# Spatial Statistics Assignment 1

Filip Schjerven and Torfinn Tyvold

January-February 2017

## Problem 1

**a)** RealiZed values from the correlation functions have been plotted in figure 1. The discontinuity of some of the lines at distance = 0 is due to starting our calculation at the minimum distance of d = 1.

The range of a GRF tells us of how intense the correlation between datapoints will be. The higher the range, the higher the correlation between distant points will be and vice versa.

The variogram function is defined as $2\gamma(\mathbf{s},\mathbf{t}) = Var(Y(\mathbf{s}) - Y(\mathbf{t}))$ for two arbitrary positions $\mathbf{s},\mathbf{t}$ when having a statinary process. For this particular problem, we get the following relation between the variogram and the correlation function:

$$\begin{aligned} 2\gamma(\mathbf{s},\mathbf{t}) &= Var(Y(\mathbf{s}) - Y(\mathbf{t})) \\ &= Var(Y(\mathbf{s})) + Var(Y(\mathbf{t})) \\ &\quad - 2Cov(Y(\mathbf{s}), Y(\mathbf{t})) \\ &= 2\sigma_1^2 - 2\sigma_1^2 Corr(\|\mathbf{s}-\mathbf{t}\|) \\ \implies \gamma(\mathbf{s},\mathbf{t}) &= \sigma_1^2(1 - Corr(\|\mathbf{s}-\mathbf{t}\|)) \end{aligned}$$

Where we in line two use that $\frac{Cov(Y(\mathbf{s}),Y(\mathbf{t}))}{\sigma_1^2} = Corr(\|\mathbf{s}-\mathbf{t}\|)$. We thus see that the two are inversely related, as the variogram function will decrease as the correlation between two points increase.

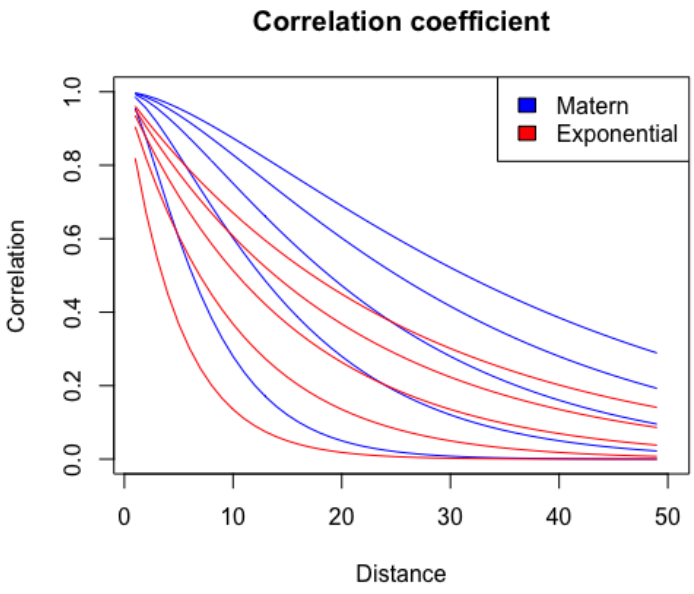



Figure 1: Correlation coefficients for the ranges 5, 10, 15, 20, 25, top to bottom.

**b)**

**c)** The data model for the temperature is $\mathbf{Z}(s_i)|Y(s_i),\sigma_\epsilon^2 \sim \mathcal{N}(Y(s_i),\sigma_e^2)$ and the process model is $Y(s_i) \sim \mathcal{N}(20, C_Y(h))$.

**d)** Having assumed that $\sigma_\epsilon^2 = 1$, we produced a realiZation of the GRF. $s^* = \{20.7, 22.2, 20\}$ were produced as a set of observations for the remainder of problem 1. This set was produced with the exponential covariance function with range 50 and the resulting conditional results is provided in plot 2.

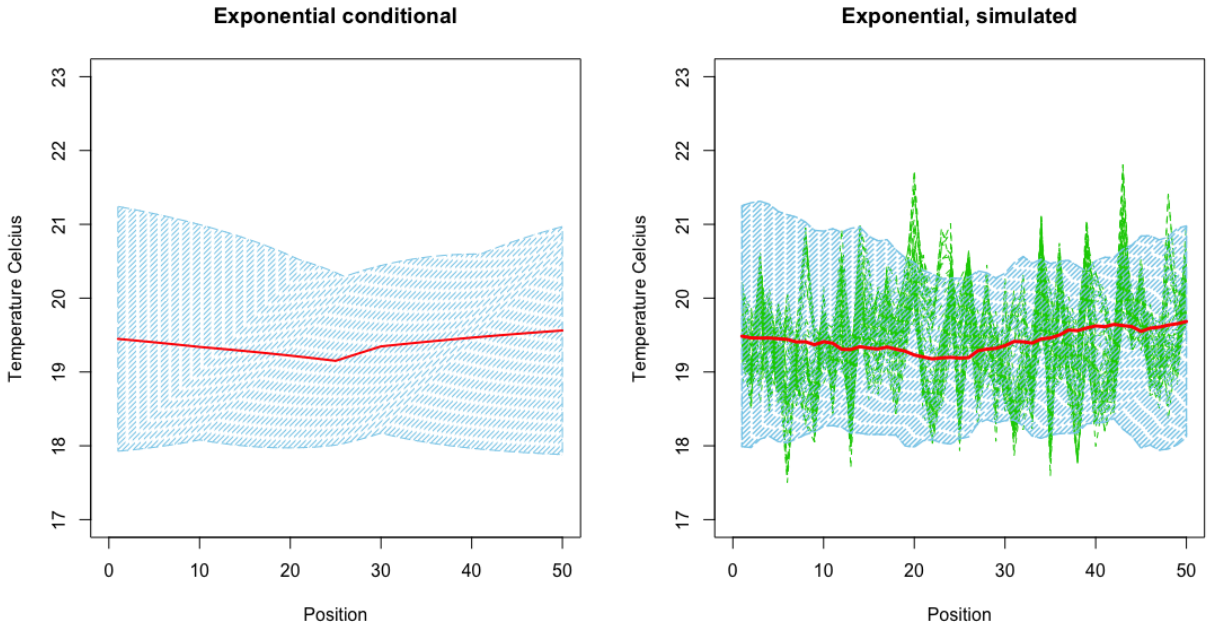



Figure 2: **Left:** The simple kriging results produced for $[Y|\mathbf{Z} = s^*]$. **Right:** 50 simulations from for $[Y|\mathbf{Z} = s^*]$ plotted over its simple kriging predictor.

**e)** 50 simulations were produced from the distribution of $[Y|\mathbf{Z} = s^*]$ produced by the simple kriging method and are displayed in 2. At first glance, we see that much of the variability of the simulated results are captured by the simple kriging results. However, we notice that many of the realiZed values take values that are outside our $2\sigma$ confidence region. By knowledge of the Gaussian pdf and assuming independent errors, this confidence region should contain approximately 96 percent of the simulated data. This is obviously not the case as many realiZations of $[Y|\mathbf{Z} = s^*]$ have values exceeding this region. However, we know that the data are correlated to a lower-bound of exp(-1) as we have chosen range as 50 and 49 being the maximum distance. The confidence region is also produced by estimating the mean and variances of the simulated data, not accounting for correlation at all, which may affect the accuracy of the variance.

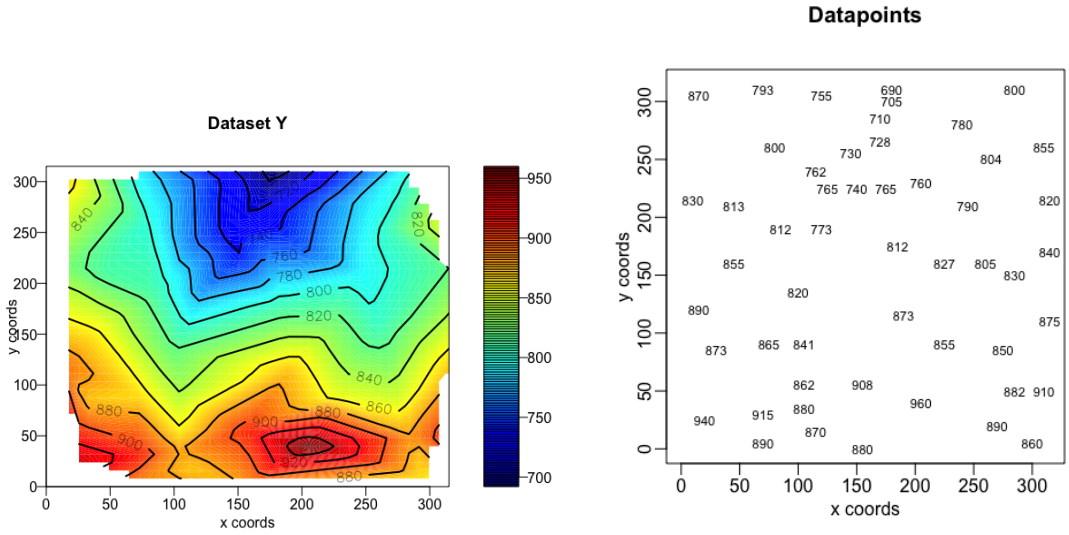
# Problem 2

**a)** An interpolated representation of the data is provided in 3a. The position of the datapoints that were sampled are portrayed in 3b.

**b)** To derive the universal kriging predictor, we'll start of by looking at the homogenously linear combination of data $\mathbf{Z}$ that minimizes the mean squared prediction error $MSPE(\lambda) \equiv E\big(\mathbf{Y}(\mathbf{s_0}) - \lambda^T\mathbf{Z}(\mathbf{s})\big)^2$, subject to the unbiasedness condition $E\big(\lambda^T\mathbf{Z}(\mathbf{s})\big) = E(\mathbf{Y}(\mathbf{s_0})) = \mathbf{x}(\mathbf{s_0})\beta$ for all $\mathbf{s_0} \subset R^2$ and our dataset $\mathbf{s}$. Rewriting the $MSPE$ we get

$$MSPE(\lambda) \equiv E\big(\mathbf{Y}(\mathbf{s_0}) - \lambda^T\mathbf{Z}\big)^2 = Var\big(\mathbf{Y}(\mathbf{s_0}) - \lambda^T\mathbf{Z}(\mathbf{s})\big) + \Big(E\big(\mathbf{Y}(\mathbf{s_0}) - \lambda^T\mathbf{Z}(\mathbf{s})\big)\Big)^2 \tag{1}$$

$$= Var\big(\mathbf{Y}(\mathbf{s_0})\big) + \lambda^T Var\big(\mathbf{Z}(\mathbf{s})\big)\lambda - 2Cov\big(\mathbf{Y}(\mathbf{s_0}), \mathbf{Z}(\mathbf{s})\big) + \Big(E\big(\mathbf{Y}(\mathbf{s_0}) - \lambda^T\mathbf{Z}(\mathbf{s})\big)\Big)^2 \tag{2}$$




(a) Interpolated dataset Y.

(b) Values of data at their respective positions.

We need to fulfill the unbiasedness criteria, which is equivalent to $\mathbf{x}(\mathbf{s_0})\beta = \lambda^T\mathbf{x}(\mathbf{s})\beta$. Incorporating this criteria with Lagrange multiplier $\kappa$, we minimize the function

$$\mathbf{C_Y}(\mathbf{s_0},\mathbf{s_0}) + \lambda^T\mathbf{C_Z}\lambda - 2\big(\mathbf{Y}(\mathbf{s_0}),\mathbf{Z}(\mathbf{s})\big) - 2\kappa\big(\lambda^T\mathbf{x}(\mathbf{s})\beta - \mathbf{x}(\mathbf{s_0})\beta\big)$$

Were we've denoted $Var(\mathbf{Y}(\mathbf{s_0})) = \mathbf{C_Y}(\mathbf{s_0},\mathbf{s_0})$, $Var(\mathbf{Z}(\mathbf{s})) = \mathbf{C_Z}$ and $Cov\big(\mathbf{Y}(\mathbf{s_0}),\mathbf{Z}(\mathbf{s})\big) = \mathbf{c_Y}(\mathbf{s_0})$. Taking derivatives with respect to $\lambda^T$ and $\kappa$ and setting equal to zero we obtain

$$-\mathbf{c_Y}(\mathbf{s_0}) + \mathbf{C_Z}\lambda = \kappa \tag{3}$$

$$\lambda^T\mathbf{x}(\mathbf{s})\beta = \mathbf{x}(\mathbf{s_0})\beta \tag{4}$$

This optimizing problem is to be solved and combined with an Generalized Least Squares (GLS) estimate of parameters $\beta$ to achieve the universal kriging predictor.

At page **REF TIL BOKEN HER**, we're given the predictor as

$$\widehat{\mathbf{Y}}(\mathbf{s_0}) = \mathbf{x}(\mathbf{s_0})^{\mathbf{T}}\widehat{\beta}_{\mathbf{GLS}} + \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big(\mathbf{Z}(\mathbf{s}) - \mathbf{X}^T\widehat{\beta}_{\mathbf{GLS}}\big)$$

From results known from regression, $\widehat{\beta}_{\mathbf{GLS}}$ is distributed Gaussian, i.e. $\widehat{\beta}_{\mathbf{GLS}} \sim \mathbf{N}(\beta, (\mathbf{X^T C_Z X})^{-1})$. The predictor variance can be found by

$$\begin{aligned}
&\mathrm{Var}\big(\widehat{Y}(s_0)\big) = \mathrm{Var}\big(\mathbf{x}(\mathbf{s_0})^T\widehat{\beta}_{\mathbf{GLS}} + \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big(\mathbf{Z}(\mathbf{s}) - \mathbf{X}^T\widehat{\beta}_{\mathbf{GLS}}\big)\\
&= \mathrm{Var}\Big(\big(\mathrm{x}(\mathrm{s}_0)^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)\widehat{\beta}_{\mathbf{GLS}}\Big) + \mathrm{Var}\big(\mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{Z}(\mathbf{s})\big)\\
&= \big(\mathrm{x}(\mathrm{s}_0)^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)\mathrm{Var}\big(\widehat{\beta}_{\mathbf{GLS}}\big)\big(\mathbf{x}(\mathbf{s_0})^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)^T\\
&+ \big(\mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big)\mathrm{Var}\big(\mathbf{Z}(\mathbf{s})\big)\big(\mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big)^T\\
&= \big(\mathrm{x}(\mathrm{s}_0)^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)\big(\mathbf{X^T C_Z^{-1} X}\big)^{-1}\big(\mathbf{x}(\mathbf{s_0})^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)^T\\
&+ \big(\mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big)\big(\mathbf{C_Z}\big)\big(\mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\big)^T\\
&= \big(\mathrm{x}(\mathrm{s}_0)^T - \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{X}^T\big)\big(\mathbf{X^T C_Z^{-1} X}\big)^{-1}\big(\mathbf{x}(\mathbf{s_0}) - \mathbf{X}{\mathbf{C_Z}}^{-1}\mathbf{c_Y}(\mathbf{s_0})\big) + \mathbf{c_Y}(\mathbf{s_0})^T{\mathbf{C_Z}}^{-1}\mathbf{c_Y}(\mathbf{s_0})
\end{aligned}$$

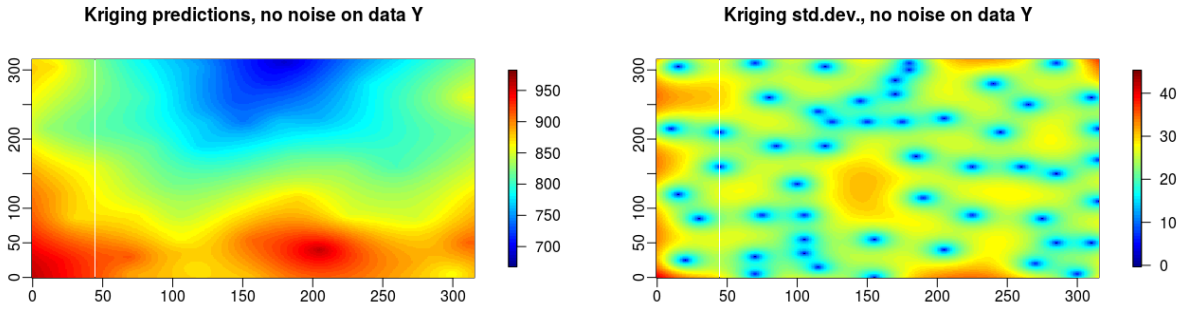



Figure 4: Kriging with no noise on Y

giving the predictor variance.

**c)** The p-dimensional vector $\mathbf{x}(\mathbf{s})$ and expected value of $Y(\mathbf{s})$ will be on the form

$$\mathbf{x}(s_u, s_v) = \begin{bmatrix} 1 \\ s_u \\ s_v \\ s_u s_v \\ s_u^2 \\ s_v^2 \end{bmatrix} \quad \text{and} \quad Y(s_u, s_v) = \beta_0 + s_u\beta_1 + s_v\beta_2 + s_u s_v\beta_3 + s_u^2\beta_4 + s_v^2\beta_5$$

We can interpret this model as an linear quadratic model of coordinates $s_u, s_v$ with continous factors $\beta$ coordinates and a linear interaction effect between the two coordinates $s_u, s_v$. The resulting kriging predictions and std.deviance are presented in 4. As evident from the results, we see that the kriging predictions nicely captures a realistic picture of how the data could look like. The variance is also greater the further away from any original datapoints that we get. We also notice that there is no "nugget effect" as the estimated variance is 0 for the original datapoints themselves. In this particular case that would make sense, as we don't expect any microscale variation to an attribute as height.

**d)** Considering the node $\mathbf{s_0} = (100, 100)$ we achieved the predicted value as 840.345 and prediction variance as 390.771. To assess the probability that the nodes real height is above 700 meters, we note that the conditional probability $[Y(\mathbf{s_0})|Z = s^*]$ is Gaussian distributed with predicted value as expectancy and prediction variance as its variance. With these data, we achieve an probability of 1, by the calculation $P(Z > \frac{700-840.345}{\sqrt{390.771}})$, that the real height is above 700 meters. To assess at which height there is a 90 percent likelihood of it being below, we solved the equation $P(Z < \frac{x-840.345}{\sqrt{390.771}}) = 0.9$ to find the value as $x = 865.6786$.

**e)** The preditcion results from the noisy data are presented in 5 and 6. We observe that the predicted difference deviates more as the noise becomes greater, which seems reasonable. However, on the scale of 800 meters, a standard deviation of $\sqrt{5}$ or $\sqrt{15}$ is very small compared to the datavalues, as we can see by how similar the kriging prediciton plots became for the noise-free as well as the noisy data. Worth to mention is that the estimated variance of the data is the same for all three datasets, which is due to the variance only depending on spatial properties and not measurements in this task. Thus the estimated variance is similar for all the datasets.

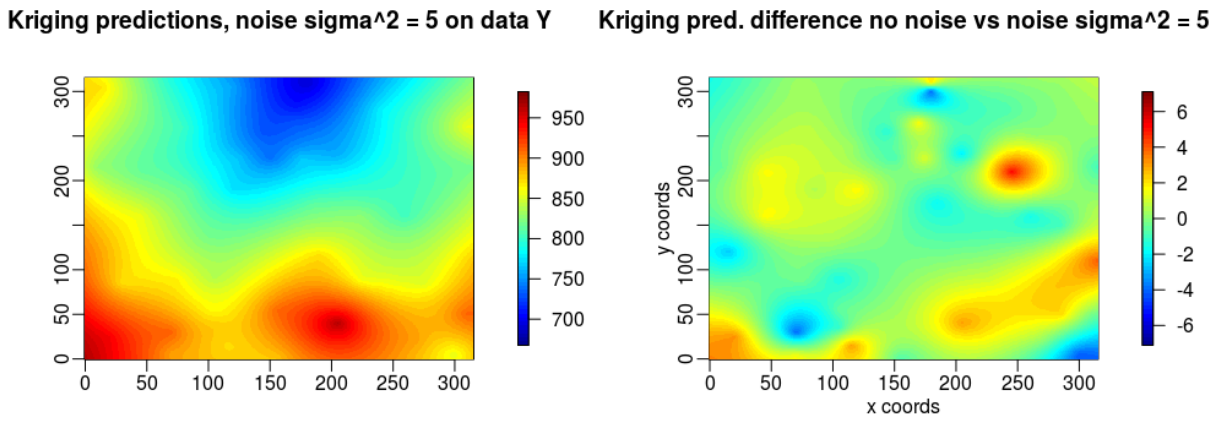



Figure 5: Kriging with noise on data $\mathbf{Y}$ with $\sigma^2 = 5$

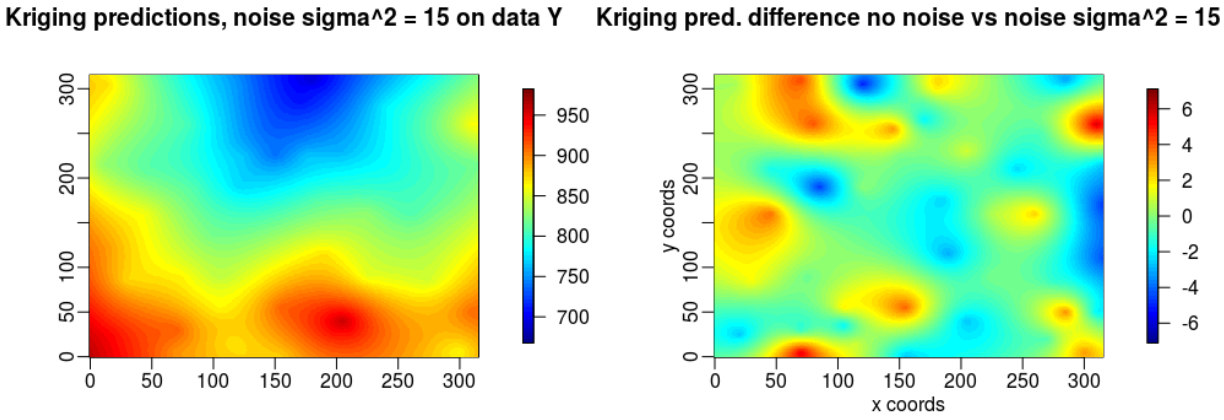



Figure 6: Kriging with noise on data $\mathbf{Y}$ with $\sigma^2 = 15$

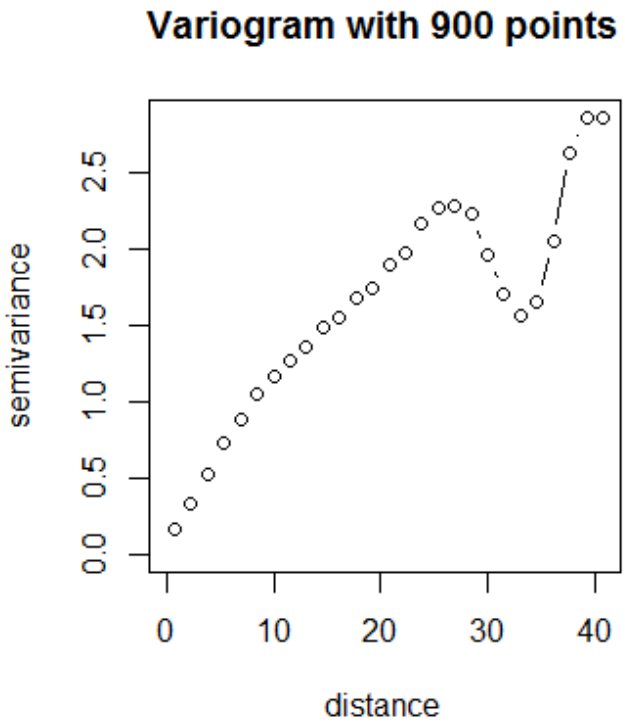
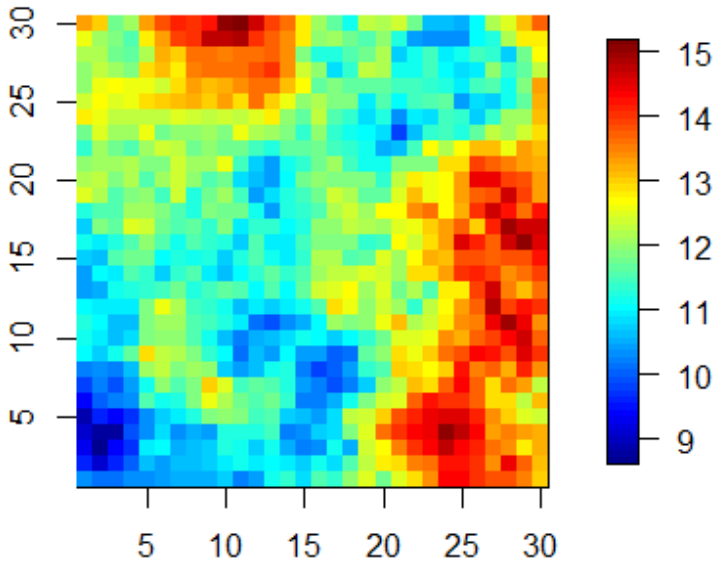



Figure 8: Variogram of the realisation of the Gaussian temperature field from figure 7 using all points in the grid.

## Problem 3

**a)** The Gaussian random temperature field is multivariate normally distributed with a covariance matrix determined by an exponential covariance function. $\mu$ is the expected value of the temperature in a random point in the field, $\sigma_1^2$ is the variance of the temperature, whereas $\theta_1$ is a parameter that is inversely proportional to how fast the correlation between two points in the field decreases as the distance between the two points increases. For a spatial covariance function to be valid it must be positive semi-definite. Figure 7 shows a realisation of the field.

**b)** Figure 8 shows the empirical variogram of the realisation from figure 7. It has the rough shape that we would expect from an exponential covariance function, but it has a strange-looking dip when the distance between two points is between 30 and 40km. This is most likely due to the inherent randomness of the model, rather than an error in the code. If a larger grid was chosen we would expect the empirical variogram to be more like the theoretical one.




**c/d)** In the top left graph in figure 9 9 random points have been drawn from the random field and these points have been used to create an empirical variogram. In the other three graphs the same has been done, but with 16, 36 and 100 random sam-

ples instead. The estimated values of the three parameters in the covariance function based on these samples were found through maximum likelihood estimation and are shown in the table below.

| Sample siZe | $\mu$ | $\sigma_1^2$ | $\theta_1$ |
|---|---|---|---|
| 9 | 12.763 | 1.117 | 4.420 |
| 16 | 12.045 | 1.624 | 8.120 |
| 36 | 12.513 | 2.104 | 8.003 |
| 100 | 11.997 | 1.616 | 8.603 |

From the figure we notice that the shape of the variograms with 9 and 16 points bear little resemblance to the theoretical variogram of the covariance function, whereas the ones with 36 and 100 points only bear superficial resemblance. From the table we notice that all samples give far to small values for $\theta_1$, although for the three largest sample sizes the estimates are quite similar. Thus this inaccuracy appears to be a consequence of the randomness of the Gaussian field, rather then a consequence of the random nature of the samples chosen. The estimates of $\mu$ and $\sigma_1^2$ are generally good, although a bit inaccurate. One additional thing we notice is that although sample size 9 clearly gives worse estimates of all three parameters then the higher sample sizes, increasing the sample size does not necessarily give better estimates, as is to be expected as the sample points are chosen randomly.

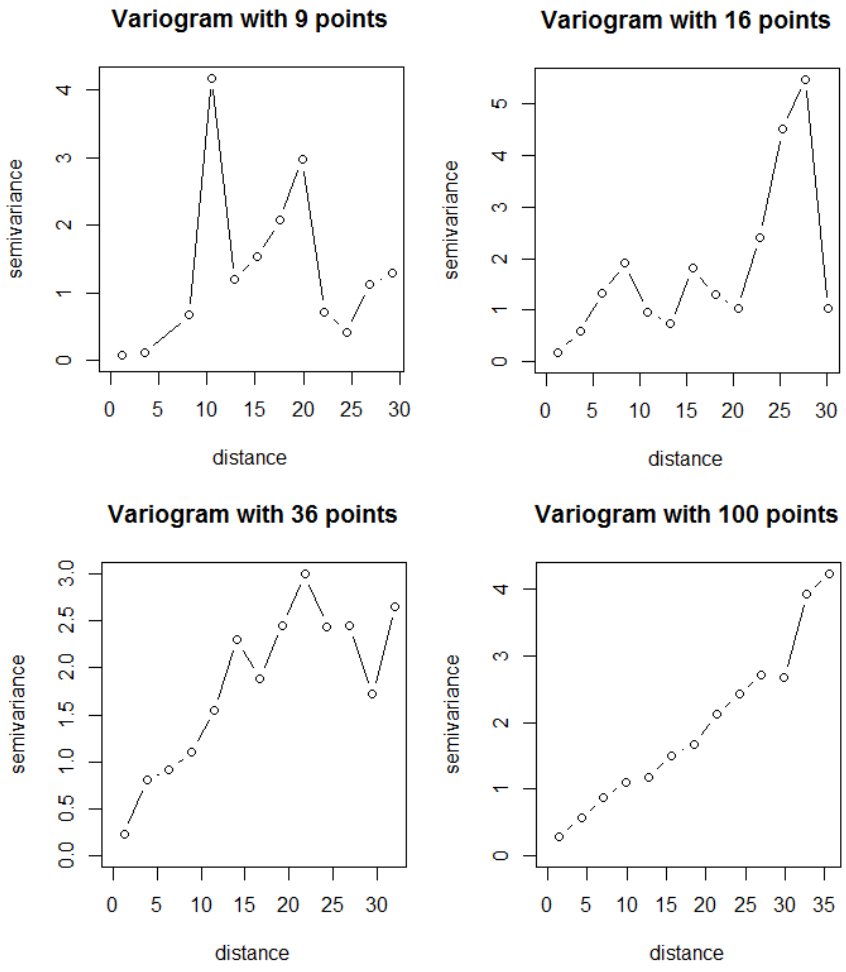



Figure 9: Variograms of random samples with four different sizes chosen from the Gaussian temperature field in figure 7.