Spatial Statistics

Assignment No 8

Variogram

Muhammad Bilal
Matriculation: - 12214473

Q.No.1

- **The following line calculates an empirical variogram from precipitation data.**
- **Part A: - Interpret in your own words (range, lag size, sill, partial sill, spatial autocorrelation).**
- **Part B: - What lag size is used.**

```
empirical_variogram <- gstat::variogram(Z_1_10MM~1, data = dat_swiss, width = 5000)
plot(empirical_variogram)
```

Part A

The variogram is a technique in spatial statistics to understand the spatial dependence or correlation between values at different locations. To determine spatial autocorrelation in precipitation dataset I applied semivariogram as statistical function of distance. So first I plotted the mean yearly precipitation data using ggplot library as shown in figure 1.

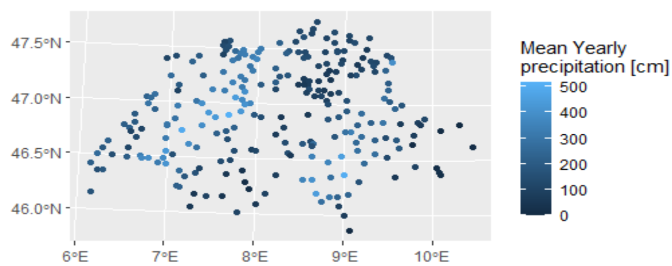

Figure 1

Range: - It is the distance at which spatial autocorrelation reaches its maximum value. In the plot, it can be identified as the point where semivariogram models flatten out. In our case the range is 70000 meters because at this distance semivariance stops increasing in the precipitation dataset as shown in figure 2.

Lag Size: - The number of data point pairs used to calculate the variogram at a specific lag distance**.** Larger lag sizes provide more data points for calculating the variogram and can improve the stability of the estimation. In our case the lag size is 5000 means this amount of precipitation points used to calculate spatial autocorrelation at specific distance Interval as shown in figure 2.

Sill: - It represents the maximum semivariance value reached as the lag distance increases. The value at which the semivariance levels off or reaches a constant value is called the sill. It indicates the maximum amount of spatial correlation present in the data. Values below the sill indicate a certain level of spatial correlation, while values at or near the sill assume that there is no additional spatial correlation beyond that distance. In our case 17000 is observed as maximum semivariance value beyond this value we assume that there is no spatial autocorrelation in the precipitation dataset as shown in figure 2

Nugget: - The **nugget** in a variogram plot represents the variability at very small distances (lag distances close to zero) that is not influenced by spatial correlation. Nugget indicates the presence of randomness or factors other than spatial proximity affecting the data. In our case the nugget effect occurs in 2000 semivariance value at y-axis in the plot as shown in figure 2.

Partial Sill: - It is the difference between Sill and Nugget. The partial sill represents the contribution of spatial correlation to the total variability or semivariance observed in the data. In our case PS is 15000 that is obtained from the difference of Sill and Nugget.

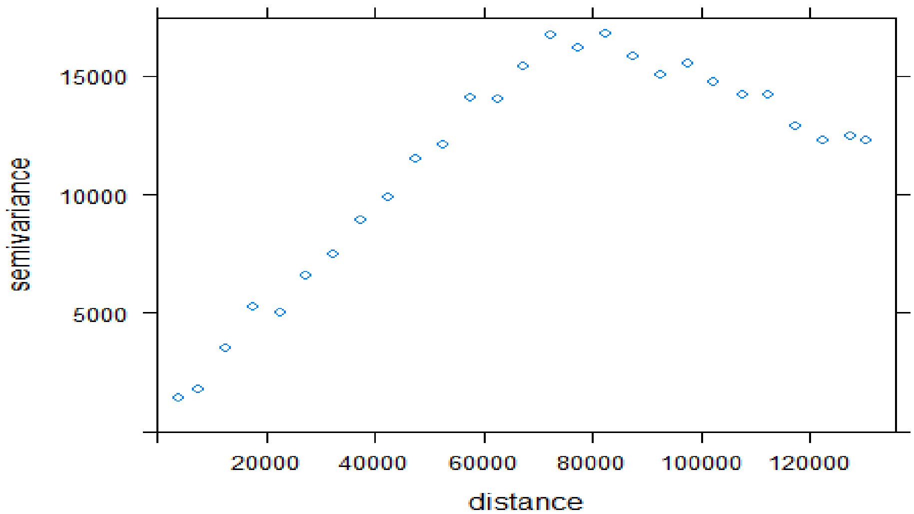

Figure 2

Spatial Autocorrelation: - It is basically the degree of similarity or correlation between values of a precipitation at different locations in space. The variogram plot visualizes the spatial autocorrelation by plotting the semivariance values against the lag distances. In our case correlation between values of precipitation decreases as the spatial lag distance increases. In our data maximum autocorrelation occurs in the dataset at 15000 semivariance value beyond this correlation assumed to be negligible or nonexistent as shown in figure 2.

Part B

Lag Size: - In our case the lag size is 5000 means this amount of precipitation point used to calculate spatial autocorrelation at specific distance as shown in figure 2.

Q.No.2

- **The next line calculates a theoretical semivariogram.**
- **Part A: - What is the difference between theoretical and empirical semi variogram?**
- **Part B: - Explain parameter values in function vgm in your own words.**

```
estimated_variogram.model <- gstat::vgm(15000, 'Sph', 70000, 0.1)
plot(empirical_variogram, model = estimated_variogram.model)
```

2.Part A

Theoretical Semivariogram: - It is based on a specific mathematical or statistical model that describes the underlying spatial correlation structure of the data. The theoretical semivariogram is calculated based on assumptions about the spatial dependence, and it provides a predicted semivariance value at each lag distance as shown in figure 3.

Empirical Semivariogram: - The empirical semivariogram is a practical measure of spatial dependence that does not assume any specific model or structure. It provides an estimate of the true semivariogram based on the observed data points**.**

2.Part B

It's a theoretical semivariogram implementation.
vgm (15000, 'Sph', 70000, 0.1)

Psill: - It is the difference between Sill and Nugget. The partial sill represents the contribution of spatial correlation to the total semivariance observed in the data. In our parameter we assume that 15000 semivariance value is Psill and beyond that variance spatial correlation assumed to be negligible or nonexistent.

Model (Sph): - This parameter specifies the model type for the variogram. In this case, the model type is set to 'Sph', which stands for spherical. The spherical model assumes that the spatial correlation decreases in a spherical manner from the origin until it reaches the range.

Range (7000): - The range is assumed to be 70000 meters at this distance semivariance stops increasing in the precipitation dataset as shown in figure 2.

Kappa (0.1): - Smoothness parameter for the Matern class of variogram models.

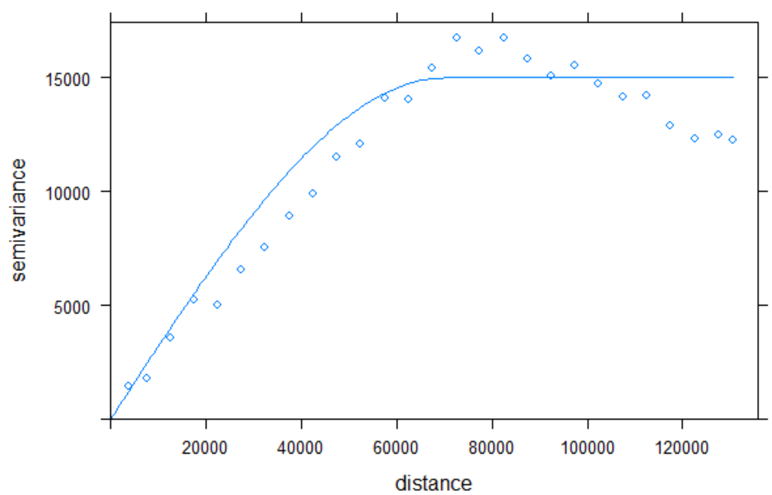

Figure 3

Q.No.3

- **Part A: - Define another variogram model of your choice to fit the sample variogram?**

estimated_variogram.model <- gstat::vgm(15000, 'Exp', 70000, 0.1)
plot(empirical_variogram, model = estimated_variogram.model)

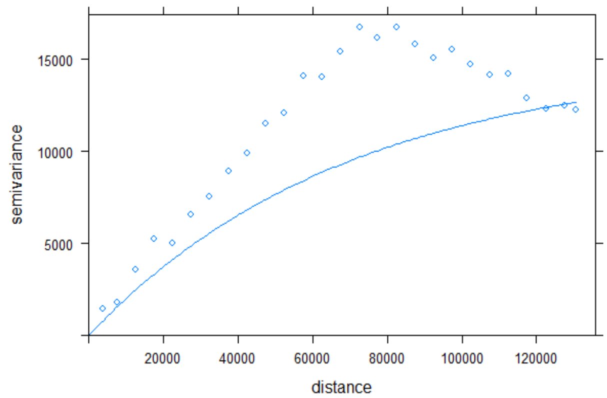

Figure 4 (Exponential Model)

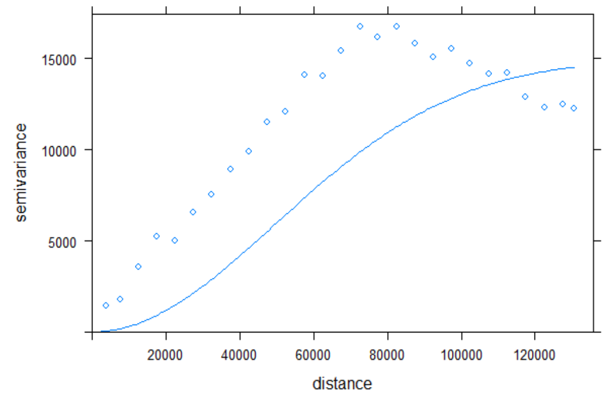

Figure 5 (Gaussian Model)

I applied a different model like Exponential, Gaussian Model to fit on the precipitation variogram. Both Model does not fit properly to our data, but Exponential model seem to more relevant that shows spatial correlation decreases exponentially with increasing distance. It does not have a distinct range but continues to decrease gradually without reaching to sill.

Q.No.4

- **Part A: - Directional sample variograms consider anisotropy in spatial autocorrelation. Why does alpha 0 and 180, 270 and 90, 315 and 135, 225 and 45 yield the same results, respectively?**

Anisotropy is the variation in spatial correlation based on direction. It means that the degree of spatial dependence can differ depending on the direction in which the distances are measured. To account for anisotropy, directional sample variograms are calculated by considering the lag distances and orientations separately in different directions.

Certain pairs of angles yield the same result due to the inherent symmetry of the variogram function. The variogram function quantifies the spatial dependence based on the distance between pairs of points and the difference in their values. It does not directly incorporate the direction of those pairs of points. As a result, if we calculate the variogram for two alternate angles (180 degrees apart), we will obtain the same variogram values because the distance and value differences remain the same.