} \in \{\hat{r}(x) : x \in \mathtt{L}\}$, with $\sigma_\epsilon^2 = 0$ such that

$$\tilde{A}_r = \sum_{x \in \mathtt{L}} I(\hat{r}(x) > 2)(\hat{r}(x) - 2).$$

From the ~~realizations~~ predictions we find $\tilde{A}_r = 0$. From Subfigure 5a we see that the empirical prediction line is always less than 2, so this coincides with our expectation.

The function

$$f(z) = I(z > 2)(z - 2) = \begin{cases} z - 2, & \text{if} \quad z > 2, \\ 0, & \text{otherwise,} \end{cases}$$

is continuous and consists of a constant part and a linear part. Since the slope of $f(z)$ is increased when $z > 2$, $f(z)$ must be convex. Using the fact that the sum of convex functions is convex, we have that $g(\mathbf{r}) := A_r$ is convex with respect to $\mathbf{r}$. Then by Jensen's inequality, we have that

$$\mathrm{E}\left[g(\mathbf{r})\right] \geq g\left(\mathrm{E}[\mathbf{r}]\right)$$

$$\mathrm{E}\left[\sum_{x \in \mathtt{L}} I(r(x) > 2)\left(r(x) - 2\right)\right] \geq \sum_{x \in \mathtt{L}} I\left(\mathrm{E}[r(x)] > 2\right)\left(\mathrm{E}[r(x)] - 2\right).$$

Assuming $\mathrm{E}[r(x)] \approx \widehat{\mathrm{E}}[r(x)] = \hat{r}(x)$ and $\mathrm{E}[g(\mathbf{r})] \approx \widehat{\mathrm{E}}[g(\mathbf{r})] = \hat{A}_r$, we expect that

$$\hat{A}_r \geq \tilde{A}_r,$$

which is the case for our computations.

g)

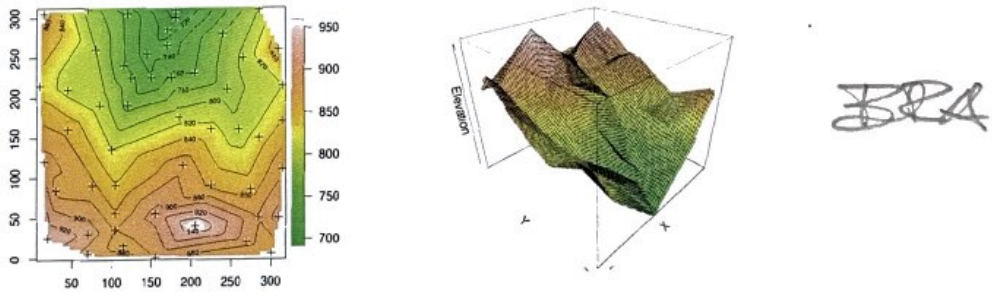
Through computations and visualization we have observed how the powered exponential and Matérn correlation function affect realizations of a 1D random field. Secondly, we saw how the choice of model parameters affects the realizations. We have experienced how exact observations in some grid points affect the prediction on a whole grid, both without and with observation error. Then we learned how observations in some points makes prediction on the whole domain more certain, and the more precise the observations, the better prediction can be made. Furthermore we saw how well the approximated prediction and prediction interval was from only 100 realizations. Lastly, we had a practical example of Jensen's inequality, when comparing estimates on the posterior.

Problem 2: Gaussian RF - real data

We consider observations of terrain elevation from the `topo` data found in the `MASS` library in `R`. The data consists of $n_d = 52$ observations located in the domain $\mathtt{D} = [(0, 315) \times (0, 315)] \subset \mathbb{R}^2$ measured in ft. Let the n_d-vector of exact observations be $\mathbf{d} = \begin{bmatrix} r(\mathbf{x}_1^o) & \dots & r(\mathbf{x}_{n_d}^o) \end{bmatrix}^\mathsf{T}$.

a)

Two plots of the observations $\mathbf{d}$ can be seen in Figure 6.



(a) Contour plot. (b) Surface plot.

Figure 6: Contour plot and surface plot showing the elevation of the terrain. The exact observations $\mathbf{d}$ are denoted by $*$. The remaining area is computed by linear interpolation.

The observations lies between 700 ft and 950 ft in elevation, with the mean elevation being about 830 ft. Although the variance seem to be constant in `D`, it appears to be a downhill trend in the y-direction, which would violate the stationary assumption. However, this trend could also be a coincidence and part of a realization of a stationary Gaussian process with a higher variance.

b)

We model the terrain elevation by the Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D} \subset \mathbb{R}^2\}$ with

$$\begin{aligned} \mathrm{E}[r(\mathbf{x})] &= \mathbf{g}(\mathbf{x})^\mathsf{T}\boldsymbol{\beta}_r, \\ \mathrm{Var}[r(\mathbf{x})] &= \sigma_r^2, \\ \mathrm{Corr}[r(\mathbf{x}), r(\mathbf{x}')] &= \rho_r(\tau), \end{aligned} \tag{11}$$

where $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} 1 & g_2(\mathbf{x}) & g_3(\mathbf{x}) & \dots & g_{n_g}(\mathbf{x}) \end{bmatrix}^\mathsf{T}$ is a n_g-vector of related spatial variables in the reference domain $\mathtt{D}$, and $\boldsymbol{\beta}_r = \begin{bmatrix} \beta_1 & \dots & \beta_{n_g} \end{bmatrix}^\mathsf{T}$ is a n_g-vector of unknown parameters. Let the variance and the spatial correlation function be

$$\begin{aligned} \sigma_r^2 &= 50^2 = 2500, \\ \rho_r(\tau) &= \exp\left(-(\tau/\xi)^\nu\right), \quad \nu = 1.5,\ \xi = 100, \end{aligned}$$

where $\tau = \|\mathbf{x} - \mathbf{x}'\|_2$.

The universal Kriging model assumes that the model parameters $\boldsymbol{\beta}$ are unknown, while the parameters (σ_r^2, ξ, ν) are known. The Kriging predictor is defined to be the best linear unbiased predictor (BLUP), and has the form

$$\hat{r}_0 = \hat{\boldsymbol{\alpha}}^\mathsf{T}\mathbf{d} \tag{12}$$

To attain an unbiased predictor $\hat{r}_0 = \hat{r}(\mathbf{x}_0)$ for an arbitrary location $\mathbf{x}_0 \in \mathtt{D}$, we require

$$\begin{aligned} & \mathrm{E}[\hat{r}_0 - r_0] = \hat{\boldsymbol{\alpha}}^\mathsf{T}\mathrm{E}[\mathbf{d}] - \mathrm{E}[r_0] = 0 \\ \Rightarrow \quad & \hat{\boldsymbol{\alpha}}^\mathsf{T}\mathbf{G}_d\boldsymbol{\beta}_r = \mathbf{g}_{\mathbf{x}_0}^\mathsf{T}\boldsymbol{\beta}_r \\ \Rightarrow \quad & \mathbf{G}_d^\mathsf{T}\hat{\boldsymbol{\alpha}} = \mathbf{g}_{\mathbf{x}_0}, \end{aligned} \tag{13}$$

where $\mathbf{G}_d = \begin{bmatrix} \mathbf{g}_{\mathbf{x}_1} & \dots & \mathbf{g}_{\mathbf{x}_{n_d}} \end{bmatrix}^\mathsf{T} = \begin{bmatrix} \mathbf{g}(\mathbf{x}_1^o) & \dots & \mathbf{g}(\mathbf{x}_{n_d}^o) \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{n_d \times n_g}$. We want to find the weights $\boldsymbol{\alpha}$ that gives the BLUP.

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}} \{\mathrm{Var}[\hat{r}_0 - r_0]\} \\ &= \arg\min_{\boldsymbol{\alpha}} \left\{\boldsymbol{\alpha}^\mathsf{T}\mathrm{Var}[\mathbf{d}]\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\mathsf{T}\mathrm{Cov}[\mathbf{d}, r_0] + \mathrm{Var}[r_0]\right\} \\ &= \arg\min_{\boldsymbol{\alpha}} \left\{\sigma_r^2\left(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{\Sigma}_d^\rho\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{\rho}_0 + 1\right)\right\}, \quad \text{constrained by (13)}, \end{aligned}$$

where $\boldsymbol{\Sigma}_d^\rho = (1/\sigma_r^2)\,\mathrm{Var}[\mathbf{d}]$, and $\boldsymbol{\rho}_0 = \mathrm{Corr}[\mathbf{d}, r_0]$. That is, we have a quadratic object function and a linear equality constraint, which then can be solved analytically by Lagrange multipliers. The solution is

$$\hat{\boldsymbol{\alpha}} = [\boldsymbol{\Sigma}_d^\rho]^{-1}\left[\boldsymbol{\rho}_0 + \mathbf{G}_d\left(\mathbf{G}_d^\mathsf{T}[\boldsymbol{\Sigma}_d^\rho]^{-1}\mathbf{G}_d\right)^{-1}\left(\mathbf{g}_{\mathbf{x}_0} - \mathbf{G}_d^\mathsf{T}[\boldsymbol{\Sigma}_d^\rho]^{-1}\boldsymbol{\rho}_0\right)\right], \tag{14}$$

giving the Kriging predictor and prediction variance

$$\hat{r}_0 = \hat{\boldsymbol{\alpha}}^\mathsf{T}\mathbf{d}, \tag{15}$$

$$\sigma_{\hat{r}}^2 = \sigma_r^2\left(\hat{\boldsymbol{\alpha}}^\mathsf{T}\boldsymbol{\Sigma}_d^\rho\hat{\boldsymbol{\alpha}} - 2\hat{\boldsymbol{\alpha}}^\mathsf{T}\boldsymbol{\rho}_0 + 1\right). \tag{16}$$

A question that might come to mind is what kind of parametrization for the expectation function $\mathrm{E}[r(\mathbf{x})]$ is suitable for this model. The more terms we allow for the parametrization, e.g. by adding more degrees to a polynomial, the more would the expectation function capture the trend from the observations, and thus it would be natural for the variance, σ_r^2, to be lower. We will consider two parametrizations, the ordinary Kriging with constant mean, and a universal Kriging where the expectation function is a 2nd order polynomial.

c)

We consider the ordinary Kriging model, that is,

$$\mathrm{E}[r(\mathbf{x})] = \beta_1, \tag{17}$$

which is a special case of the universal Kriging model. That is, when $n_g = 1$, such that (11) gives us

$$\mathbf{g}(\mathbf{x}) = 1, \ \forall \mathbf{x} \in \mathtt{D},$$
$$\mathbf{G}_d = \mathbf{i}_{n_d},$$
$$\boldsymbol{\beta}_r = \beta_1.$$

Inserting this into (14) gives us the BLUP weights

$$\hat{\boldsymbol{\alpha}} = [\boldsymbol{\Sigma}_d^\rho]^{-1}\left[\boldsymbol{\rho}_0 + \left(\mathbf{i}_{n_d}^\mathsf{T}[\boldsymbol{\Sigma}_d^\rho]^{-1}\mathbf{i}_{n_d}\right)^{-1}\left(1 - \mathbf{i}_{n_d}^\mathsf{T}[\boldsymbol{\Sigma}_d^\rho]^{-1}\boldsymbol{\rho}_0\right)\mathbf{i}_{n_d}\right], \tag{18}$$

which we insert into (15) and (16), to attain the Kriging predictor and prediction variance, respectively.

The result of this model is seen in Figure 7, where the estimated prior mean is given by

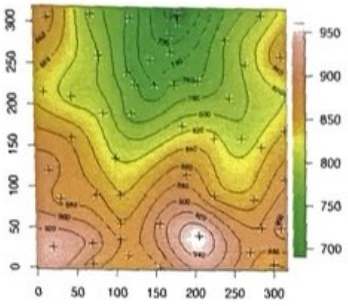
$$\hat{\mathrm{E}}[r(\mathbf{x})] = \hat{\beta}_1 = 845.2479.$$

Both the contour plot and surface plot of the Kriging predictor $\hat{r}$ shown in Subfigure 7a and Subfigure 7b, is similar to the plots in Figure 6, but is smooth and looks more like a realistic terrain. In the Subfigure 7c and Subfigure 7d, we see the prediction standard deviation $\sqrt{\sigma_{\hat{r}}^2}$, showing that in the darker areas where we have the observations, the variance is zero.

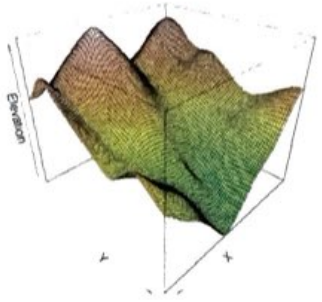
d)

Let the reference variable $\mathbf{x} \in \mathtt{D} \subset \mathbb{R}^2$ be denoted $\mathbf{x} = (x_v, x_h)$. We now consider the universal Kriging model, where the n_g-vector of related spatial variables is given by

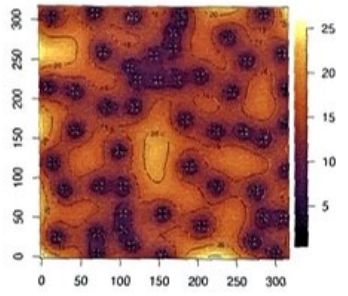
$$\mathbf{g}(\mathbf{x}) = \mathbf{g}(x_v, x_h) = \begin{bmatrix}1 & x_v & x_h & x_v x_h & x_v^2 & x_h^2\end{bmatrix}^\mathsf{T} \in \mathbb{R}^{n_g}, \tag{19}$$



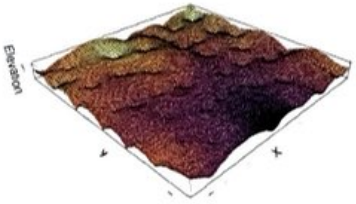
(a) Kriging predictor contour plot.



(b) Kriging predictor surface plot.



(c) Prediction std. dev. contour plot.



(d) Prediction std. dev. surface plot.

Figure 7: Contour plot and surface plot of the ordinary Kriging predictor $\hat{r}$ in the upper row (a) and (b), and prediction standard deviation $\sqrt{\sigma_{\hat{r}}^2}$ in the lower row (c) and (d), together with the exact observations $\mathbf{d}$ in the contour plots, denoted by $*$.

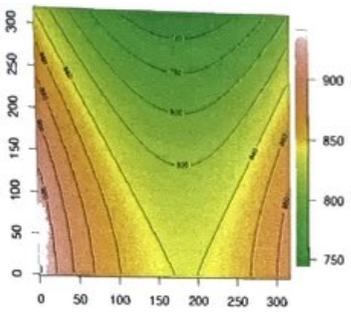
with $n_g = 6$. Then the expected value of $\mathbf{r}(\mathbf{x})$ becomes

$$\mathrm{E}[r(\mathbf{x})] = \mathbf{g}(\mathbf{x})^\mathsf{T}\boldsymbol{\beta}_r = \beta_1 + \beta_2 x_v + \beta_3 x_h + \beta_4 x_v x_h + \beta_5 x_v^2 + \beta_6 x_h^2. \tag{20}$$

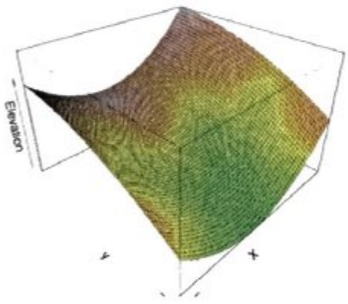
The result of this model is seen in Figure 8, where the estimated parameters are given by

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 941.66805 & -1.10908 & -0.05182 & 0.00013 & 0.00301 & -0.00087 \end{bmatrix}^\mathsf{T}.$$

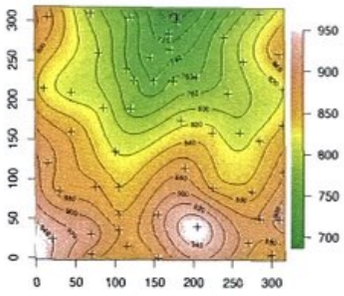
Figure 8a and Figure 8b shows an estimation of the expected prior given by (20). We see that this prior captures the downhill trend along the y-axis, which we mentioned in Subsection **a**). The Kriging predictor $\hat{r}$ and prediction standard deviation $\sqrt{\sigma_{\hat{r}}^2}$ shown in the middle row and bottom row, respectively in Figure 8 is almost identical to the one in the ordinary Kriging model shown in Figure 7. However, one can see a slight difference in the neighborhood of the coordinate $(x_v, x_h) = (0, 0)$. On the ordinary Kriging model, we can see the mode in this corner, but on the universal Kriging it keeps increasing according to the prior



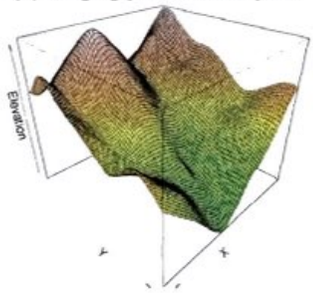
(a) Kriging prior contour plot.



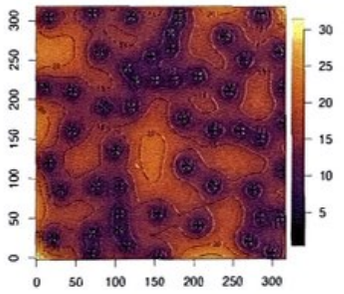
(b) Kriging prior surface plot.



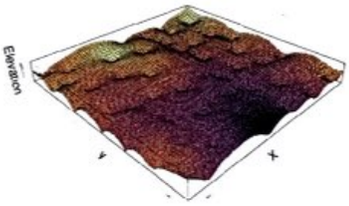
(c) Kriging predictor contour plot.



(d) Kriging predictor surface plot.



(e) Prediction std. dev. contour plot.



(f) Prediction std. dev. surface plot.

Figure 8: Contour plot and surface plot of the universal Kriging prior $\mathrm{E}[r(\mathbf{x})]$ in the upper row (a) and (b), the Kriging predictor $\hat{r}$ in the middle row (c) and (d), and prediction standard deviation $\sqrt{\sigma^2_{\hat{r}}}$ in the lower row (e) and (f), together with the exact observations $\mathbf{d}$ in the contour plots, denoted by $*$.

mean. This results in a slight increase in the variance(standard deviation) at this location, and can be thought of as a mild version of Runge's phenomenon.

e)

Let us consider the ordinary Kriging model specified in subsection **c)**, and the grid node $\mathbf{x}_0 = (100, 100)$. The Kriging predictor and prediction variance at this location is found by (15) and (16) with $\hat{\boldsymbol{\alpha}}$ computed from (18).

$$\hat{r}_0 = 838.6781$$
$$\sigma^2_{\hat{r}} = 100.8866.$$

We wish to calculate the probability for the elevation to be higher than 850 m at this location. That is,

$$\begin{aligned}
\Pr(r_0 > 850 \mid \mathbf{d}) &= \Pr(r(\mathbf{x}_0) > 850 \mid \mathbf{d}) \\
&= 1 - \Phi\left(\frac{850 - \mathrm{E}[r(\mathbf{x}_0) \mid \mathbf{d}]}{\sqrt{\mathrm{Var}[r(\mathbf{x}_0) \mid \mathbf{d}]}}\right) \\
&= 1 - \Phi\left(\frac{850 - \hat{r}_0}{\sqrt{\sigma^2_{\hat{r}}}}\right) \\
&\approx 0.129829,
\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cdf.

We also want to compute the elevation z for which there is a 0.90 probability the true elevation is below it. That is,

$$\begin{aligned}
0.90 &= \Pr(r_0 < z \mid \mathbf{d}) \\
&= \Phi\left(\frac{z - \hat{r}_0}{\sqrt{\sigma^2_{\hat{r}}}}\right) \\
\Rightarrow \quad z &= \hat{r}_0 + \sqrt{\sigma^2_{\hat{r}}}\Phi^{-1}(0.90) \\
&= 851.5503.
\end{aligned}$$

f)

We have investigated real data with 52 observations of terrain elevation, and applied two Kriging models on the data, where in the first model we assumed the prior model to be a stationary Gaussian RF, with mean $\mathrm{E}[r(\mathbf{x})] = \beta_1$, and in the other we assumed the prior to have a quadratic trend. Both of the models gave almost identical outputs, with a minor difference in one of the corners in the domain $\mathtt{L}$. It would be interesting to investigate higher order trends on the data, to check if there are significant differences. Also, we assumed the parameters σ^2_r, ν and ξ to be known. It would be interesting to estimate these parameters and test their significance.

Problem 3: Parameter estimation

We consider a stationary Gaussian random field, $\{r(\mathbf{x}); \mathbf{x} \in \mathtt{D} \subset \mathbb{R}^2\}$, where $\mathtt{D} : [(1,30) \times (1,30)]$. The field parameters are

$$\begin{aligned} \mathrm{E}[r(\mathbf{x})] &= \mu_r = 0 \\ \mathrm{Var}[r(\mathbf{x})] &= \sigma_r^2 \\ \mathrm{Corr}[r(\mathbf{x}), r(\mathbf{x}')] &= \exp(-\tau/\xi_r), \end{aligned} \tag{21}$$

with $\tau = \|\mathbf{x} - \mathbf{x}'\|_2$.

a-c)

We discretize the random field to $\{r(\mathbf{x}) : \mathbf{x} \in \mathtt{L}\}$ where $\mathtt{L}$ is the grid over $\mathtt{D}$ with 100×100 grid points, and let the variance $\sigma_r^2 = 2$ and the scale parameter $\xi_r = 3$. Figure 9 shows three realization of this Gaussian random field.

Based on the realizations in Figure 9 we compute the empirical variogram functions. A comparison of the empirical variogram functions to the exact function is shown in Figure 10. This shows that the different realizations give a similar result for each empirical variogram. They are all following the real variogram function quite closely for small distances. For distances larger than 30 they are all differing from the exact function. This is due to the number of points to base the estimation on. For such distances there are far less grid points to estimate with, compared to distances shorter than the length of the domain.

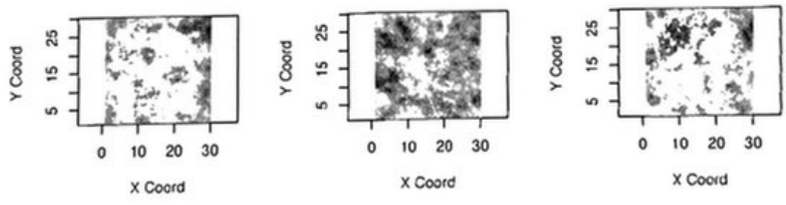


Figure 9: Three realizations of a Gaussian random field with parameters specified in (21).

d, e)

We now refer to $\mathtt{L}$ as an integer grid on $\mathtt{D}$, that is a 30×30 grid. We generate a realization on this grid. From this realization we draw 36 uniformly random locations (without replacement) and use exact observations in these points to

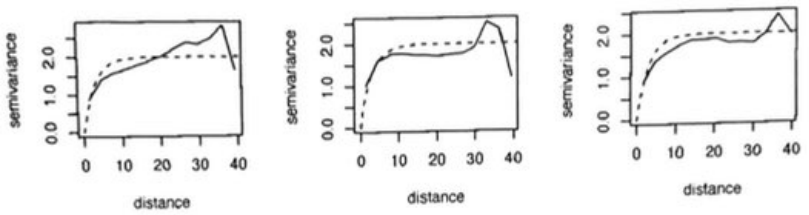


Figure 10: Three empirical variograms based on the realizations shown in Figure 9.

compute the empirical variogram function and compare it to the exact function. The result is displayed in Figure 11. From this, we see that the empirical variogram is close to the exact function for smaller distances, but for distances larger than 20 we observe that it differs from the exact function. The reason behind this is, as discussed before, that we have fewer points to base our estimation on, and the random points might not have any points with some distances. The computations are therefore worse than when we have multiple sets of points with the desired distance.

Next we repeat the procedure, but with 9, 64 and 100 uniformly random locations in the grid. The empirical variogram estimates are shown in Figure 12. To the left, we see that the empirical variogram is far off, compared to the exact function. This is, as discussed for the result in Figure 11, likely due to the lack of points to do the estimation on with the desired distances. With only 9 randomly chosen points in the grid, it is difficult to say something about the relation between points on the whole grid. We note that there are no pair of points with distance smaller than 5, and the empirical variogram has no estimates for these distances. The middle and right subfigures shows empirical variogram for 64 and 100 uniformly random points on the grid, respectively. It is clear that these empirical variograms are better approximations to the exact variogram function than the ones with 9 and 36 points. It is reasonable from this result to say that the estimated variogram is better if it is based on more randomly chosen points.

From now, we consider the model parameters σ_r^2 and ξ_r to be unknown. We would like to estimate these using the maximum likelihood criterion. To do so, we use the `likfit()` function from the `geoR` library.

We estimated the model parameters σ_r^2 and ξ_r using the third realization from Figure 9, and the same uniformly random chosen points from the grid as in Figure 11 and Subfigure 12. Then we use these model parameters in the exact variogram function, and the result is shown in Figure 13. We note that estimation based on 100 random points performs the best. Then we have the estimation based on 9 random grid points, which differ from the exact function.

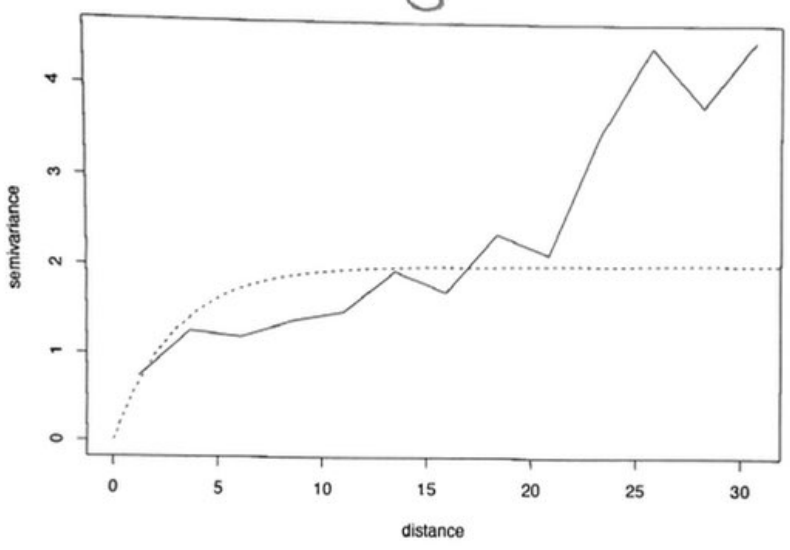


Figure 11: The empirical variogram based on 36 uniformly randomly chosen points from the third realization in Figure 9.

It increases drastically for the first values for τ, and then stabilizes right beneath 2.0. This means that the correlation quickly decreases and stabilizes at a low level for larger distances. The estimated variogram based on 36 and 64 points both have a lower rate than the exact variogram function. For larger distances, the first seems to be closing in on the exact function, while the one based on 64 points stabilizes quickly at about 1.1. This means that the correlation is constant for distances larger than about 20. And this correlation is just below 0.5. Why this is the case for these 64 points could be random, and we might have a bad foundation to make assumptions about larger distances. As we saw in Figure 12, the empirical variogram function is way off for the larger distances, already at $\tau = 20$ it starts to differ more from the exact function. Then, when the estimation of σ_r^2 and ξ_r is based on this set of points, the estimation will not be satisfactorily. From this we note that the set of points we use for the estimate is crucial to get a good estimate of the field parameters. The set of points should represent the field well, to make a good estimation.

f)

In this section we have experienced how well empirical variograms based on realizations might perform. The empirical variogram functions based on the

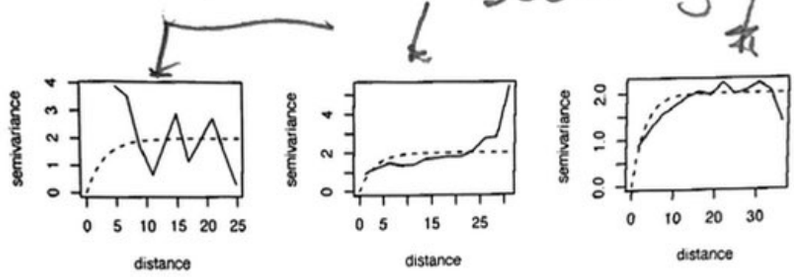


Figure 12: The empirical variograms, from left to right, based on 9, 64 and 100 uniformly randomly chosen points from the third realization in 9.

full, smooth grid performs very well for multiple realizations. Further, we looked into how only choosing a number of uniformly random selection of points from the grid might perform for this estimate. We also compared how the number of points we select has an impact on the estimate as well. In our experience 100 uniformly random points on the grid $(1, 30) \times (1, 30)$ performed the best. With this result in mind, it could be interesting to see if a larger number of randomly chosen grid points could perform even better than the 100 we used here. Further, this could be repeated on more realizations to see if there could be some optimal choice - or determine if such optimal choice exists.

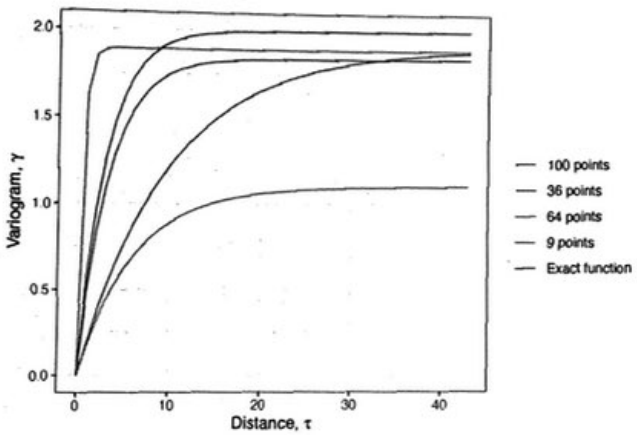


Figure 13: Estimations of the parameters σ_r^2 and ξ_r based on the realization shown in 9, and 9, 36, 64 and 100 uniformly random chosen grid points used to compute estimated variogram functions. They are plotted against the true variogram function from (21).