MODULE 12 GEOSTATISTICS

REPORT

ON

MAPPING EXERCISE

SUBMITTED BY

GROUP5: MWANAIDI MOHAMEDI, NIVEDHITHA AJITHKUMAR, QIAO REN, RAKTIM GHOSH

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

LIST OF TABLES

LIST OF FIGURES

GROUP 5: MWANAIDI MOHAMEDI, NIVEDHITHA AJITHKUMAR, QIAO REN, RAKTIM GHOSH

Question (a)

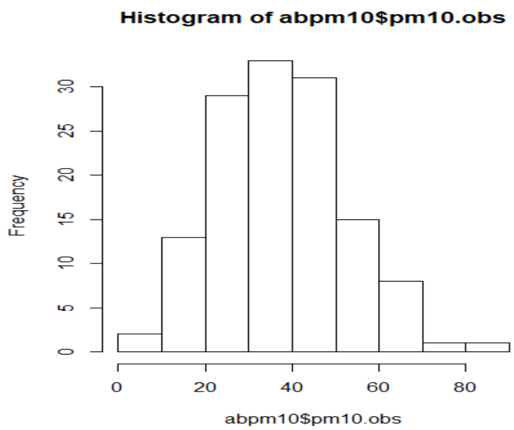

Figure 1: Histogram of PM10.obs data

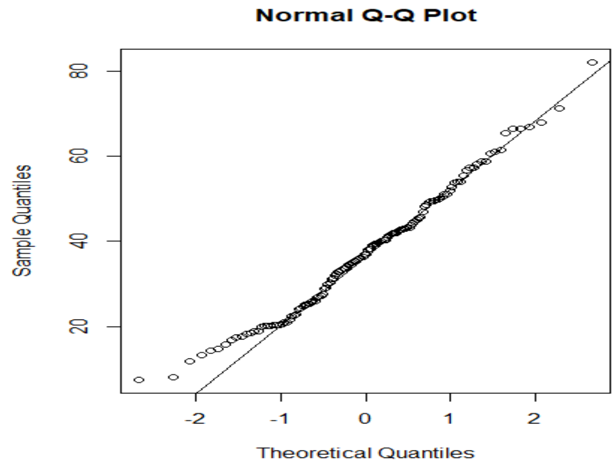

Figure 2: Normal QQ Plot of PM10.obs data

```
hist(abpm10$pm10.obs)
library(e1071)
skw1 = abpm10$pm10.obs
skewness(skw1)
lt <- log(abpm10$pm10.obs)
head(lt)
hist(lt)
skewness(lt)
```

```
> skewness(skw1)
[1] 0.3223979
```

Figure 3： Screenshot of the histogram codes and its skewness

```
qqnorm(skw1)
qqline(skw1)
qqnorm(lt)
qqline(lt)
```

Figure 4: Screenshot of the code for plotting QQ plot

The data were analysed by using QQ plot and the histogram. The results show that the data were skewed to the right in the histogram in figure 1. Log transformation was applied to make it normally distributed, but the skewness was increased after the data were transformed using logarithm function. The absolute value of the skewness after log-transformation (0.74) is larger than the absolute value of the skewness before log-transformation (0.32).

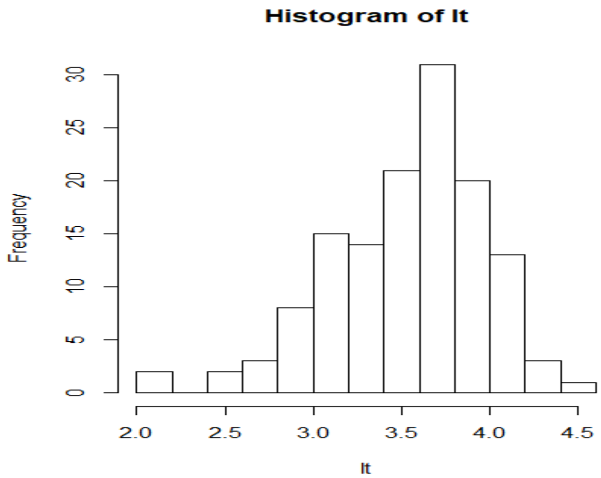

Figure 5: Histogram of log Transformation(PM10.abs) data

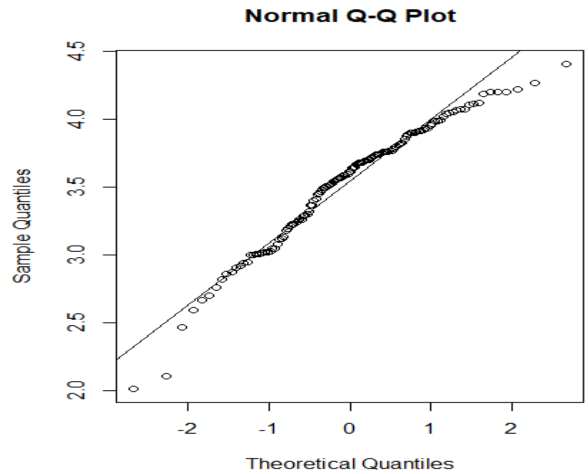

Figure 6: QQ Plot of Log transformation (PM10.abs)

Figure 7: Screenshot of the skewness of the histogram in figure 6

Question (b)

- Yes, the number of points at each lag is sufficient to obtain a robust estimate of the sample variogram.
- Yes, the variogram approaches a sill.
- Case 1: The cut-off was set to 1200 while the bin width was changed to 25, 30, 35 and 40. In this case, the sill had reached at all bin width except at 25 (see Table 1). Case 2: The bin width was kept 30 while the cut-off was changed to 500,600,1000 and 1200. In this case at 500 and 600 cut off the sample variogram was increasing without bounds but between 1000 and 1200 the sill was reached. Although in case 1, when increasing the bin width from 30 to 40, 50 etc. the standard error decreases, but RMSE value increases. So, it cannot be concluded to take a particular width as the best choice. Our choice of sample variogram was at the cut off angle 1020 and bin width 30 because the sill was reached, and the points were fitted nicely in a variogram. See figure 8

Cut-off = 1200		
Bin width	**Partial sill**	**Range**
25	176.8537	108.9925
30	203.60757	346.3699
35	207.1142	365.0068
40	221.7182	454.883

Table 1: Showing the variation of Sill and range of case 1

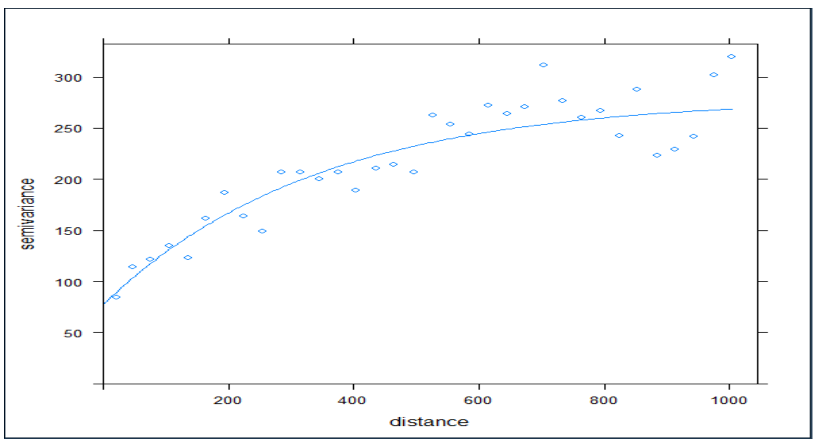

Figure 8: Variogram when cut-off is 1020 and width is 30

Question （c）

Three variogram models were taken into consideration namely Spherical, Matern, and Gaussian by comparing with exponential model. The estimated parameters, accuracy and the ME and RMSE is shown in a table 2 below

Table 2: Estimated parameters, accuracy, ME and RMSE values of different variogram models

Model	Psill	Range	ME	RMSE	Standard error
Spherical	171.94879	648.1431	-0.1693073	9.486547	45.35488
Matern	201.52703	339.9803	-0.2023866	9.416813	37.00299
Gaussian	114.18956	147.1899	-0.2012574	9.591495	85.04432
Exponential	201.52704	339.9804	-0.2023866	9.416813	37.00299

Exponential and Matern variogram models are considered to be most appropriate models compared to others because of the low RMSE and standard error as shown in the table 2 above.

Question (d)

The sampled variogram is shown in Figure 9. The modelled variogram is shown in Figure 10. After we apply the exponential model to fit the variogram, we get this model (Figure 10). The nugget is 77.66456. The nugget is a non-spatial variability. The nugget indicates the measurement error, including instrument error and operator variability. Sill = nugget + partial sill. In this case, sill= 77.66456+201.52704= 279.1916. It means that the total variability is 279.1916. This means that the variogram sizes increasing when the range reaches the 339.980. The range, 339.980, is an effective range. When the semi variance reaches 95% of the sill, the lag is 339.980. When the lag exceeds this range, the regionalized variables are no more auto-correlated. Because there is no dependency between the unsampled location and the observations.

In the modelled variogram, the slope is gradual and not steep. This indicates that, with increase of lag, the semi-variation between the observations will increase gradually. Therefore, in interpolation, the predicted values will slowly variate from each other.

Table 3: showing the value of model variogram

model	psill	effective range
Nug	77.66456	0.0000
Exp	201.52704	339.980

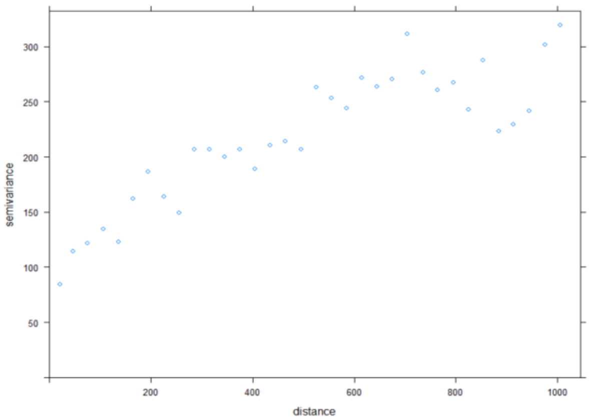

Figure 9: Sample Variogram

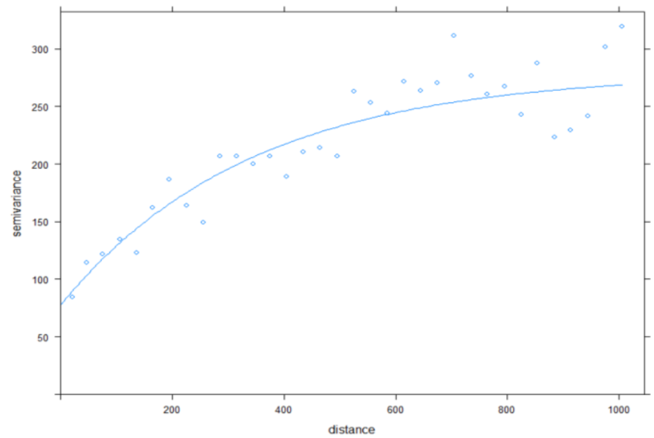

Figure 10: Modelled Variogram

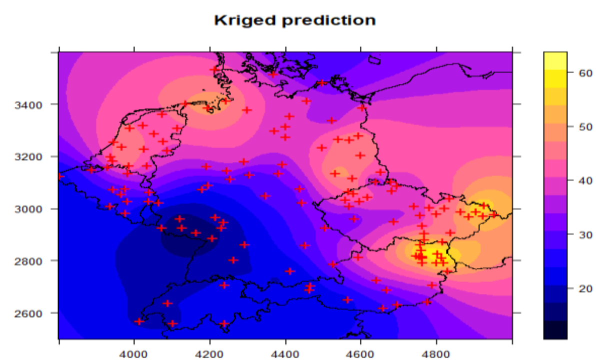

Figure 11: Graph showing the kriged prediction

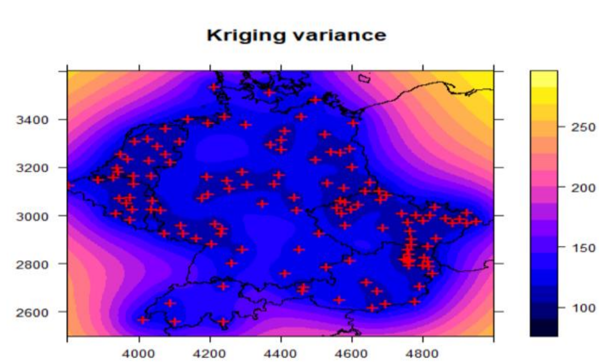

Figure 12: Graph showing the kriged variance

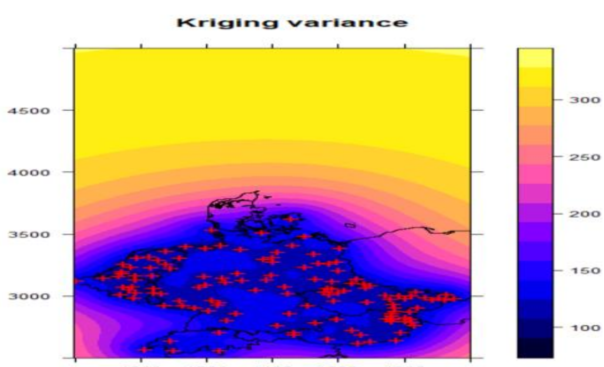

Figure 13: Graph showing the kriged variance

Question(e)

Table 4: Showing the maximum and minimum predicted value of from data PM10

	value	easting	northing	description of the location
highest predicted concentrations of PM10	57.4350	4803.5075	2830.0915	located in Vienna, Austria
minimum kriging variance	91.7702	4758.8263	2819.1860	located in Vienna, Austria
maximum kriging variance	192.4693	4454.2832	3623.0754	located in Denmark

The table above shows the maximum and minimum values with their corresponding coordinates. The highest predicted concentrations of PM10 is located in Vienna, Austria. This matches with what we observed in figure 11. The smallest variance is located in Vienna, Austria. The reason is very likely that the observations at this region is highly dense. In general, the closer to the observations, the lower the kriging variance is. The further away from the observations, the higher the kriging variance is. The highest variance is located in Denmark. If the lag exceeds the range, then the semi-variance will keep constant and keep being super large.

Question(f)

The Dataset used is of air quality observation for the European Economic Area. It provides the 24-hour data of PM10 (particulate matter less than 10 um in diameter) concentration were extracted.

Co-kriging is used when estimating one variable with two variables. Spatial and cross covariance model of the variable is required, and the two variables should have strong relationship between them. Here our dataset contains the PM10 concentration of each location in Europe, whose value does not have any relationship with each other.

Regression kriging is used when there is spatial variability of mean in the data with respect to a variable, which can be a location, elevation etc. Thus, there is an increase in variogram model without any bound(non-stationary). When auxiliary information is available in the form of maps of covariates which can explain part the variation in the target variable, then regression kriging will give more information compared to ordinary kriging (Hengl, Heuvelink, & Rossiter, 2007). But as mentioned earlier, the dataset contains only the extracted PM10 concentration and the coordinate values for which ordinary kriging is worth considered than regression kriging.

References

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers and Geosciences*, *33*(10), 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001