Comparative Analysis Of Different Kriging Techniques: Implication On The Prediction Of Moisture Variability Over The Mander Area, The Netherlands

GROUP 5:
MWANAIDI MOHAMEDI (s6035493)
NIVEDHITHA AJITHKUMAR (s6038182)
QIAO REN (s6036341)
RAKTIM GHOSH (s6038514)

1. Introduction

Moisture deficit can be defined as a difference between evapotranspiration and precipitation (Kerebel, Cassidy, Philip, & Holden, 2010). One of the major cause of moisture deficit is groundwater extraction. Therefore, quantifying ground water fluctuation (MHW) will have a plausible consequence on statistical interpretation of effect on land quality moisture (MD30) deficit which in turn will help in exploring the correlation between the two. There are various underlying simulation models associated with agriculture which are explicitly used to estimate land quality moisture deficit. However, in view of economic viability, those aforementioned models are not lucrative. Interestingly, kriging and cokriging has gained ground over the field of mathematical modelling and interpolation techniques explicitly on geostatistical applications like soil mapping (McBRATNEY & WEBSTER, 1983). Therefore, quantifying the groundwater fluctuation is of paramount importance for carrying out exhaustive kriging interpolation methods because MD30 and MHW are inextricable correlated with each other (Stein, Hoogerwerf, & Bouma, 1988). In this report, a comparative analysis vis-à-vis of regionalized moisture deficit has been carried out with the help of kriging and co-kriging interpolation techniques over the area of Netherlands.

PROBLEM STATEMENT

This project aims at comparing the performance of three methodologies namely Ordinary, Regression and Co Kriging on predicting unsampled locations of average moisture deficit (MD30).

OBJECTIVE

1. To predict 30-years average moisture deficits (MD30) using Ordinary Kriging at unsampled locations
2. To identify if the fluctuation of Groundwater in terms of Mean Height Water Level (MHW) variable has an influence on MD30 using Regression Kriging
3. To determine the correlation between MD30 and MHW variable using CO-Kriging

Questions

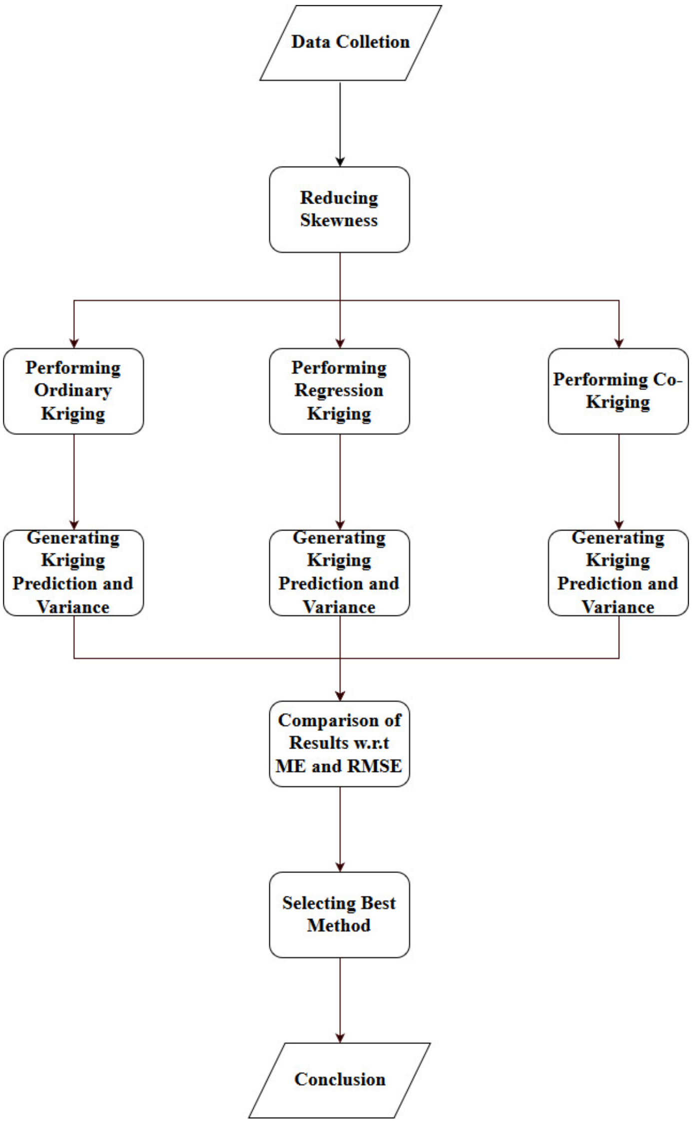
1. Is the mean constant and no fluctuation around the study area?

2. Is the fluctuation of Ground Water Level in terms of MHW has an influence in MD30?

3. Is there any correlation between MD30 and MHW?

2. Methods

In this report, three types of kriging interpolation (Ordinary Kriging and Regression Kriging and cokriging) have been performed by considering the predictor variables in order to predict the land use moisture deficit on unsampled locations. The following flowchart shows the process of this report (figure…). Undoubtedly, Ordinary Kriging has been performed by considering the spatial mean to be same everywhere in the sampled location. After that, Regression Kriging has been carried out by observing the correlation (0.715) between MD30 and MHW. As the auxiliary environmental variable (MHW) has shown a significant correlation with the target variable (MD30), Regression Kriging can be more effectively used to predict the values of moisture deficit at unvisited location. Interestingly, as all values of variables have been recorded at all sampling points, the coherence can be aligned with the fundamental need of co-kriging (Webster & Oliver, 2002). Therefore, Co-kriging has also been explored with the objective of predicting more accurately at unvisited sampling location.

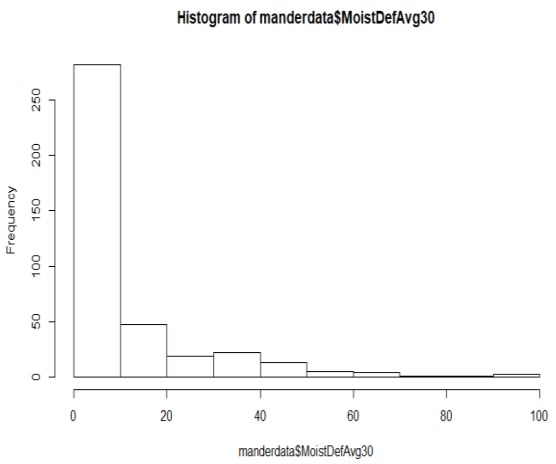


3. Data.

In this study, 400 samples were collected in Mander area of the Netherlands (Stein et al., 1988). Presumptively, the kriging prediction largely depends on the configuration of sampling procedures. Furthermore, the sampling grid plays a crucial role in determining the overall variability of soil map delineations. Moreover, the pre-requisite of applying kriging over a regionalized area is: the data should be normally distributed, the data should meet stationarity condition; the data should not consist of any trend (Webster & Oliver, 2002). Therefore, the skewness has to be closed to zero. In the preliminary dataset, earlier the skewness of MD30 was 2.51. Therefore, the log transformation has been taken to reduce the skewness (-0.28). Similarly, the preliminary skewness of MHW was 1.88 but, interestingly, the log transformation has been carried out thrice resulting into skewness of 0.11. Collectively, the skewness of both log transformed dataset (MD30 and MHW) were tending towards zero. Therefore, this dataset can be used for interpolation. However, even if the statistical distribution of sampling data has always been expected to be normally distributed, in respect of mathematical frameworks, the statistical prediction will always be having a similar implication of predicted result. Additionally, the statistically interpolated results wouldn't have a greater consequence in respect of skewness, i.e., even though, with the help of log-transformed data, the interpolation can be performed, it does not have much effect on the end result.

4. Results

The data were visually inspected to see if they are normally distributed before further manipulation using Histogram and QQ plot so that it will be easy for interpretation. The data were not normally distributed because of the skewness to the right of the histogram. Log transformation was applied to normalize the data and removing skewness tending towards zero.



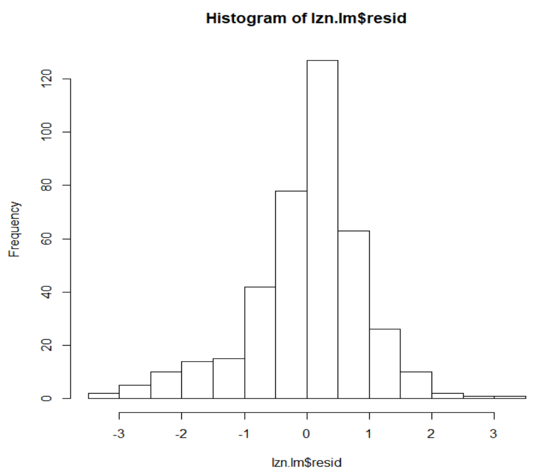


Figure 1; Histogram with skew

Figure 2; Histogram without skew

ORDINARY KRIGING

Assuming the stationarity condition, ordinary kriging was performed to predict the average moisture deficit (MD30) in the soil in unsampled locations. The variogram for OK was performed for MD30. The cut off was 1200 and the width was 70. An exponential model was used to fit the variogram. The model was fitted in such a way that the sum of squared errors of prediction (SSE) is as close to zero as possible. The parameters of the fitting model are: the nugget is 0.7008, the partial sill is 0.835 and the range is 258.21. Prediction for the unsampled value was performed. Closely located samples causes clusters in the prediction. In the kriging variogram, the further away from the observations, the larger the variance is. The cross validation was executed in order to assess the accuracy of OK. The mean error and RMSE are given in the table below.

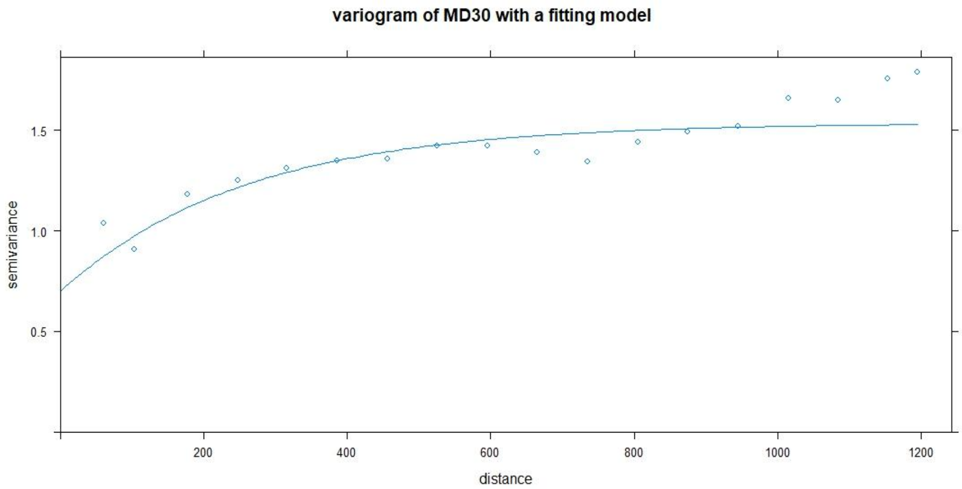
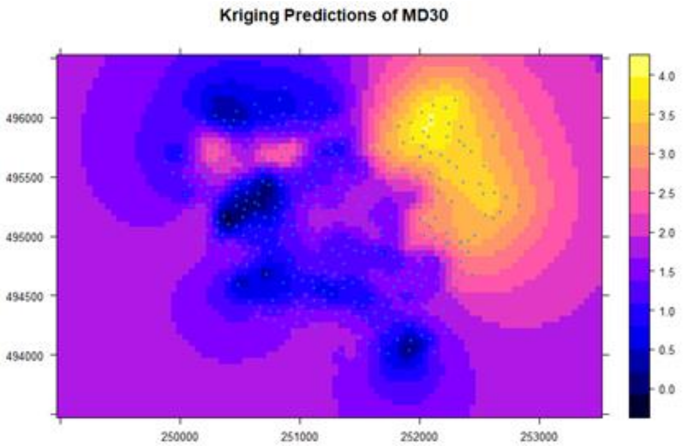


Figure 3; Exponential model variogram



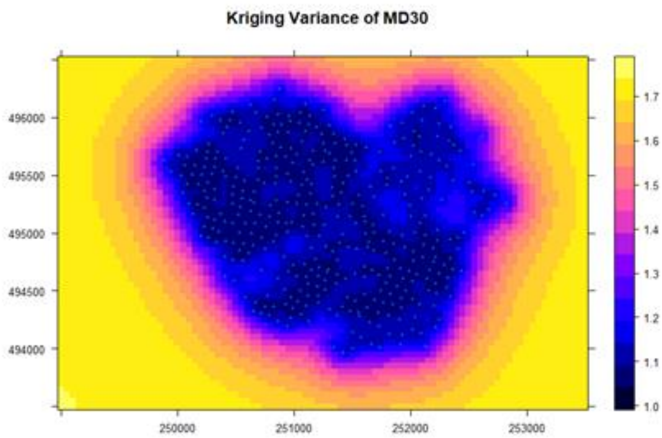


Figure 4; Kriging prediction of MD30

Figure 5; Kriging variance of MD30

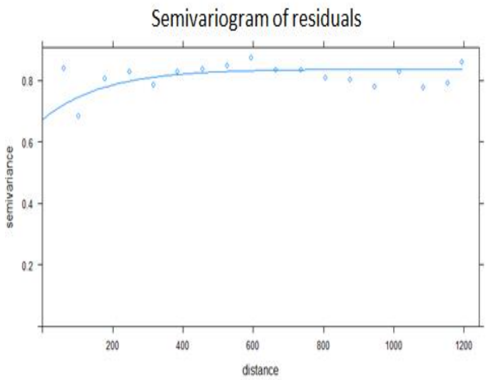
Table1; Cross validation results of Ordinary Kriging

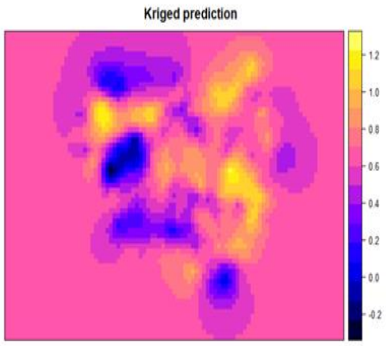
Standard Error of model variogram	Mean Error	RMSE
0.00151	-0.001456	1.01949

REGRESSION KRIGING

Regression kriging was performed to see if the target variable (MD30) has fluctuation with respect to the MHW. Regression Kriging gives more information for prediction than Ordinary Kriging (OK) because in OK only observation data is used for prediction. The variogram in Regression Kriging is estimated using the residuals. The setting for the variogram of the residuals was 1200 for cutoff and width 70. The parameters used to fit the model were 200, 0.2 and 0.2 for range, sill and nugget respectively. The variogram curve becomes smooth when the cut off angle is low.

The cross-validation was also performed for the purpose of evaluating the statistical analysis of independent dataset the results were shown on a table below.





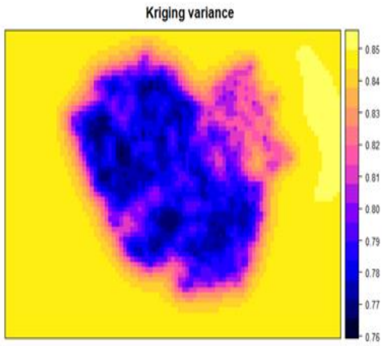
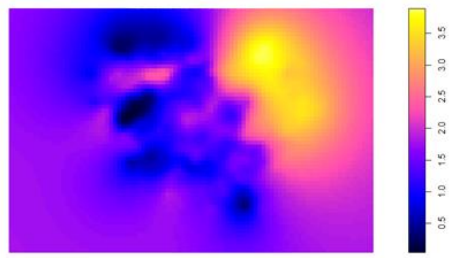


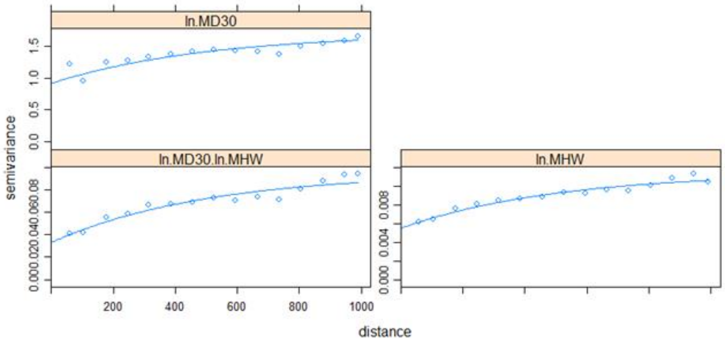
Table2; Cross validation results of Regression Kriging

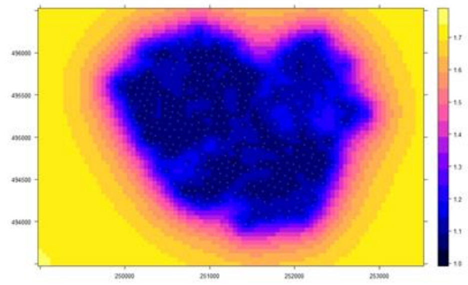
Mean Error	Root Mean Square Error	Variability
-0.001246	0.777800	0.5641274

CO-KRIGING

In CO-Kriging the variable MHW was used as co-variable because of high-correlation (i.e. 0.7155) and due to its clear spatial correlation. We have taken 30 sub-samples points for validation in such a way that the sampling points are highly distributed all over the region. This highly distributed data will help in validating the sampling points more accurately. The variogram and cross variogram of the primary and auxiliary variables are shown in the figure 5b.







5. Discussion

The main purpose of this study was to predict the 30 years average moisture deficit (MD30) at unsampled locations using ordinary kriging (OK). Furthermore, using Regression kriging (RK) identifying if fluctuation of groundwater has an influence on moisture deficit and also using co-kriging (CK) determining the correlation between MD30 and MHW variables. Additionally we also compared the results of different kriging techniques which was performed earlier to conclude which technique was giving the best result.

The sample data of the study is highly skewed distribution. It was transformed to normal distribution for easy interpretation, as it does not give any significance to the techniques used during analysis purpose. Ordinary kriging predicted the MD30 for the unsampled location by producing the RMSE value of 1.019. Similarly, we did RK as there was correlation between the MD30 and MHW. Linear Regression model was fitted to find the fluctuation between the variables. But the result was not as expected. From Figure 4b. the pink patches are all over the region even when we increase the distance. But theoretically and practically it should be like as we increase the distance the prediction value should tend towards the mean, which is yellow patches. So we can conclude , it is not necessary that if there is correlation between the variable, we have to perform Regression kriging.

Co-kriging can be done in two situations, one is under sampling and other is fully sampled case. We have considered fully sampled case i.e. all the variables are recorded at sampling points. Here the principle advantage of fully sampled case is coherence. Co-kriging ensures coherence. As mentioned earlier, in co-kriging we have taken the primary variable as the MD30 and auxiliary variable as MHW as they are highly correlated to each other. Co-kriging was performed and the RMSE produced was 0.68384. Thus as a comparative study, the RMSE produced in co-kriging is less compared to the other two techniques. Low RMSE produces accurate prediction so we can conclude that co-kriging produces best results in our study.

6. Conclusion

The following were found after performing analysis;

The (30-years) average moisture deficits (MD30) at unsampled locations was predicted.

The fluctuation of Groundwater in terms of Mean Height Water Level (MHW) variable has an influence on MD30

There is a correlation between 30-years average moisture deficits (MD30) and MHW which is high about 0.7155

Co-kriging performed better compared to Ordinary and Regression Kriging, its RMSE was small compared to other two methods. The low RMSE indicate accurate predictions.

7. References

Kerebel, A., Cassidy, R., Philip, J., & Holden, M. (2010). The relationship between field soil water content variability and soil moisture deficit prediction from meteorological data, (August), 179–181.

McBRATNEY, A. B., & WEBSTER, R. (1983). Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science*, *34*(1), 137–162. https://doi.org/10.1111/j.1365-2389.1983.tb00820.x

Stein, A., Hoogerwerf, M., & Bouma, J. (1988). Use of soil-map delineations to improve (Co-)kriging of point data on moisture deficits. *Geoderma*, *43*(2–3), 163–177. https://doi.org/10.1016/0016-7061(88)90041-9

Webster, R., & Oliver, M. A. (2002). *Geostatistics for Environmental Scientists. Vadose Zone Journal* (Vol. 1). https://doi.org/10.2136/vzj2002.0321