Spatial Statistics - Exercise 1

Karine Hagesæther Foss, Martin Outzen Berild, Sindre Henriksen

February 2019

Problem 1

We consider a continuous spatial variable in $\mathbb{R}^1$ s.t.

$$\begin{aligned}\{r(x) : x \in \mathrm{D} : [1, 50] \subset \mathbb{R}^1\}\\ \mathbf{E}\{r(x)\} = \mu_r = 0\\ \mathrm{Var}\{r(x)\} = \sigma_r^2\\ \mathrm{Corr}\{r(x), r(x')\} = \rho_r(\tau)\,,\end{aligned}$$

where $\tau = |x - x'|/10$. Discretization: $\mathrm{L} \in \{1, 2, ..., 50\}$. The discretized random field (RF) is

$$\{r(x) : x \in \mathrm{L}\}.$$

a) A function $c(\boldsymbol{\tau}) : \mathbb{R}^q \to \mathbb{R}$ is positive definite (non-negative definite, to be precise) for $\boldsymbol{\tau} \in \mathbb{R}^q$ if

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j c(\mathbf{x_i} - \mathbf{x_j}) \geq 0\\ &\forall[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{q\times n}\\ &\forall\boldsymbol{\alpha} \in \mathbb{R}^n\\ &\forall n \in \mathbb{N}_+ \setminus \{1\}\,.\end{aligned}$$

In one dimension we replace the vectors $\boldsymbol{\tau}, \mathbf{x}_i$ and $\mathbf{x}_j$ with scalars τ, x_i and x_j. Now assume that $\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j\rho_r(\tau) < 0$ for some $[x_1, x_2, \ldots, x_n]$, $\boldsymbol{\alpha}$ and n. Then

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n}\alpha_i r(x_i)\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\mathrm{Cov}\{r(x_i), r(x_j)\}\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\sigma_r^2\mathrm{Corr}\{r(x_i), r(x_j)\}\\ &= \sigma_r^2\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_r(\tau)\\ &< 0\,.\end{aligned}$$

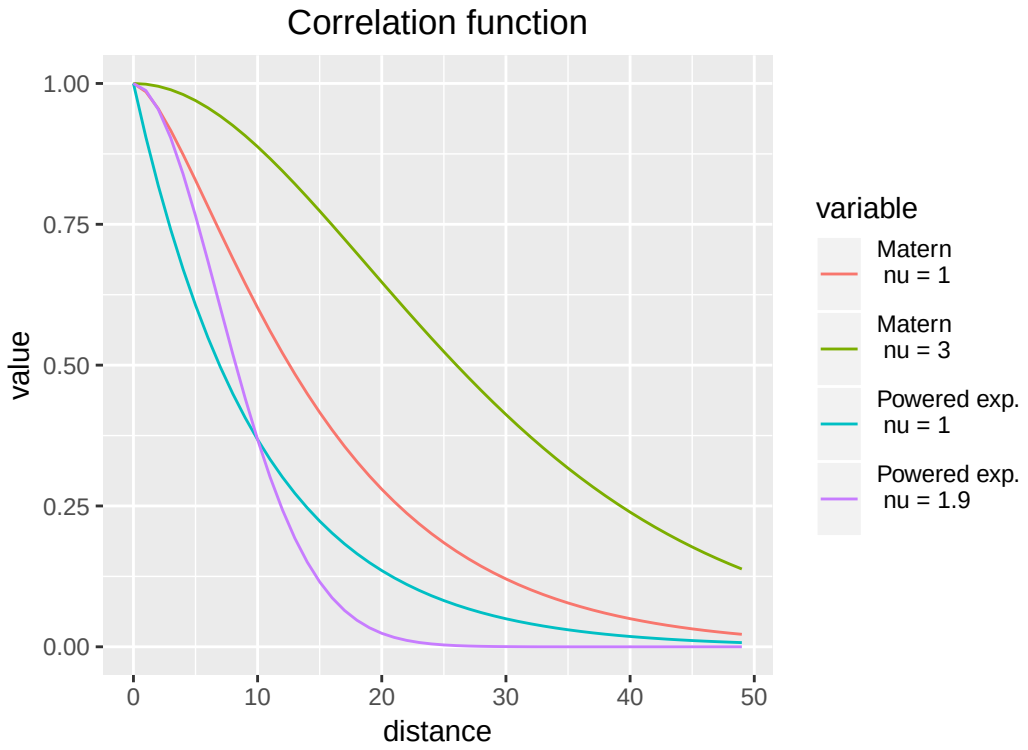

Figure 1: The Powered exponential and the Matern spatial correlation functions for some parameter values.

But the variance of a weighted sum of random variables is non-negative, so this is a contradiction. Thus the correlation function needs to be positive definite.

Two possible correlation functions, $\rho_r(\tau)$, are the Powered exponential and the Matern. The functions take values for τ, and a parameter ν_r. Let $\nu_r \in \{1, 3\}$ for the Matern function and $\nu_r \in \{1, 1.9\}$ for Powered exponential.

The correlation functions for each ν_r are shown in figure 1. We see that the correlation is 1 for $\tau = 0$, which is a criterium for the correlation functions. Also, the correlation goes to 0 as τ increases for all four displayed functions. This means, values far from each other will be independent, and the stationary Gaussian Field will be ergodic.

The variogram function is

$$\begin{aligned}\gamma_r(\tau) &:= \frac{1}{2}\mathrm{Var}\{r(x) - r(x')\} \\ &= \sigma_r^2 - \mathrm{Cov}\{r(x), r(x')\} \\ &= \sigma_r^2[1 - \rho_r(\tau)]\,,\end{aligned}$$

i.e. one minus the covariance scaled by the variance. This is displayed for the given parameter values in figure 2.

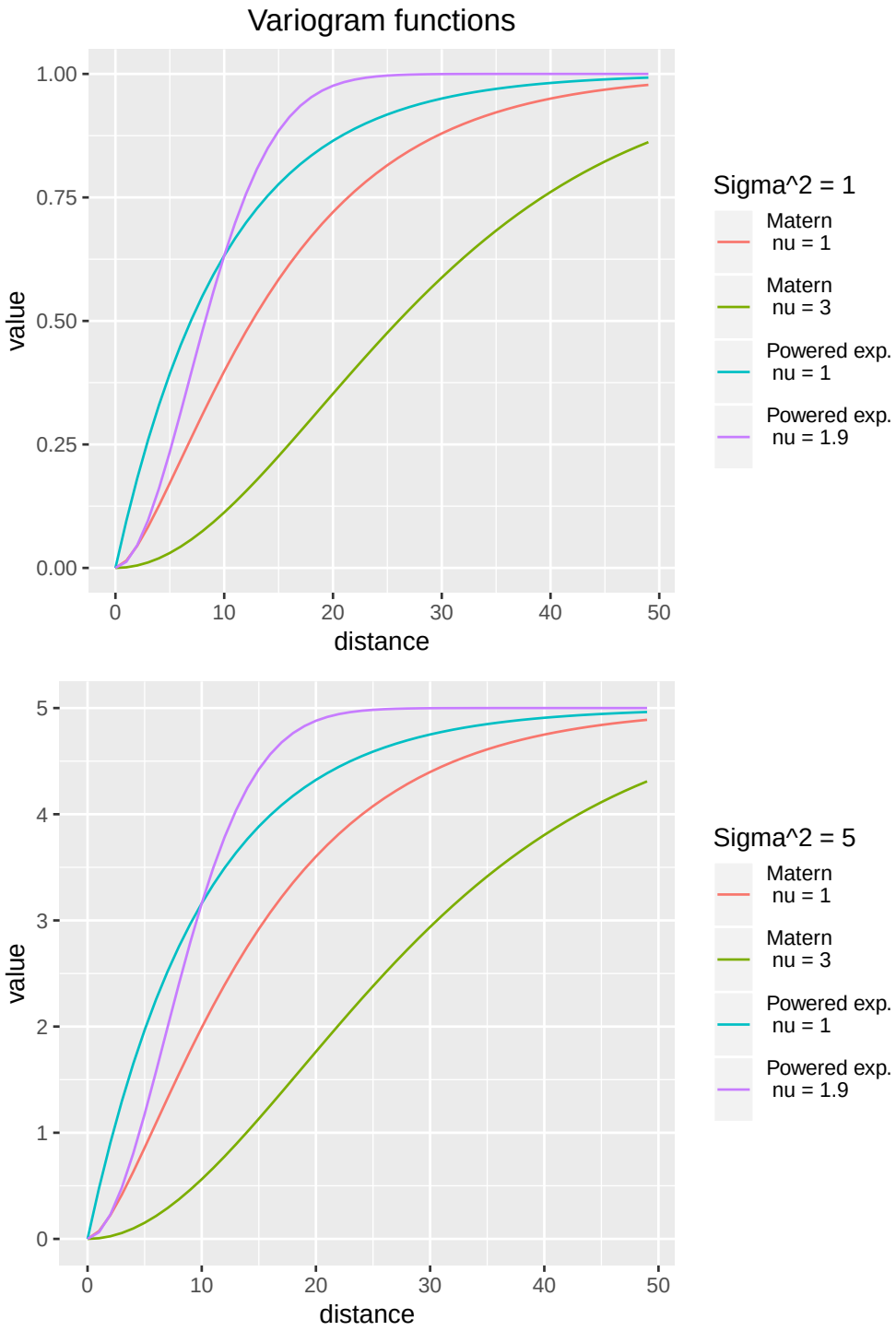

Figure 2: The variogram functions corresponding to Powered exponential and the Matern spatial correlation functions for some parameter values. The variance, $sigma^2$, is purely scaling the functions.

Figure 3: Realizations from prior Gaussian RF using Powered exponential correlation function. Model parametere values are shown the each plot, respectively.

b) The discretized prior Gaussian model has pdf

$$\mathbf{r} \sim p(\mathbf{r}) = \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \mathbf{\Sigma}_r^\rho), \tag{1}$$

where $n = 50$ is the number of grid points in L, $\mathbf{i}_n$ is a n-vector with 1s and $\mathbf{\Sigma}_r^\rho$ is the correlation matrix for the grid points, defined from the chosen spatial correlation function applied to all τs in the system.

Ten realizations for each of the eight different parameter combinations are displayed in figures 3 and 4. The priors in figure 3 are using the Powered exponential function as correlation function, while the ones in figure 4 use the Matern correlation function.

The influence of the parameter values are illustrated in the plots. Large σ^2 means large variability in values, and the realizations vary more along the y-axis in these displays. Increased ν_r increases the correlation between space points, as shown in figure 1. Thus, this results in smoother realizations, where changes are done more gradually.

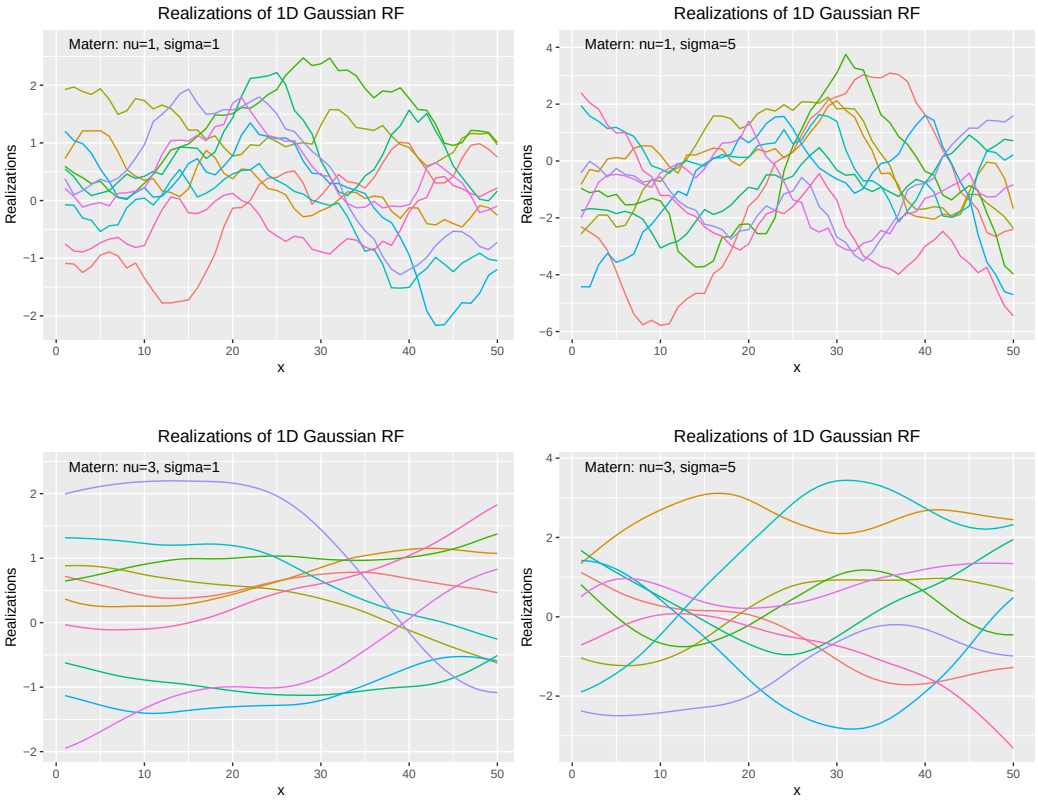

Figure 4: Realizations from prior Gaussian RF using Matern correlation function. Model parametere values are shown the each plot, respectively.

c) Now assume the spatial variable is observed as

$$\{d(x) : x \in \{10, 25, 30\} \subset \mathrm{L}\}$$
$$d(x) = r(x) + \epsilon(x),$$

where the measurement errors, $\epsilon(\cdot)$, are independent and identically distributed.

$$\epsilon(x) \sim \phi(\epsilon; 0, \sigma_\epsilon^2).$$

$r(x)$ and $\epsilon(x')$ are also independent for all x, x'.

The likelihood function for this model is the Gauss-linear likelihood.

$$[\mathbf{d}|\mathbf{r}] = \mathbf{Hr} + \boldsymbol{\epsilon} \sim p(\mathbf{d}|\mathbf{r}) = \phi_m\left(\mathbf{d}; \mathbf{Hr}, \boldsymbol{\Sigma}_{d|r}\right). \tag{2}$$

m is the dimension of $\mathbf{d}|\mathbf{r}$, $\mathbf{H}$ is the observation matrix with dimensions $m \times n$, and $\boldsymbol{\Sigma}_{d|r} = \sigma_\epsilon^2 \mathbf{I}_m$, $\mathbf{I}_m$ being the $m \times m$ identity matrix.

The likelihood $p(\mathbf{d}|\mathbf{r})$ would be a pdf as a function of the data, but in this setting, the data is known and constant and $\mathbf{r}$ is unknown. Thus, the likelihood should be considered a function of $\mathbf{r}$. As a function of $\mathbf{r}$, it does noe necessarily integrate to 1, and hence it is not a pdf. Because of this, there is no need to normalize the likelihood function.

d) To simulate that $\{d(x) : x \in \{10, 25, 30\} \subset \mathrm{L}\}$ is observed, one realization from problem 1b) is saved. The realization is chosen from the prior with parameter combination $\sigma^2 = 5$ and Powered exponential correlation function with $\nu_r = 1.9$. We also let $\sigma_\epsilon^2 \in \{0, 0.25\}$.

The Gaussian RF is a conjugate prior class to the Gauss-linear likelihood. Thus, the posterior distribution is also a Gaussian RF. The Gaussian distribution is completely determined by the mean and the variance. For the discretized version we have

$$[\mathbf{r}|\mathbf{d}] \sim p(\mathbf{r}|\mathbf{d}) = \phi_n\left(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d}\right), \text{ with}$$
$$\boldsymbol{\mu}_{r|d} = \mu_r \mathbf{i}_n + \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T \left[\sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} \left[\mathbf{d} - \mu_r \mathbf{H} \mathbf{i}_n\right]$$
$$\boldsymbol{\Sigma}_{r|d} = \sigma_r^2 \boldsymbol{\Sigma}_r^\rho - \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T \left[\sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} \sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho$$

For prediction, the posterior mean is commonly used. The associated $(1 - \alpha)$ prediction interval is

$$\left[\boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}} - z_{\alpha/2}\sqrt{diag(\boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}})}, \quad \boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}} + z_{\alpha/2}\sqrt{diag(\boldsymbol{\Sigma}_{\mathbf{r}|\mathbf{d}})}\right].$$

Predicted values with corresponding prediction intervals for the two measurement errors are shown in figure 5.

Having measurement errors means that different values can be drawn from the same system in the same positions, and the error affects both the mean and the variance in the posterior distribution. In particular, notice that when the measurement error $\sigma_\epsilon^2 = 0$, the posterior mean in the measured positions matches the measured values in this places, and there is no uncertainty in these positions. The uncertainty increases with the distance from the measured positions.

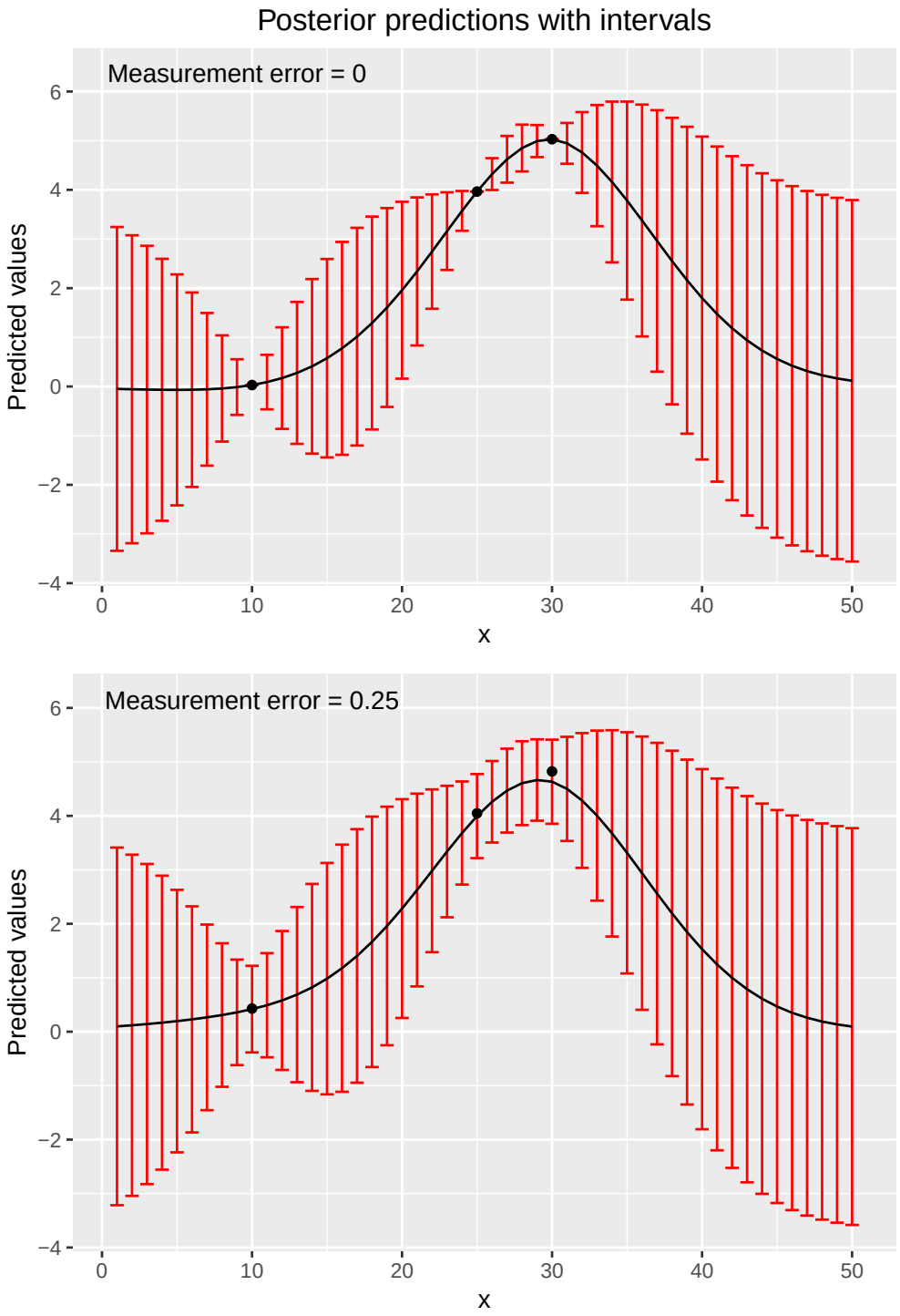

Figure 5: Posterior mean and prediction intervals for two different measurement errors when measured in $x = \{10, 25, 30\}$. The measured values are shown in black dots.

e) The predictions from problem 1d) can be estimated by simulate realizations from the posterior distribution and computing the empirical means, $\hat{\boldsymbol{\mu}}_{r|d}$, and standard deviations ($\hat{\boldsymbol{\sigma}}^2_{r|d}$) from the realizations.
Consider the prediction of one grid point in L, y_i. Then $y_i - \hat{\mu}_{r|d,i}$ is Gaussian distributed with $\mathrm{E}[y_i - \hat{\mu}_{r|d,i}] = 0$ and $\mathrm{Var}(y_i - \hat{\mu}_{r|d,i}) = \sigma^2_{r|d,i}(1 + 1/n)$, with n being the number of simulations. Thus

$$\frac{y_i - \hat{\mu}_{r|d,i}}{\sigma_{r|d,i}\sqrt{1+1/n}} \sim \phi(0,1)$$
$$\frac{y_i - \hat{\mu}_{r|d,i}}{\hat{\sigma}_{r|d,i}\sqrt{1+1/n}} \sim t_{n-1}$$
$$\implies \hat{PI}_\alpha = \left[\hat{\mu}_{r|d,i} - t_{n-1,\alpha/2} \cdot \hat{\sigma}_{r|d,i}\sqrt{1+1/n}, \hat{\mu}_{r|d,i} + t_{n-1,\alpha/2} \cdot \hat{\sigma}_{r|d,i}\sqrt{1+1/n}\right]$$

The realizations, empirical prediction and empirical prediction interval are shown in figure 6. The figure shows how the measurement error affects the posterior distribution and the realizations thereof. When the measurement error is 0, there is not uncertainty in the measured points, and all realizations will have the same value in $x = \{10, 25, 30\}$. There is more variability in the realizations of the model where some measurement was present, as expected. The empirical mean depends on the realizations drawn, but with with increased number of realizations, the empirical mean will converge to the theoretical mean. The empirical prediction intervals are larger than the theoretical one, because of the larger tail in the t-distribution and added uncertainty. This can be seen on the scales of figure 6 versus figure 5.

f) Consider the following function on $\{r(x); x \in \mathrm{D}\}$:

$$A_r = \int_{\mathrm{D}} I(r(x) > 2)dx, \tag{3}$$

with I being the indicator function that is 1 for $r(x) > 2$ and 0 otherwise. Let $A_{r,d} = \sum_{x \in L} I(r(x) > 2)$ be the discrete version of A_r.
We will use the posterior model for $\{r(x); x \in \mathrm{D}\}$ problem 1d)-e) where the $\sigma^2_\epsilon = 0$ and consider two predictors for A_r. One predictor is made from the 100 simulated realizations, $\{r_i(x) : x \in \mathrm{L}, i \in [1, 100]\}$, in problem 1e),

$$\hat{A}_r = \frac{1}{100}\sum_{i=1}^{100}\left(\sum_{x \in L} I(r_i(x) > 2)\right) := \frac{1}{100}\sum_{i=1}^{100} A_{r_i,d},$$

and the other estimates A_r using the predicted spatial variable $\hat{r}(x)$ from 1d):

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

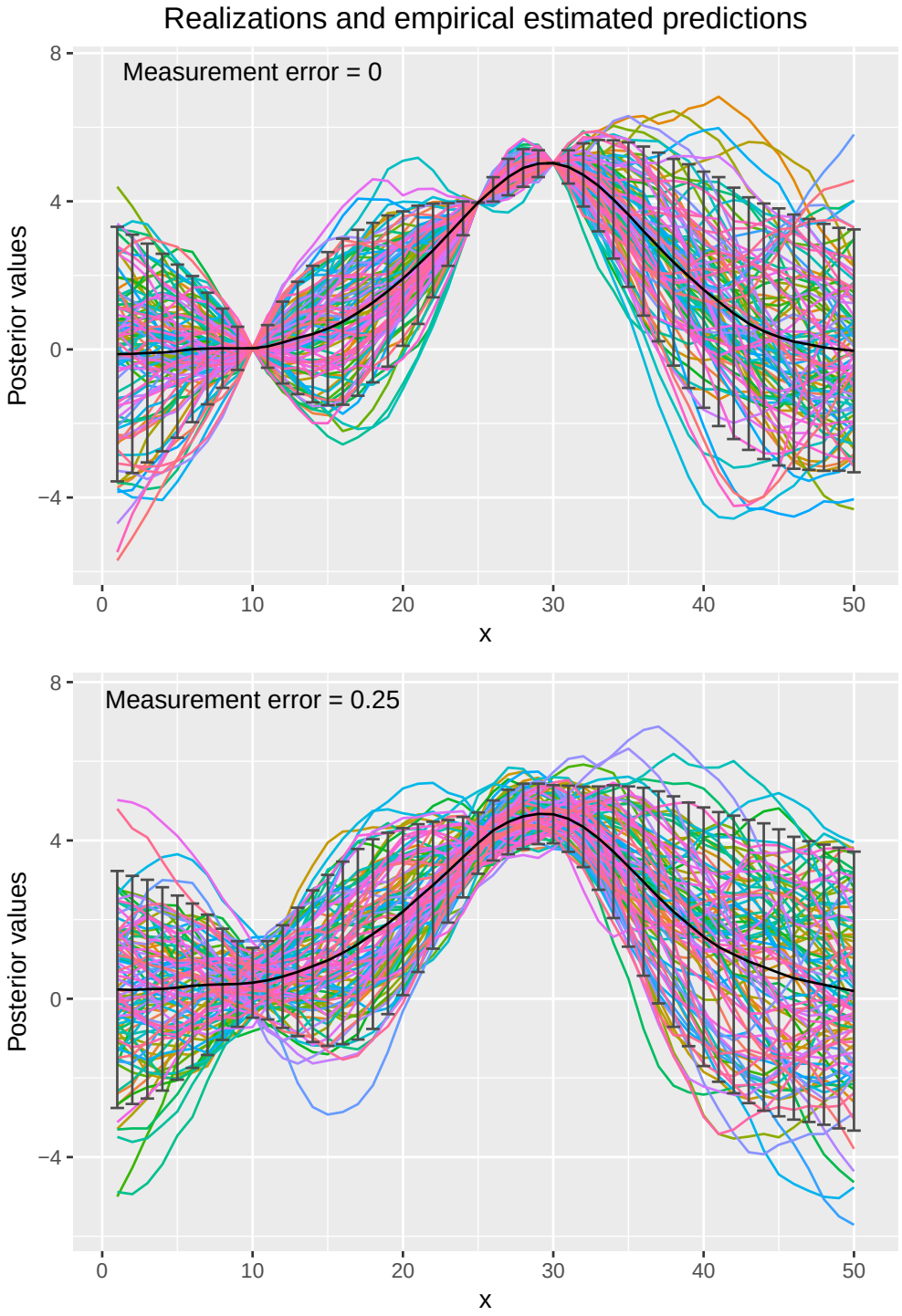

Figure 6: 100 realizations from the posterior distribution for two different measurement errors when measured in $x = \{10, 25, 30\}$. Empirical estimated means and prediction intervals are printed on top of the realizations.

For the former predictor, we would also like to give the associated prediction variance. The variance of the predictor $\hat{A}_r$ is

$$\mathrm{Var}(\hat{A}_r) = \frac{1}{100^2}\sum_{i=1}^{100}\mathrm{Var}(A_{r_i,d}) = \frac{1}{100}\mathrm{Var}(A_{r,d}),$$

since the realizations r_i are independent.

Now let $\hat{a}_r$ be the prediction of $\hat{A}_r$ given the 100 realizations from 1e), i.e a numerical value. Then the empirical variance for $A_{r,d}$ is

$$\hat{\mathrm{Var}}(A_{r,d}) = \frac{1}{100-1}\sum_{i=1}^{100}(A_{r_i,d} - \hat{a}_r)^2,$$

and the estimated variance of the predictor is

$$\hat{\mathrm{Var}}(\hat{A}_r) = \frac{1}{100}\hat{\mathrm{Var}}(A_{r_i,d})$$

Inserting numerical values for the realizations gives the following predictions and the prediction variance:

$$\begin{aligned}\hat{a}_r &= 21.83\\ \tilde{a}_r &= 19\\ \hat{\mathrm{Var}}(\hat{A}_r) &= 0.3535\end{aligned}$$

We observe that the $\hat{a}_r > \tilde{a}_r$. Is it to be expected for the predictors in general that $\hat{A}_r > \tilde{A}_r$?

Jensen's inequality states that for convex functions $g(Y)$, the following inequality holds:

$$\mathrm{E}[g(Y)] \geq g(E[Y]).$$

As $\hat{r}(x) = \mathrm{E}[r(x)|d]$, $\quad \tilde{A}_r = g(\mathrm{E}[r(x)|d])$ with $g(Y) = \sum_{x\in L} I(Y(x) > 2)$. Also, $\hat{A}_r = \mathrm{mean}_i(g(r_i(x)))$, which is a common estimator for $\mathrm{E}[g(r(x))]$. Thus, if $\sum_{x\in L} I(Y(x) > 2)$ is a convex function, one would expect $\hat{A}_r > \tilde{A}_r$.

g) In this problem, we have seen how parameter choices in the prior and the accuracy of measurements for the likelihood affect the posterior distribution we end up with. The *nu*-parameter in the correlation function determines how smooth the prior distribution is, which also affects the characteristics of the posterior distribution.

The measurement error tells how much we can trust the measured values, which is reflected in the posterior variance in these points and thus how much realizations of the posterior distribution are allowed to differ from the measured values.

We have seen that analytically computed predictions and simulated approximation gives similar result, but the analytical prediction will always be more

accurate and has smaller uncertainty. In addition, the prediction for the posterior distribution is the posterior mean, and we use this when we sample from the posterior. Thus we do the same computations in both cases, and there is no reason to do the numerical approximation.

However, when analytical computations become non-trivial, it is good to be able to do numerical approximations instead.

Problem 2

a) We have 52 observations contained in the domain $\mathcal{D} = (0, 315) \times (0, 315)$. In Figure 7 we see a interpolation of the data points in *"topo.dat"*. A stationary Gaussian RF model would not be a suitable for the elevation, since the mean is changing with the reference point $\mathbf{x}_0$. We let the field $\mathcal{D}$ be modelled by a Gaussian RF.

b) Consider the continuous spatial variables $\{r(\mathbf{x}); \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^2\}$ s.t.

$$\begin{aligned} E[r(\mathbf{x})] &= \mathbf{g}(\mathbf{x})^T \boldsymbol{\beta}_r \\ \text{Var}\{r(\mathbf{x})\} &= \sigma_r^2 \\ \text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} &= \rho_r(\boldsymbol{\tau}/\xi) \end{aligned} \tag{4}$$

with $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), ...g_{n_g}(\mathbf{x}))$ a vector of known spatial variables and $\beta_r = (\beta_1, ..., \beta_{n_g})$ an unknown parameter. $\sigma_r^2 = 2500$ and the spatial correlation function have $\xi = 100$ with $\tau = |\mathbf{x} - \mathbf{x}'|$. The prediction is

$$\hat{r}_0 = \boldsymbol{\alpha}^T \mathbf{r}_{n_g}, \tag{5}$$

with $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{n_g})$ being the weights. The predictor error needs to be unbiased, which mean that the expected value of the error must be zero, giving

$$\begin{aligned} \mathbf{E}[\hat{r}_0 - r_0] &= \boldsymbol{\alpha}^T \mathbf{E}[r^g] \\ &\Downarrow \\ \boldsymbol{\alpha}^T \mathbf{G}_g \beta_r^+ &= \mathbf{g}(\mathbf{x}_0) \\ &\Downarrow \\ \mathbf{G}_g^T \boldsymbol{\alpha} &= \mathbf{g}(\mathbf{x}_0) \end{aligned}$$

with the $(n_g \times n_g)$-matrix $\mathbf{G}_g = (\mathbf{g}(\mathbf{x}_1, ..., \mathbf{g}(\mathbf{x}_{n_g})^T$. The optimization problem comes from minimizing the prediction error variance. It is derived from

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \arg\min_{\alpha} \text{Var}\{\hat{r}_0 - r_0\} \\ &= \arg\min_{\alpha} \{\sigma_r^2 - 2\boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\rho}_0 + \boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\Sigma}_g^\rho \boldsymbol{\alpha}\} \\ \text{constrained by} \quad & \mathbf{G}_g^T \boldsymbol{\alpha} = \mathbf{g}(\mathbf{x}_0) \end{aligned}$$

The analytical solution using optimization with Lagrange multipliers is

$$\hat{\boldsymbol{\alpha}} = [\boldsymbol{\Sigma}_g^\rho]^{-1} \left[\boldsymbol{\rho}_0 - \mathbf{G}_g^T [\mathbf{G}_g [\boldsymbol{\Sigma}_g^\rho]^{-1} \mathbf{G}_g^T]^{-1} [\mathbf{G}_g [\boldsymbol{\Sigma}_g^\rho]^{-1} \boldsymbol{\rho}_0 - \mathbf{g}(\mathbf{x}_0)]\right],$$

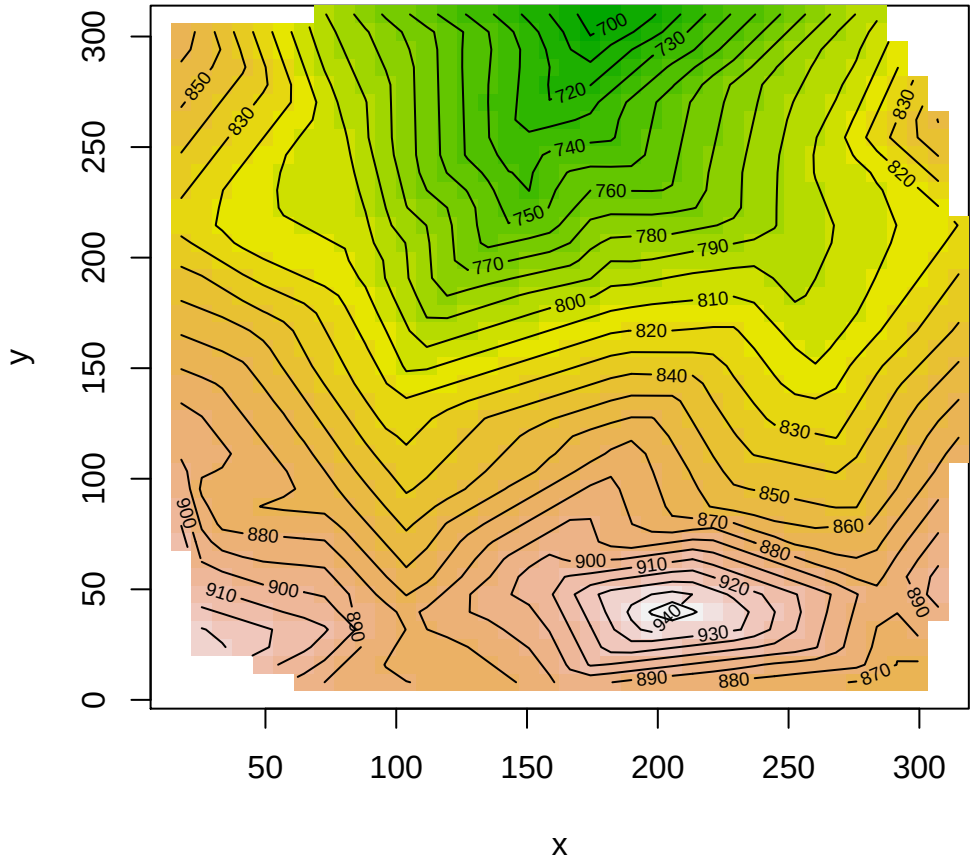

Figure 7: Contour plot with a heat-map of the interpolated observed data in $\mathcal{D}$

which yields the universal kriging predictor and the associated prediction variance at an arbitrary location $\mathbf{x}_0 \in \mathcal{D}$.

$$\begin{aligned} \hat{r}_0 &= \hat{\boldsymbol{\alpha}}^T \mathbf{r}_{n_g} \\ \sigma^2_{\hat{r}} &= \sigma^2_r \left[1 - 2\hat{\boldsymbol{\alpha}}^T \boldsymbol{\rho}_0 + \hat{\boldsymbol{\alpha}}^T \boldsymbol{\Sigma}^\rho_g \hat{\boldsymbol{\alpha}}\right] \end{aligned} \tag{6}$$

c) Let the reference variable $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^2$ be $\mathbf{x} = (x_v, x_h)$ with $n_g = 6$, and the expectation function $\mathbf{g}(\mathbf{x}) = x_v^k x_h^l$ for $(k,l) \in \{(0,0),(1,0),(0,1),(1,1),(2,0),(0,2)\}$. Which means that the n_g-vector $\mathbf{g}(\mathbf{x}) = (1, x_v, x_h, x_v x_h, x_v^2, x_h^2)^T$. Looking at the expected value for $r(\mathbf{x})$ we get

$$\mathbf{E}[r(\mathbf{x})] = \mathbf{g}(\mathbf{x})^T \boldsymbol{\beta}_r = \beta_r^1 + x_v \beta_r^2 + x_h \beta_r^3 + x_v x_h \beta_r^4 + x_v^2 \beta_r^5 + x_h^2 \beta_r^6.$$

The values of $\boldsymbol{\beta}_r$ is unknown, but we don't need them to find the prediction. This is because of the unbiasedness of the expectation of the error in the prediction.

The predicted value does not include the variance, and the variance does not include the actually values of the observations. In Figure 8 we see a heat map with contours of the predicted values $\hat{r}_{\mathbf{x}_0}$ and it resembles what we got in the interpolation made in Figure 7. We also see the variance of the prediction, where the values are low close to the observed data points and higher far away from them. This is what we should expect from the formula of predicted variance. Now we consider the model with a 1st degree expectation function. Which means we remove the 2nd order terms from $\mathbf{g}(\mathbf{x})$. The comparison of the result is shown in Figure 9. Here one can see a clear difference in predicted height between the models with 1st and 2nd degree expectation function. In the predicted variance you can't see much difference close to the data point, the further away the reference $\mathbf{x}_0$, the more difference one can observe. This also correlates with the deviations seen in the predicted value. The reason for this is that the 2nd degree terms of $\boldsymbol{\beta}_r$ is small and the difference in the variance of the 1st and 2nd degree model, is only visual when the variances gets large. Changing the variance σ_r^2 will off course change the variance in the model, but it doesn't change the predicted value.

d) The probability for the height to be higher than 700 at the point $\mathbf{x}_0 = (100, 100)$ is a normal distribution. The calculation yields

$$P\left(Z > \frac{700 - \hat{r}_{\mathbf{x}_0}}{\sigma_{\hat{r}_{\mathbf{x}_0}}}\right) = 1 - P\left(Z \leq \frac{700 - \hat{r}_0}{\sigma_{\hat{r}_{\mathbf{x}_0}}}\right) = 1 - P\left(Z \leq \frac{700 - 838.2}{\sqrt{595.5}}\right) = 1$$

So the probability for the height to be higher than 700 in $\mathbf{x}_0$ is 100%. The height at which the probability is 0.90 of the true height being lower than the predicted height at $\mathbf{x}_0$ is

$$Z|_{P=0.9} = \frac{alt - \hat{r}_{\mathbf{x}_0}}{\sigma_{\hat{r}_{\mathbf{x}_0}}} = 1.28 \Rightarrow alt = 1.28 \cdot \sqrt{595.5} + 838.2 = 869.4.$$

e) We want to look at the predicted value and variance when we add a observation error to the observation. The result from this is displayed in Figure 10. Where you can see that this error doesn't change the predicted value, but in the variance you can see a clear difference. The predicted variance gets larger as the observation error gets larger, which is a natural response.

f) Using the kriging predictor formulated above, we achieved close to the same field of predicted values as with the interpolation. The β-values corresponding to the 2nd degree values of the expectation function was really small, so this resulted in a small difference in the prediction using a first degree expectation function. In the latter part of this exercise we introduced a observation error. This didn't affect the predicted value at all, but it resulted in a bigger predicted variance.

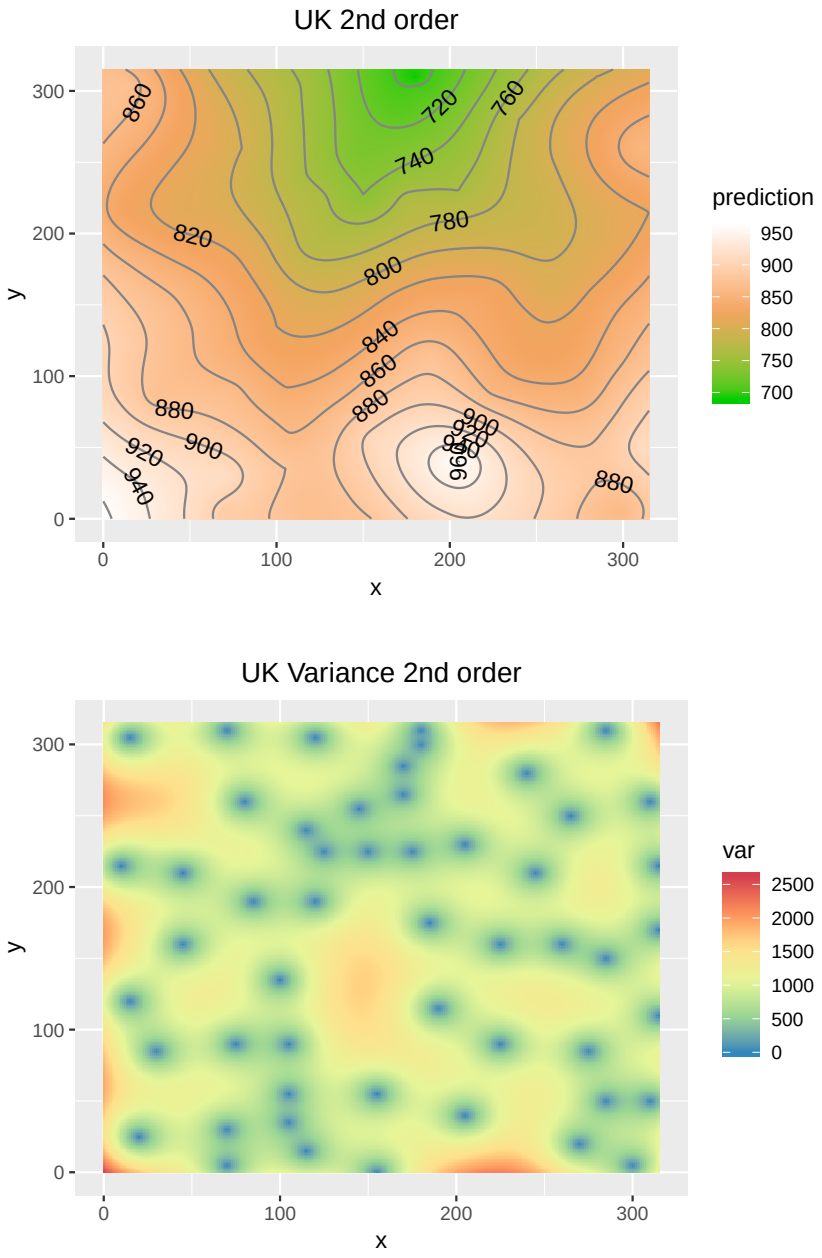

Figure 8: Top: Heat-map with contours of the predicted value $\hat{r}_{\mathbf{x}_0}$ on the grid $(0, 315) \times (0, 315)$ with second order expectation function. Bottom: Heat-map of the predicted variance $\sigma_{\hat{r}_{\mathbf{x}_0}}$ on a grid $(0, 315) \times (0, 315)$ with second order expectation function.

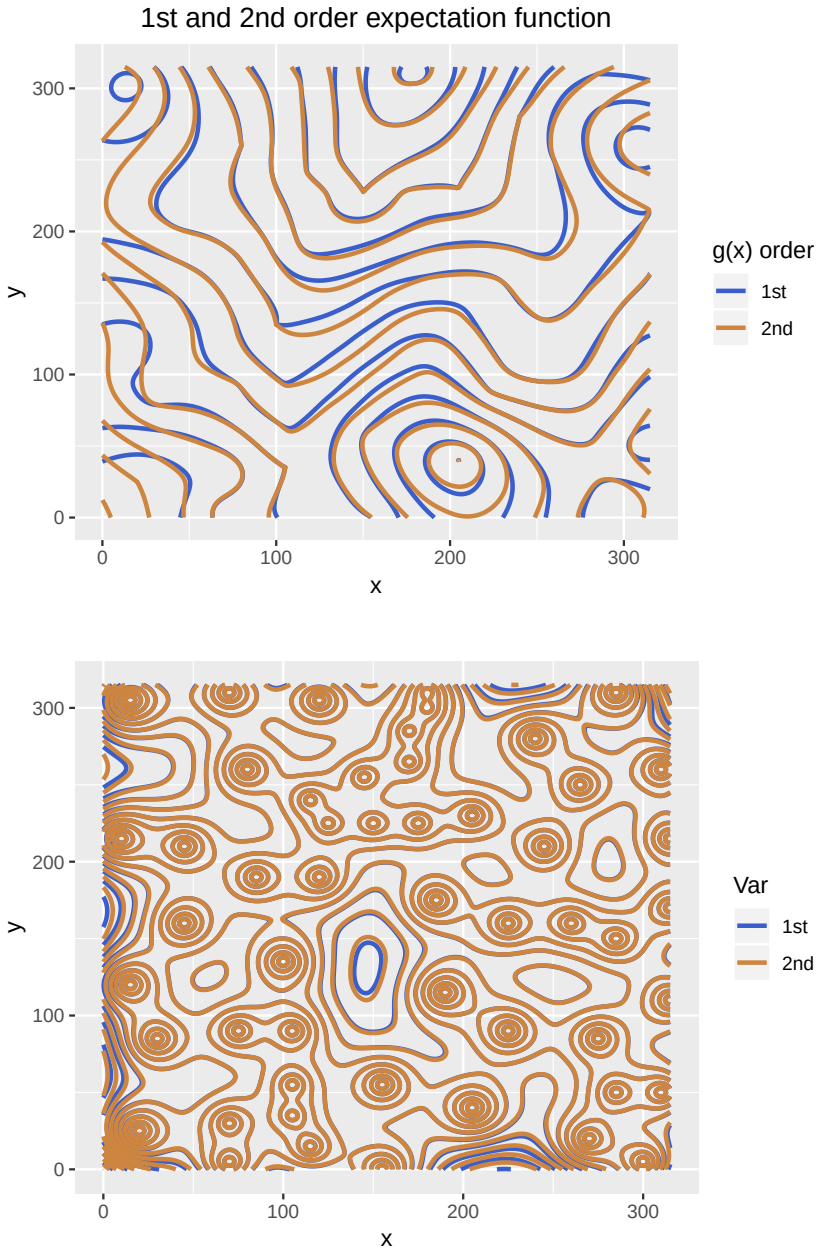

Figure 9: Top: A contour comparison of the predicted value with 1st and 2nd order expectation function $\mathbf{g}(\mathbf{x})$. Bottom: A contour comparison of the variance of the predicted variance with 1st and 2nd order expectation function.

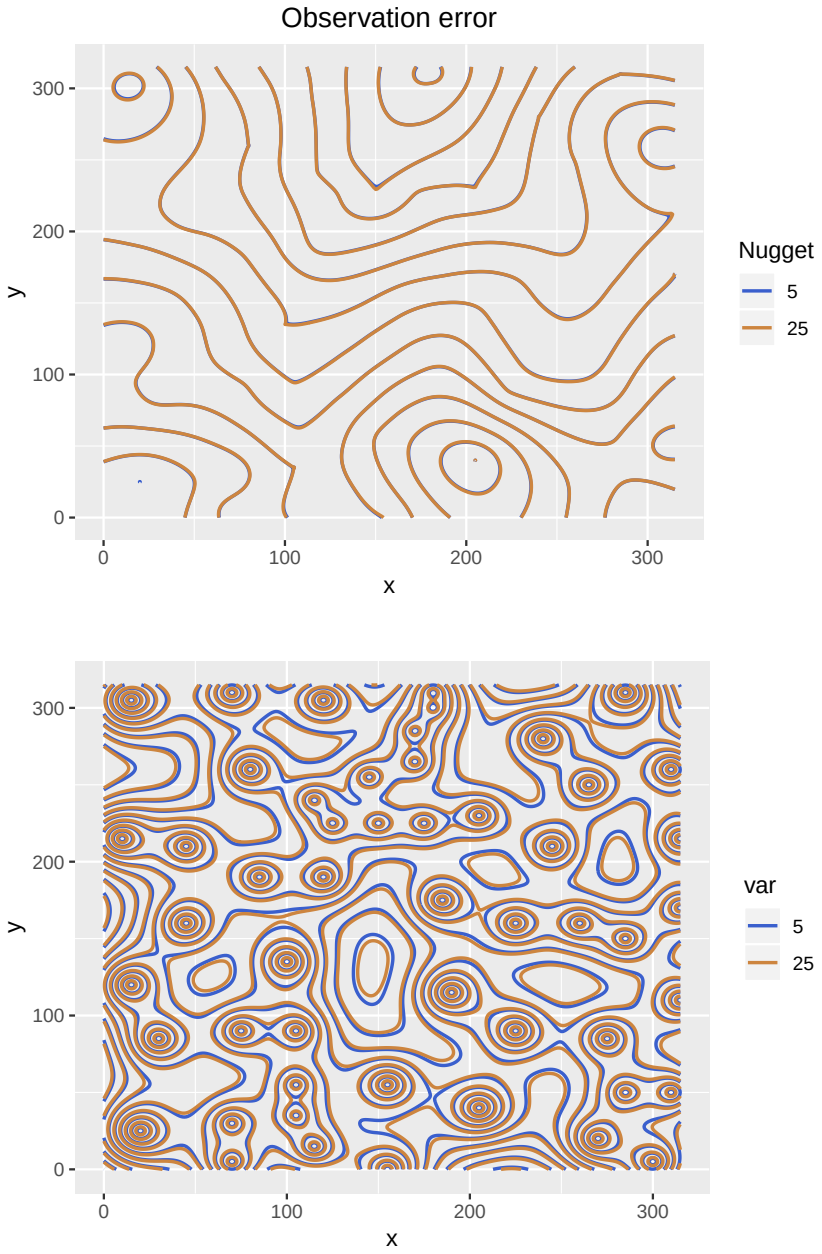

Figure 10: Top: The predicted value of the observation with observation error. Bottom: predicted variance of the observation with observation error. The blue contours is $\sigma_\epsilon^2 = 5$ and the brown contours is $\sigma_\epsilon^2 = 25$

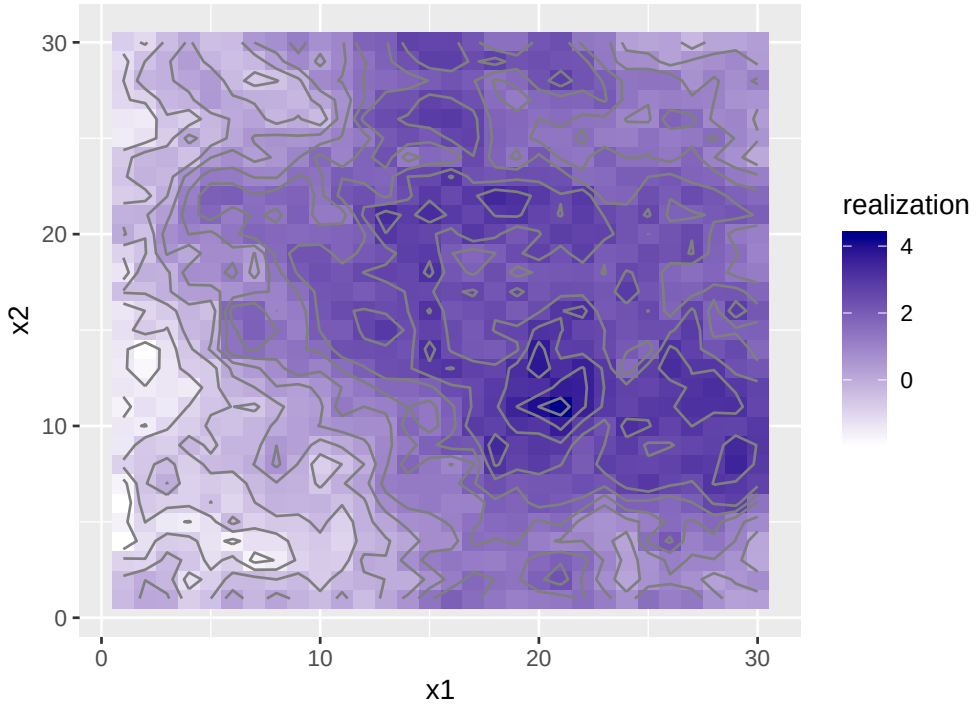

Figure 11: One realization of the discretized Gaussian RF given in Problem 3a.

Problem 3

We consider the stationary Gaussian RF $\{r(\mathbf{x})\,;\,\mathbf{x}\in \mathrm{D}\subset\mathbb{R}^2\}$ with $\mathrm{D}\colon [(1,30),(1,30)]$,

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

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

a) We consider the discretized Gaussian RF $\{r(\mathbf{x})\,;\,\mathbf{x}\in \mathrm{L}\}$ on the grid $\mathrm{L}\colon [30\times 30]\in \mathrm{D}$ with model parameters $\sigma_r^2 = 2$ and $\xi_r = 15$. In figure 11 we see one realization of the discretized Gaussian RF.

b) The variogram function for the Gaussian RF is

$$\begin{aligned}\gamma_r(\tau) &\coloneqq \frac{1}{2}\mathrm{Var}\{r(x)-r(x')\}\\ &= \sigma_r^2 - \mathrm{Cov}\{r(x), r(x')\}\\ &= \sigma_r^2[1-\exp\{-\tau/\xi_r\}]\,.\end{aligned}$$

This is shown in figure 12 together with the empirical variogram based on exact observations of the full realization in a). The empirical variogram is precise for

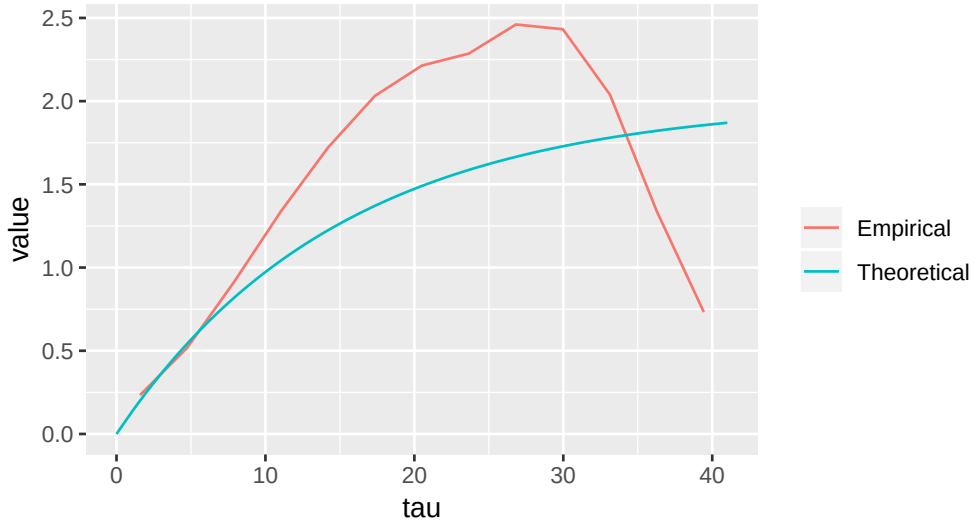

Figure 12: The theoretical variogram and empirical variogram from one realization of the discretized Gaussian RF given in Problem 3a.

small distances, but differs from the theoretical variogram for larger distances. This is not surprising, since there is less data for higher distances.

c) We generate 36 locations uniformly randomly on the grid L and compute the empirical variogram based on the realization values in these locations. See figure 13It is less smooth than the empirical variogram computed from the entire realization, but still quite close to the theoretical variogram for small distances.

Now we consider σ_r^2 and ξ_r to be unknown and estimate their values by maximum likelihood estimation. We do this for the full realization and the 36 exact observations. This gives $\sigma_r^2 \approx 2.220$ and $\xi_r \approx 19.950$ using the full realization, and $\sigma_r^2 \approx 2.374$ and $\xi_r \approx 20.558$ using the observations.

From these values we estimate the variogram function. They are displayed in figure 14. The estimates are way more accurate and smooth than the empirical variograms. We can get a similarly smooth estimate by simulating several times and using the mean of the empirical variograms. This, however, takes a lot more time, and it is not possible if one doesn't know the true distribution and thus can't simulate repeatedly. For realizations of a higher resolution the empirical variogram would likely fare better.

d) Figure 15 shows empirical variograms computed from 9, 64 and 100 exact observations from the same realization. The empirical variogram from 9 observations seems very arbitrary, while there is not a huge difference between the ones from 64 and 100 observations.

Figure 16 shows estimated variograms using estimates ML estimates of σ_r^2 and ξ_r from 9, 64 and 100 exact observations from the same realization. Again, it is

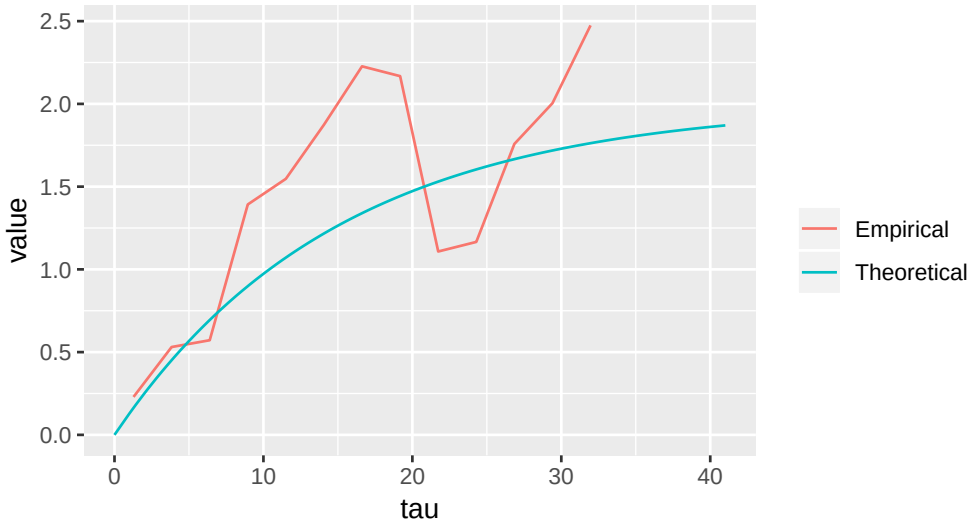

Figure 13: The theoretical variogram of the discretized Gaussian RF given in Problem 3a and the empirical variogram based on 36 exact randomly chosen observations of a realization.

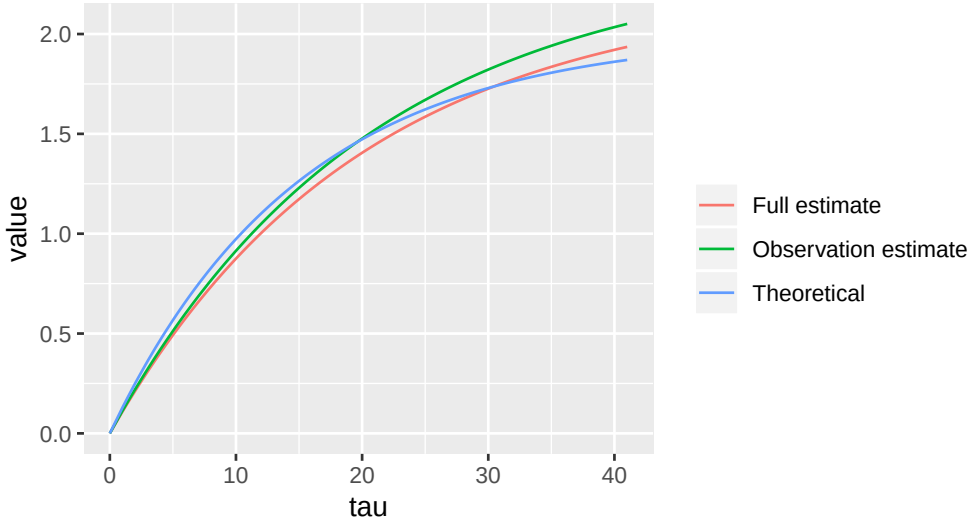

Figure 14: The theoretical variogram of the discretized Gaussian RF given in Problem 3a and estimated variograms based on estimates of σ_r^2 and ξ_r using the whole realization and using 36 exact observations.

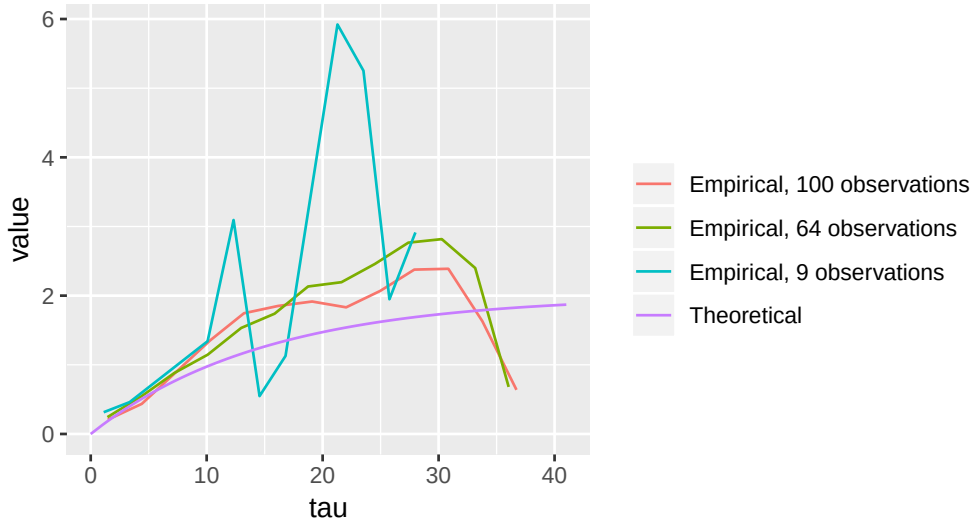

Figure 15: The theoretical variogram of the discretized Gaussian RF given in Problem 3a and the empirical variogram based on 9, 64 and 100 exact randomly chosen observations of a realization.

not easy to distinguish the ones from 64 and 100 observations in terms of quality, while the estimate from 9 observations is quite off.

e) In this problem we have seen that we can estimate parameters of a Gaussian RF with an exponential correlation function quite accurately from only few realizations. We have also seen that it may be beneficiary to use estimated parameters to estimate variograms rather than use empirical variograms, and so probably also for correlations. That is, if one knows the theoretical variogram (or correlation function) in the first place.

An important thing to note is that for random fields where the correlation depends only on the distance between points, the estimates of the correlation are more accurate for short distances. This might not be too problematic though, since the correlation usually decreases with distance.

Furthermore, we have seen that when increasing the number of observations to get better estimates, the factor of the number of observations before and after the increase is more important than how many observations one adds. So it is, not very surprisingly, easier to increase the accuracy of estimates by gathering more data when we only have a few observations.

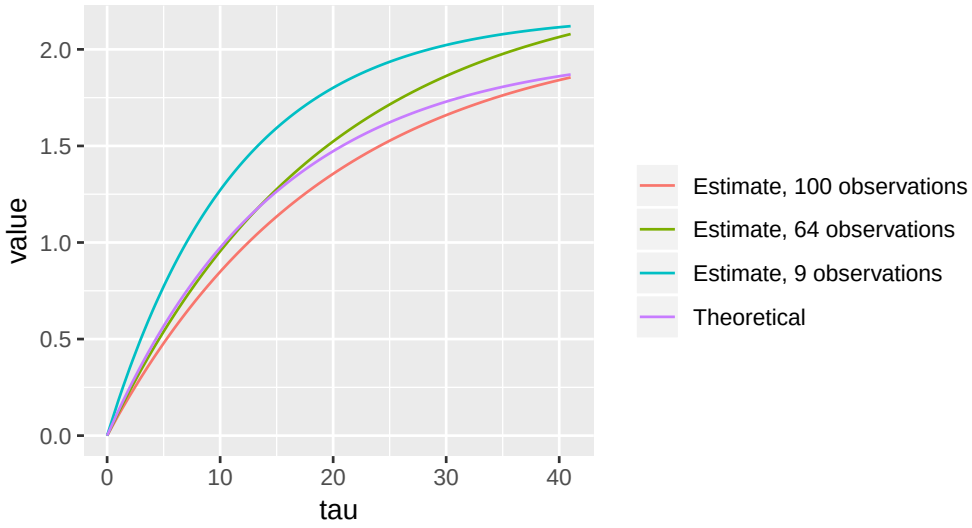

Figure 16: The theoretical variogram of the discretized Gaussian RF given in Problem 3a and estimated variograms based on estimates of σ_r^2 and ξ_r using the whole realization and using 9, 64 and 100 exact observations.