Project 1 - SPATIAL STATISTICS

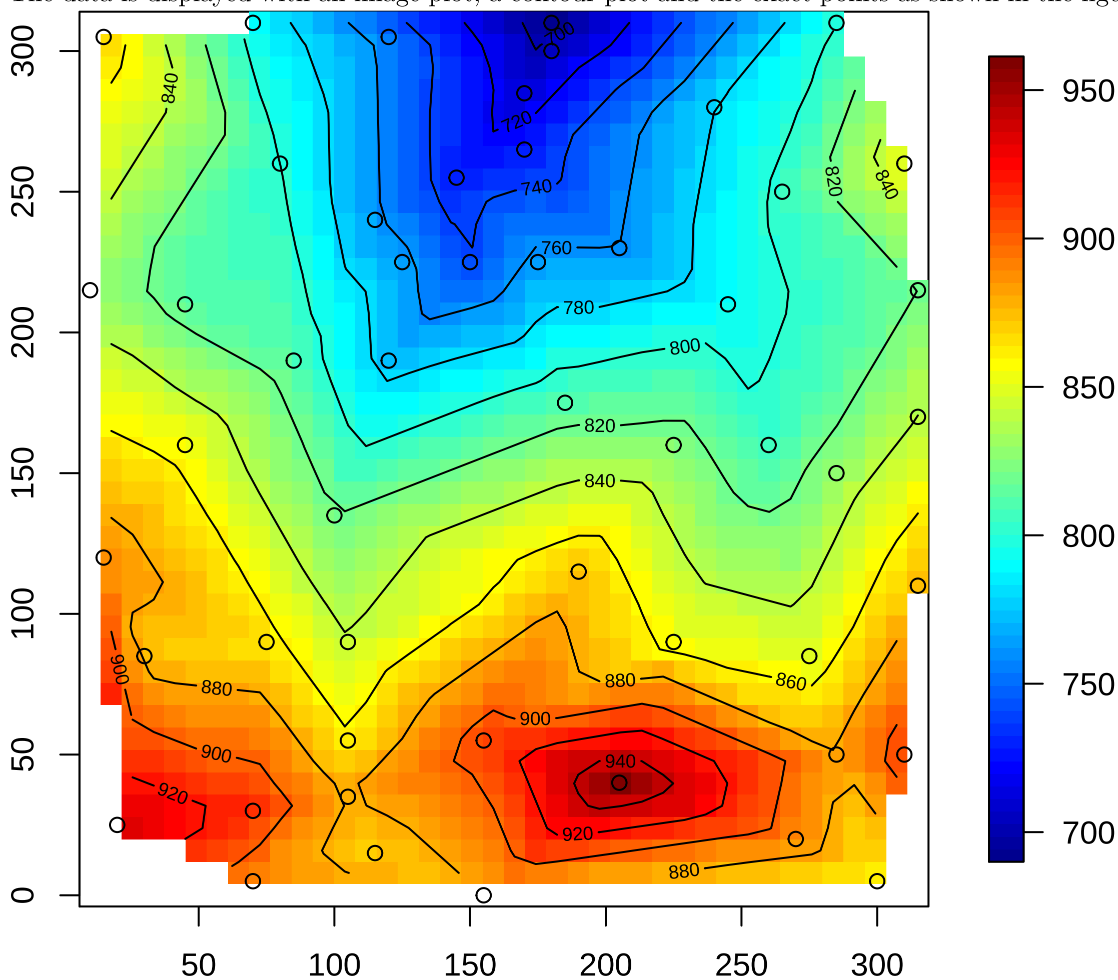
Kwaku Peprah Adjei, Amir Ahmed

Problem 2) Gaussian RF - real data

Given the domain $D = [(0, 315) \times (0, 315)] \subset \mathbb{R}^{\nvDash}$. We let $\mathbf{d} = r(\mathbf{x_1^0}), ..., r(\mathbf{x_{52}^0}))^T$.

a) Display of the data

The data is displayed with an image plot, a contour plot and the exact points as shown in the figure **??** below.



It was observed that the data points did not move in the same direction as with the x and y cordinates (see figure 1 a;b). This suggest that the data is not mean stationary. Moreover, a density plot for the data points in figure {fig:fig2} c) shows a skewness in the data, making the Guassianity assumption doubtful. Hence a stationary Gaussian RF may not be appropriate.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

b) Let

$$\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^{\nvDash}\}$$

be the Gaussian RF that is used to model the domain D.

Given that $E\{r(x)\} = (\mathbf{gx})^T \boldsymbol{\beta}_\mathbf{r}$, $Var\{r(\mathbf{x})\} = \sigma_r^2$ and $Corr(r(\mathbf{x}), r(\mathbf{x'})) = \rho_r(\frac{\tau}{\xi})$. We assume that σ_r^2, ξ are assumed known but $\boldsymbol{\beta}_\mathbf{r}$ is unknown. This is therefore a universal krigging problem.

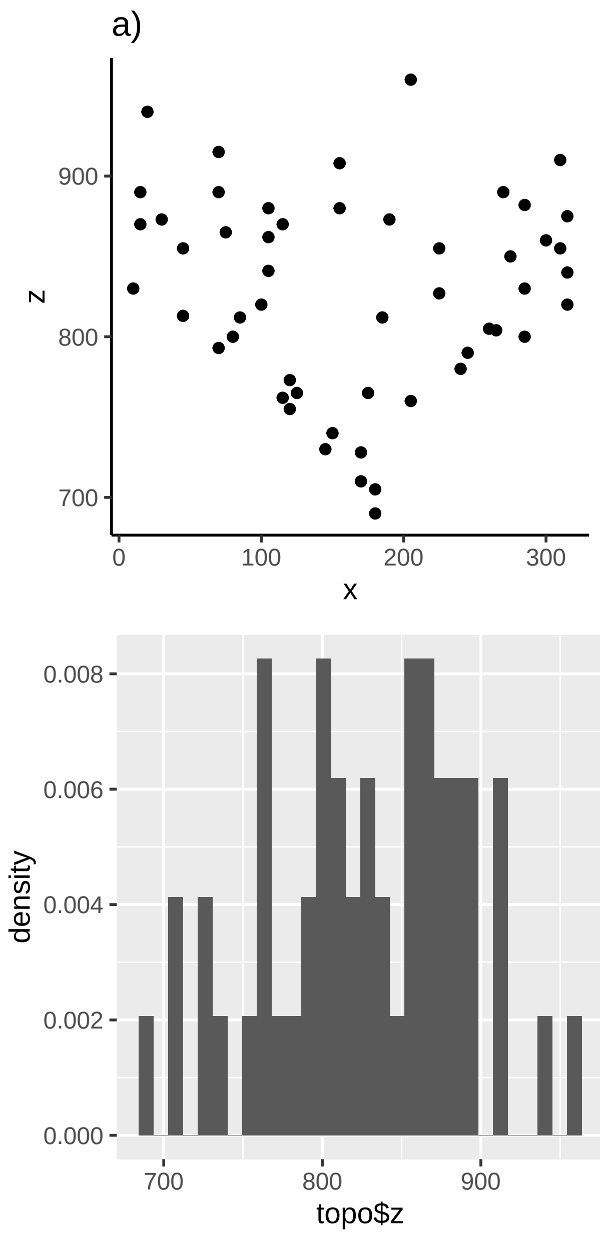


Figure 1: Plot of the data points with respect to their a) x-cordinates and b) y-cordinates; and c) shows the density distribution of the terrain elevation observations.

Let the krigging predictor be:

$$\hat{\mathbf{r}}_{\mathbf{0}} = \boldsymbol{\alpha}^T \mathbf{r}^{\mathbf{d}}$$

We discretise the predictor as:

$$\{\mathbf{r}_{\boldsymbol{\Delta}}(\mathbf{x}) = r(\mathbf{x}) - \mu_r^0 - \sum_{i=1}^{n_g} \beta_r^j g_j(\mathbf{x}); \mathbf{x} \in D\}$$

For the estimator to be unbiased,

$$E\{\hat{\mathbf{r}}_{\mathbf{0}} - \mathbf{r}_{\mathbf{0}} = 0\} \implies \sum_{i=1}^{m} \alpha_i E\{r_i^d\} - E\{\mathbf{r}_{\mathbf{0}}\} = 0$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_g} \alpha_i \beta_r^j g_j(\mathbf{x}_i) = \sum_j \beta_r^j g_j(\mathbf{x}_0)$$

For the estimator to be unbiased,
$\sum_{i=1}^{m} \alpha_i g_j(\mathbf{x}_i) = \sum_j \beta_r^j g_j(\mathbf{x}_0)$.

$$\begin{aligned} Var\{\hat{\mathbf{r}}_{\mathbf{0}} - \mathbf{r}_{\mathbf{0}}\} &= E\{(\hat{\mathbf{r}}_{\mathbf{0}} - \mathbf{r}_{\mathbf{0}})^2\} \\ &= Var\{\alpha_i\{r_i^d\} - \mathbf{r}_{\mathbf{0}}\} \\ &= \sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \alpha_j \rho_{ij} + \sigma^2 + 2\sigma^2 \sum_{j=1}^{m} \alpha_j \rho_{j0} \end{aligned}$$

Hence, we find $\hat{\boldsymbol{\alpha}}$ such that

$$\hat{\boldsymbol{\alpha}} = argmin_{\boldsymbol{\alpha}} Var\{\hat{\mathbf{r}}_{\mathbf{0}} - \mathbf{r}_{\mathbf{0}}\}$$

and subject to the constraint $\sum_{i=1}^{m} \alpha_i g_j(\mathbf{x}_i) = \sum_j \beta_r^j g_j(\mathbf{x}_0)$ for $j = 1, 2, ..., n_g$.

Problem (c)

Considering the case with $E(r(\mathbf{x})) = \beta_1$, we estimated the universal krigging predictor and variance as follows:

```
## krige.conv: model with constant mean
## krige.conv: Kriging performed using global neighbourhood
```

Problem (d) The resulting polynomial function becomes:

$$(\mathbf{gx}) = (1, x_v, x_h, x_v x_h, x_v^2, x_h^2)$$

The expected value of $r(\mathbf{x})$ then is:

$$E\{r(\mathbf{x})\} = \beta_1 + \beta_2 x_v + \beta_3 x_h + \beta_4 x_v x_h + \beta_5 x_v^2 + \beta_6 x_h^2.$$

We present the predictions and the associated variance in the figure below:

```
## krige.conv: model with mean given by a 2nd order polynomial on the coordinates
## krige.conv: Kriging performed using global neighbourhood
```

Problem e)

```
## krige.conv: model with constant mean
## krige.conv: Kriging performed using global neighbourhood
```

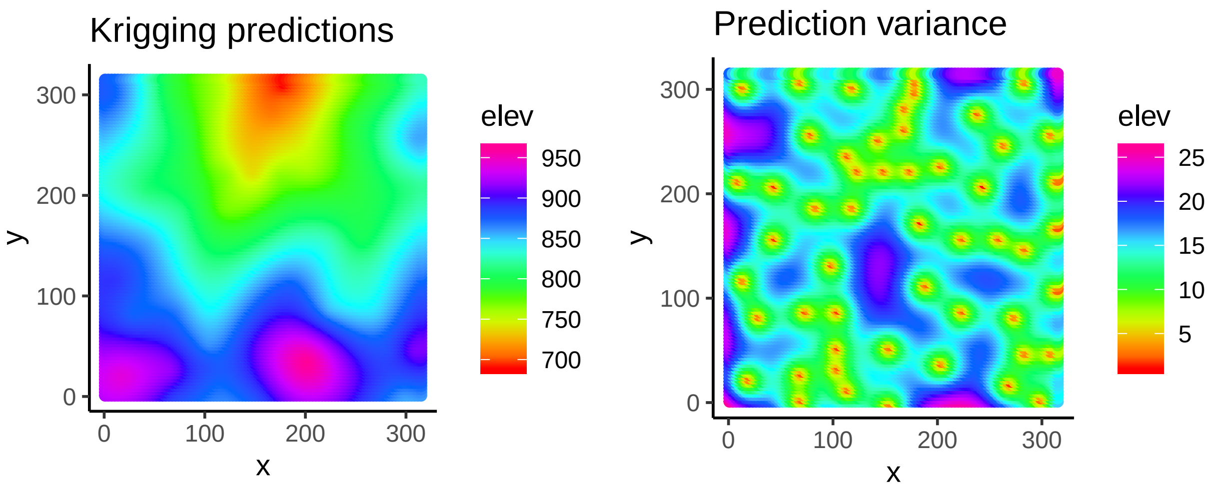


Figure 2: Krigging predictions and prediction variance of the ordinary krigging method.

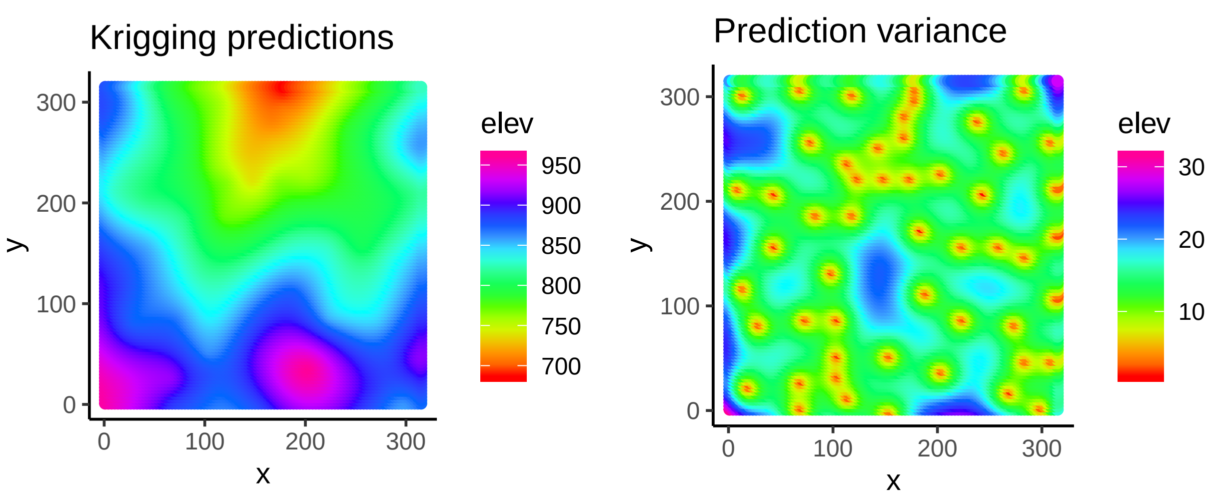


Figure 3: Universal krigging predictions and prediction variance.

We now consider the grid node, $\mathbf{x_0} = (100, 100)$. Using the ordinary krigging, we estimated the predicted mean as 838.6781414 and the predicted variance as 10.044233. Assuming Gaussianity of the data,

$$\begin{aligned} P(r\{\mathbf{x_0}\} > 850) &= P\left(\frac{r\{\mathbf{x_0}\} - E(r\{\mathbf{x_0}\})}{\sqrt{Var(r\{\mathbf{x_0}\})}} > \frac{850 - E(r\{\mathbf{x_0}\})}{\sqrt{Var(r\{\mathbf{x_0}\})}}\right) \\ &= 1 - \Phi\left(\frac{850 - E(r\{\mathbf{x_0}\})}{\sqrt{Var(r\{\mathbf{x_0}\})}}\right) \end{aligned}$$

The resulting probability is 0.13.

To obtain the elevation for which it is 0.90 probability that the true elevation is below it, we used the formular,

$$\begin{aligned} P(r\{\mathbf{x_0}\} > r\{\mathbf{x_{new}}\}) &= 0.90 \\ r\{\mathbf{x_{new}}\} &= E(r\{\mathbf{x_0}\}) + \phi(0.90)\sqrt{Var(r\{\mathbf{x_0}\})} \end{aligned}$$

We obatained 851.55m to be that elevation that satifies the preamble.