Geostatistical Characterisation of seasonal Soil Moisture Variability

Edwige Mukundane
NOVA Information Management School
Campus de Campolide, 1070-312 Lisboa-Portugal
m20190649@novaims.unl.pt

Abstract;
The Variation of soil moisture matter over distinct Spatio-temporal scales and under changing climatic conditions has in recent times become a focus of interest for understanding underlying spatial processes responsible for soil water content variation. Geostatistical approaches provide ways of characterizing and quantifying soil moisture variability in space and they are increasingly used as inputs to hydrological and meteorological models. This project undertakes a geostatistical analysis of the 5cm soil moisture percentile parameter across to two seasons (spring and summer) for the year 2019 by employing several exploratory Spatial data analysis tools and interpolation methods to characterize soil moisture variability across New York in the United States of America (USA). Results from exploratory spatial data analysis provide an overview of the spatial variability. Results from interpolation techniques included interpolated or prediction surfaces of the soil moisture variable for the different seasons and are compared.

Keywords; Soil moisture, Spatial Variability, Geostatistics.

1. Introduction

Soil moisture is the water that is held in the spaces between soil particles measured as volumetric water content. The importance of soil moisture in climate dynamics has been carefully explored over the last few decades [1,4]. Soil moisture is a vital factor affecting energy interactions between the atmosphere and the earth's surface, and thus it is considered as a significant parameter by international scientific domains in characterizing global climate [3]. Comprehensive soil moisture information is necessary for various applications including drought prediction, weather forecasting, predicting flood conditions, agronomy, hydrology modeling, climate modeling to mention but a few [1,2,4]. Soil moisture conditions vary with time, space, soil texture, and soil type and relates greatly to the geological structure, land use and land cover, topography and geographic location and much or scarce amount of soil moisture could influence catastrophic occurrences and thus the need of methods to carefully understand soil moisture spatial processes in order to improve periodic monitoring [2,3]. Combinatorial spatial and temporal analysis using geostatistical approaches provides for forecasting values at unsampled locations by exploiting spatial associations between estimated and known sampled points while minimizing the variance assessment error [4]. This paper provides a geostatistical Spatio-temporal analysis on a seasonal scale for soil moisture variability in new york using the 5cm soil moisture percentile parameter (sm_5) which is the volumetric water content measured at 5cm depth. Results give a characterization of soil-moisture dynamics in the spring and summer seasons that would be applied for further analysis or decision making.

1.1 Objective

To provide a characterization of soil-moisture variability across the Spring and Summer season for the year 2019 in New York.

2. Study Region and Data

New York is a city in the USA with location coordinates of 40° 43' 50.1960'' N and 73° 56' 6.8712'' W covering an area of 783.8km2. The NewYork soil moisture data was obtained from the national soil moisture network of the (USA), specifically for the spring and summer season. Data extracted was for the 5cm of the soil-moisture percentile parameter (sm_5). This was renamed to indicate sm_SPG and

sm_SMR to represent the sm_5 parameter for the spring and summer season respectively. The daily soil moisture was aggregated to obtain average values for each season. The data can be found on the following link;
http://nationalsoilmoisture.com/test/VWC_QAQC/Daily%20soil%20moisture%20percentile/

Figure 1 showing study area (Newyork)

3. Methodology

The methods employed are classified into three parts below;

- Exploratory data analysis; this is a collection of descriptive methods for discovering patterns, detecting unusual or interesting elements in a dataset. This is usually the first step, and in this study, we used descriptive statistics which gives a summary overview of the variable.

- Exploratory spatial data analysis (ESDA) tools are an extension of exploratory data analysis (EDA) since they provide for spatial properties existing in a dataset. Tools applied include Dataposing; which shows spatial distributions by using a scale of graduated colors, the Regional histogram which enables graphical visualization of the dataset, Voronoi maps that facilitate exploration of local variability, the local Moran´s I statistic that enables identification of spatial outliers and statistically significant local clusters and the Global Moran´s I statistic that helps investigate spatial autocorrelation in the study area.

- Interpolation techniques followed such as; Inverse distance weighting that uses known values surrounding an unsampled location to predict unknown locations. The other technique is kriging that produces a prediction surface from sampled points of a continuous variable.

4. Results

4.1 Descriptive Statistics;

Results from descriptive statistics computation help to understand patterns in the sm_SPG and sm_SMR variables;

sm_SPG	
Mean	0.805714286
Standard Error	0.00857566
Median	0.8
Standard Deviation	0.096261548
Sample Variance	0.009266286
Kurtosis	1.020816708
Skewness	-0.738748746
Range	0.55
Minimum	0.44
Maximum	0.99
Sum	101.52
Count	126

sm_SMR	
Mean	0.270873016
Standard Error	0.012716703
Median	0.26
Standard Deviation	0.142744638
Sample Variance	0.020376032
Kurtosis	1.950757096
Skewness	1.023672436
Range	0.84
Minimum	0.03
Maximum	0.87
Sum	34.13
Count	126

Figure 2 showing descriptive statistics results for the sm-SPG on left and sm-SMR on right.

Descriptive statistics of ***sm_SPG*** *and* ***sm_SMR*** help us deduce the following;

- The variables were measured at 126 sensor locations across New York.
- The mean is 0.8057cm^3, the median is 0.8 cm^3 and the mean is 0.2708cm^3, the median is 0.26cm^3 for the Sm-SPG and Sm-SMR variable respectively.
- The typical deviation from the mean value 0.0962 and 0.1427 for the Sm-SPG and Sm-SMR variable respectively
- The Maximum values obtained for the Sm-SPG and Sm-SMR variable are 0.99cm^3 and 0.87cm^3 respectively while the minimum value is 0.44cm^3 and 0.3cm^3 respectively.
- 50% of the sampled points have lower than 0.8cm^3 and 0.26cm^3 of the Sm-SPG and Sm-SMR variable respectively.

4.2 Data posting

This gave an indication of the overall spatial distribution and trend of the Sm-SPG and Sm-SMR variable across the study area.

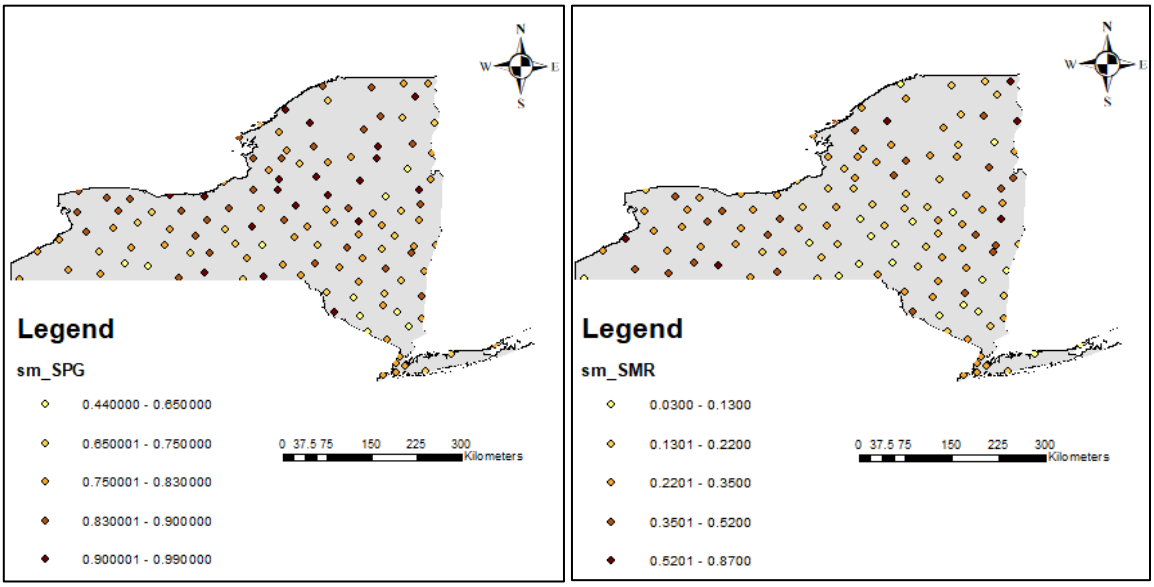

Figure 3 shows data posting maps for the sm-SPG on left and sm-SMR on right.

Figure 3 shows that the Sm-SPG variable has short instances of fair homogeneity in the west and south parts were the low values are located. The south has the lowest values ranging from 0.44cm^3 to 0.83cm^3. High values are concentrated at the upper center and north of the study area. The sm-SMR variable shows a relatively homogeneous distribution at the center of the study area having low volume values in the range of 0.03cm^3 to 0.35cm^3 except for four locations that have high values. Also, the sm-SMR has the highest values in the east and western part of the study region. There is no apparent trend over the study area in both seasons. In both seasons, the lowest variable volumes are in the southern tail of the study region.

4.3 Regional Histogram

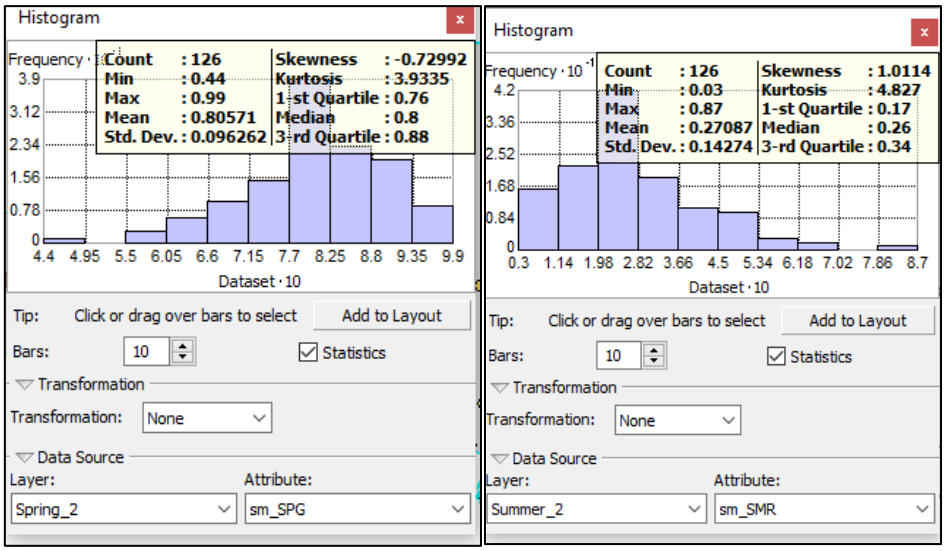

Figure 4 regional histogram for the sm-SPG on left and sm-SMR on right.

The descriptive statistics are identical to those attained in descriptive statistics. The sm-SPG histogram shows negative asymmetry, meaning that there are a few points with low sm-SPG values. While the sm-SMR shows positive asymmetry. All instances have extreme that could be possible outliers. The spatial distribution is analyzed in figure 5 below;

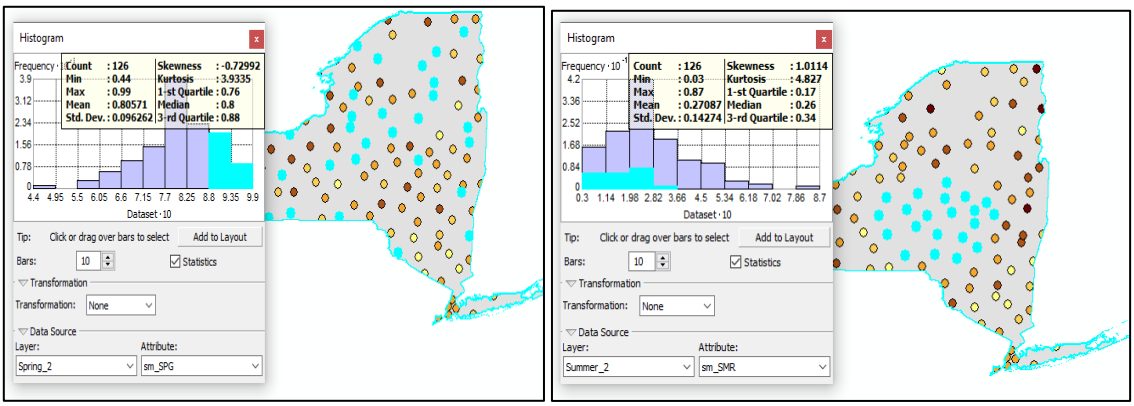

Figure 5 showing the investigation of spatial patterns and regimes across the study area for the sm-SPG on left and sm-SMR on right.

Visualization of spatial distribution was carried out. In both instances, points on map and histogram bars were sampled and locations selected could be seen spread out across the study area or in the histogram. We can deduce that the variables exhibit more spatial heterogeneity than homogeneity in all seasons across the study region with no outstanding spatial regimes in the study area.

4.4 Voronoi Maps

Simple and cluster Voronoi maps were created to investigate local variability. The Simple Voronoi map was creat to help visualize spatial patterns while the Cluster Voronoi map was to indicate the location of possible outliers.

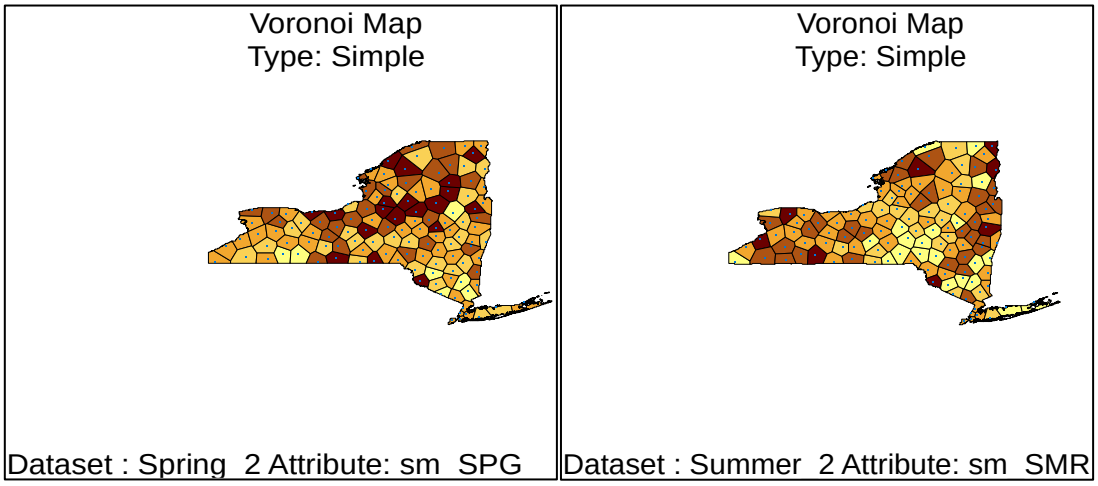

Figure 6 showing simple Voronoi Map for the sm-SPG on left and sm-SMR on right.

The simple Voronoi indicates similar situations as data posting. In both maps, there on major directional continuity pattern observed hence no anisotropy of the variable in the study area.

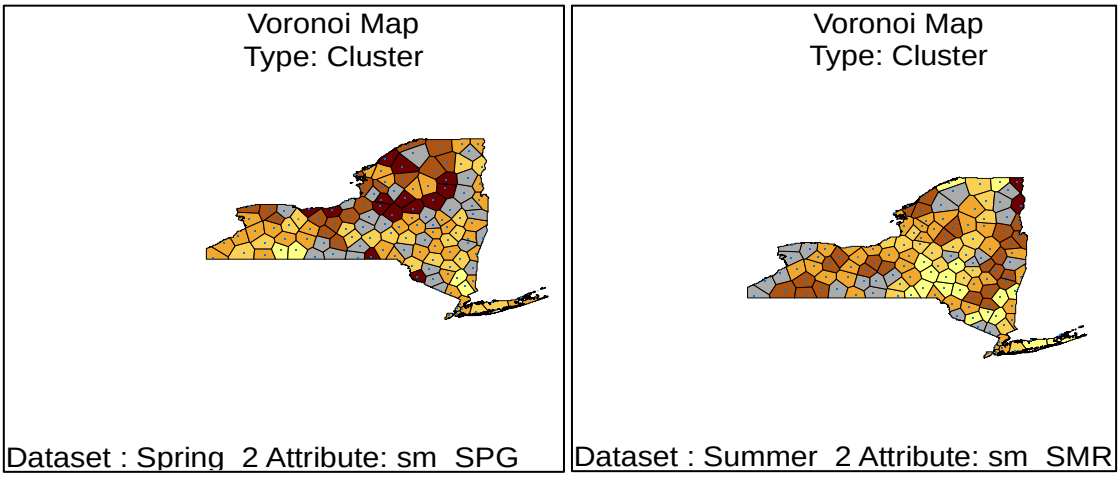

Figure 7 showing cluster Voronoi Map for the sm-SPG on left and sm-SMR on right.

The cluster map shows 29 polygons and 23 polygons for the spring and summer seasons respectively that imply possible spatial outliers. The spatial outliers are investigated further using the local morans I statistic.

4.5 Local Moran´s 1 static

Before applying the Anselin local morans I tool, threshold distances were determined by applying the Incremental spatial autocorrelation tool in ArcGIS.

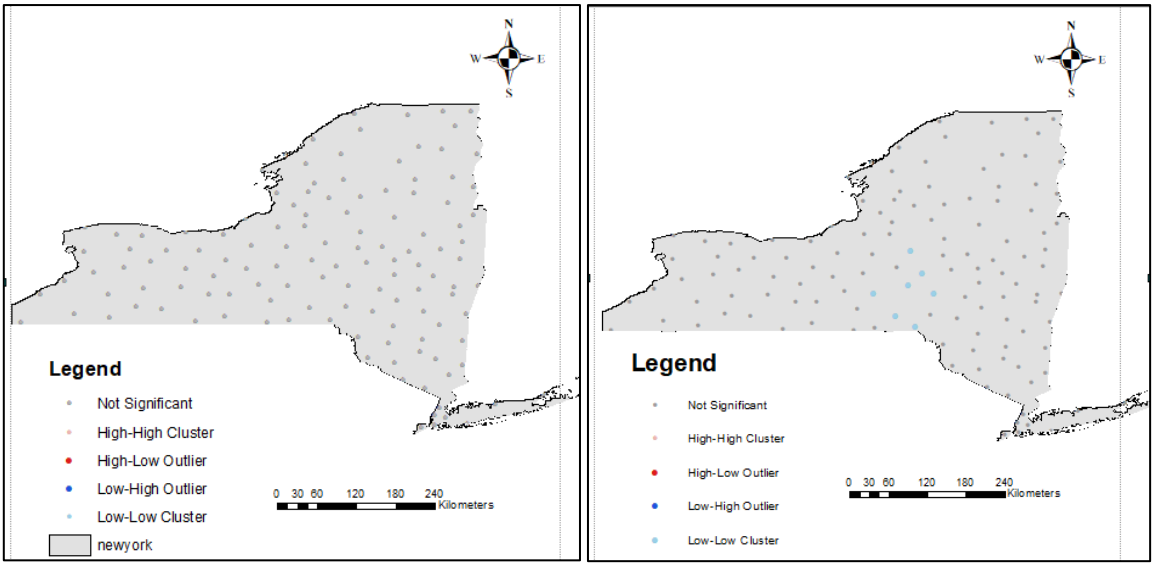

Figure 8 shows Local Moran's I statistic results for the sm-SPG on left and sm-SMR on right.

The local morans I maps show only statistically none-significant values of the sm-SPG variable for the spring season while for the sm-SMR we can see a cluster of low - low values in the central region of the study area. We can also see that for both seasons, we don't have any resulting spatial outliers and thus, spatial outliers assumed in the previous exploration findings can be disregarded. With most of the variable values seen as not statistically significant, this would imply that spatial autocorrelation is not statistically significant, this is investigated further using the Global Morans I statistic.

4.6 Global Moran´s 1 static

The figures in 9 below were extracted from the spatial autocorrelation report and global moran´s I summary;

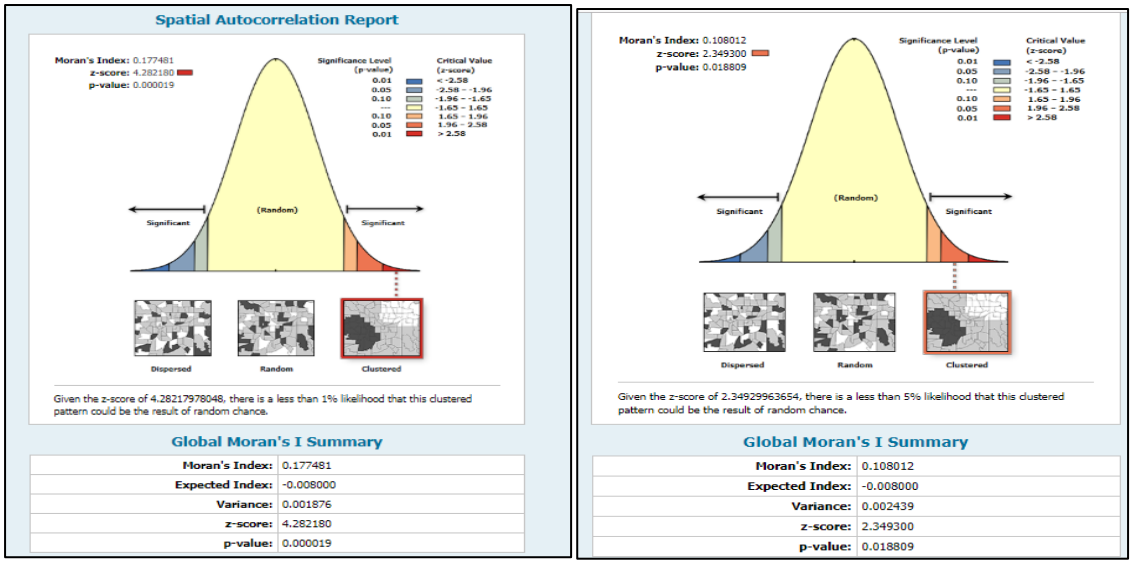

Figure 9 showing Global Moran's I statistic results for the sm-SPG on left and sm-SMR on right.

Results for the Global Moran´s I Statistic show statistically significant Moran´s index is equal to 0.177481 and 0.108012 for the sm-SPG and sm-SMR Variables respectively. P-value = 0.000019 and 0.018809 for sm-SPG and sm-SMR respectively is lesser than the accepted 0.05 alpha level which shows that the spatial autocorrelation is statistically significant in both seasons. The resultant Z-score = 4.282180 and 2349300 for the sm-SPG and sm-SMR respectively that are greater than +1.96 indicate the presence of a clustered pattern for the soil moisture variable in both seasons. A p-value can

represent the probability of having the observed spatial pattern to be random, therefore, a P-value = =0.000019 and 0.018809 for sm-SPG and sm-SMR respectively that is lower than the accepted alpha level of 0.05 is considered low meaning that there is a low probability that the observed spatial pattern is random which justifies the existence of a global cluster of the variables in the study area. With the resultant p-values, I may conclude that there is a 95% confidence level to reject the complete spatial randomness hypothesis and that the spatial distribution of the sm-SPG and sm-SMR is not created by random spatial process. These results indicate that there a high degree of spatial dependence of the soil moisture parameter across New York in both the spring and summer season.

4.7 Inverse distance weighting (IDW)

The IDW interpolation/ prediction surface was achieved by determining appropriate neighborhood specifications for variables. The major and minor semi-axes were chosen to be 600,000m. This distance was attained by calculating the highest distance across the whole study area. Both default and trail cross-validations gave great mean prediction errors closest to zero. The row highlighted green shows the final cross-validation with a mean prediction error of 0.000052 and 0.00146 for the sm-SPG and sm-SMR Variables respectively. The first row was computed using default values.

Maximum Neighbors	Minimum Neighbors	Sector type	Major semi-axis	Minor semi-axis	Mean Error	Root Mean Square Error
15	10	1	2.086538	2.086538	-0.00023	0.0940
15	10	4	600000	600000	0.00075	0.0924
10	5	8	600000	600000	0.00099	0.0921
10	5	1	600000	600000	0.000052	0.0951

*Table 1 showing cross-validation trials and results of IDW-spring (*sm-SPG)

Maximum Neighbours	Minimum Neighbours	Sector type	Major semi-axis	Minor semi-axis	Mean Error	Root Mean Square Error
15	10	1	2.086538	2.086538	0.00181	0.1326
10	5	4	600000	600000	0.00177	0.1332
10	5	1	600000	600000	0.00178	0.1342
15	10	8	600000	600000	0.00146	0.1332

*Table 2 showing cross-validation trials and results of IDW-summer (*sm-SMR)

The prediction and contour line maps in figures 18 below help to show how different areas across the study region are influenced by the variability of the sm-SPG and sm-SMR variable. The interpolated IDW surfaces give an estimation of the sm-SPG and sm-SMR variable at unknown locations where they were not measured.

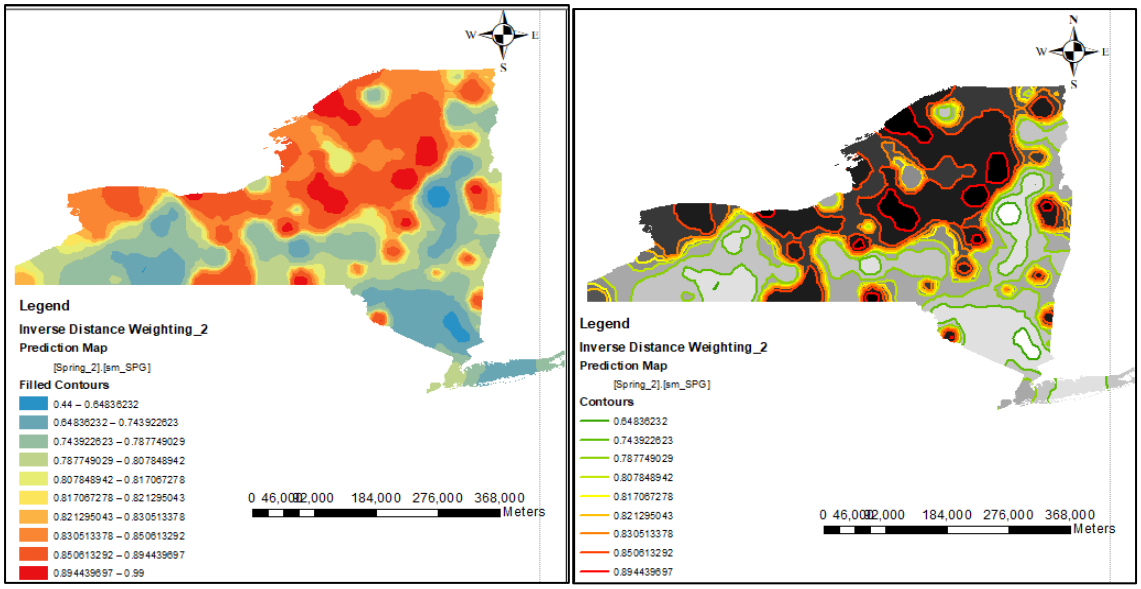

Figure 10 showing the prediction pattern map on the left and the contour line map on right for the Sm-SPG variable -Spring

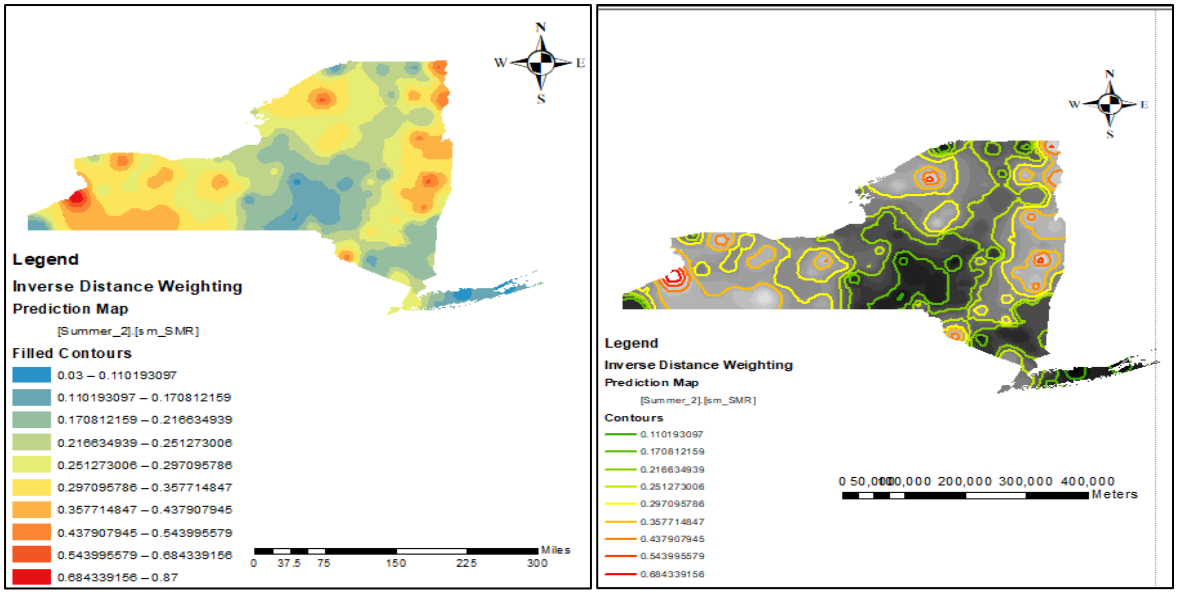

Figure 11 showing the prediction pattern map on the left and the contour line map on right for the Sm-SMR variable -Summer

Looking at figure 10 showing the interpolated surface, it can be seen that a major part of the north is dominated by high values of the Sm-SPG variable. This is seen in the isoline map that shows contour lines with high values ranging from (0.81706cm^3 to 0.8944cm^3) dominating a larger part of the northern region with a few instances in other areas. In figure 10, the situation of the Sm-SMR variable is somewhat the opposite of the Sm-SPG variable. Locations that tend to have high moisture volumes in the spring are seen having the lowest values in the summer especially near the upper center of the study area. In both seasons the lowest soil moisture volume locations are in the southern region especially the tail. The prediction maps and contour line maps in both seasons show us that there is no apparent trend across the study area.

4.8 Kriging

Geostatistical methods= Kriging/Cokriging were followed closely in the ArcGIS geostatistical wizard in order to achieve the empirical semivariogram model using the kriging type as Ordinary. Construction of the semivariogram model involved both analyzing default parameters and determining parameters for

the best-fitting model. The model type used is exponential. The model was analyzed using both default parameters and estimated parameters;

Nugget	False
Model type	Exponential
Major Range	0.5431
Anisotropy	False
Partial Sill = Total Sill	0.00961
Lag size	0.0654
Number of Lags	12

Table 3 showing default parameters for the empirical Semivariogram for Sm-SPG variable -Spring season

Nugget	False
Model type	Exponential
Major Range	1.157
Anisotropy	False
Partial Sill = Total Sill	0.0195
Lag size	0.2860
Number of Lags	12

Table 4 showing default parameters for the empirical Semivariogram for Sm-SMR variable -Summer season.

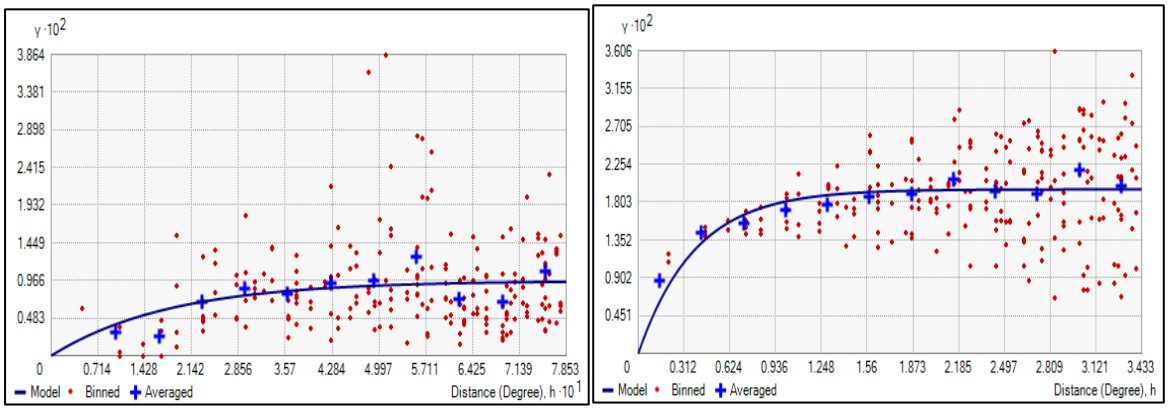

Figure 12 showing the default view for the Semivariogram for the sm-SPG on left and sm-SMR on right.

The default parameters did not fit well to the binned and averaged values, attempts were made to estimate parameters that best fit the model to the averaged values. Final parameters of best variogram are presented for both seasonal datasets in tables in table 2 below;

Nugget	False
Model type	Exponential
Major Range	0.5447
Anisotropy	False
Partial Sill = Total Sill	0.00958
Lag size	0.260
Number of Lags	10
Empirical Semi variogram values	Averaged.

Table 5 showing the best-fit model parameters for the empirical Semivariogram for Sm-SPG variable -Spring season

Nugget	False
Model type	Exponential
Major Range	0.80256
Anisotropy	False
Partial Sill = Total Sill	0.019185

Lag size	0.386
Number of Lags	10
Empirical Semi variogram values	Averaged.

Table 6 showing the best-fit model parameters for the empirical Semivariogram for Sm-SMR variable -Summer season.

Knowing that anisotropy is absent in the sm-SPG and sm-SMR variables, the omnidirectional experimental variogram was fitted for Isotropic pattern of the Sm-SPG and Sm-SMR variables.

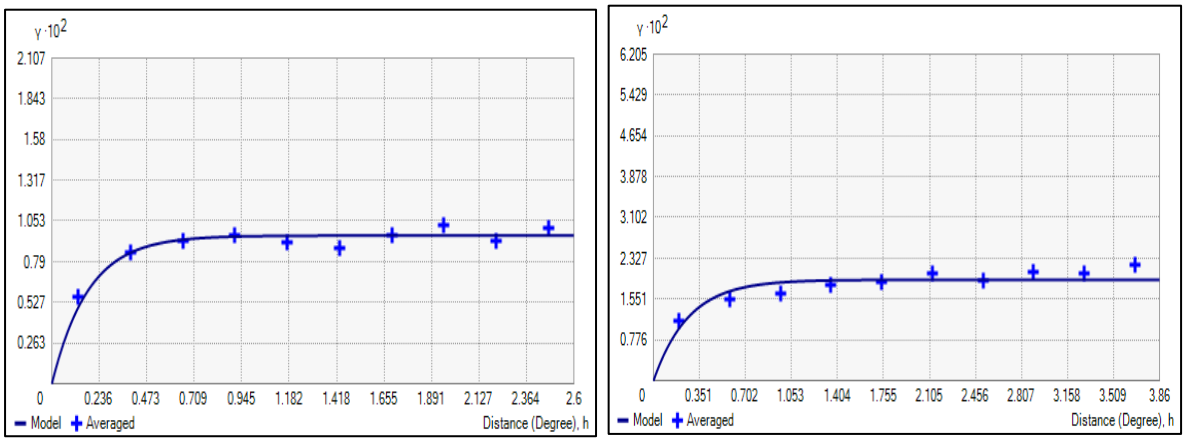

Figure 13showing the best-fit model for the Semivariogram for the sm-SPG on left and sm-SMR on right.

The nugget was set to be zero, thus the model originates from the zero-point origin of the experimental variogram graph in both figures. It is not clear to me whether the soil moisture data has errors, on assumption that the soil moisture sensors are periodically calibrated and that sensors measure and record values electronically I assumed the absence of measurement errors in the data.

The partial sill implied the total sill since the nugget was set to Zero. This gave the sill to be 0.00958and 0.019185 for the sm-SPG and sm-SMR Variables respectively. Looking at figure 13 the sill is the value at which the semivariogram model achieves the range.

The range was adjusted repeatedly until the best fit model was attained. The best range parameters obtained were 0.5447 and 0.80256 for the sm-SPG and sm-SMR Variables respectively.

The final lag sizes obtained were sufficient to get representative averages for the variogram values. Lag sizes obtained were 0.26 and 0.386 for the f the sm-SPG and sm-SMR Variables respectively with the number of lags=10.

Defining the neighborhood search parameters was done to produce prediction surfaces. The major and minor semi-axes were chosen to be 600,000m. Both default and trail cross-validations gave great results and I had unbiased prediction errors of mean prediction error nearest to zero and the root mean square standardization error nearest to 1. All trails gave great results and the most precise result was chosen.

Sector-type	Maximum-neighbors	Minimum-Neighbours	Major semi-axis	Minor semi-axis	Mean Error	Root Mean Square Error	Root Mean Square Standardised Error
4 with 45^0 offsets	5	2	0.80255	0.80255	0.00185	0.1347	1.0646
4 with 45^0 offset	10	5	600000	600000	0.00171	0.1340	1.0679
1	15	5	600000	600000	0.00154	0.1334	1.0614
1	10	5	600000	600000	0.00074	0.1339	1.06251

Table 7 showing cross-validation trials for neighborhood definitions and prediction error statistics for the sm-SPG variable

Sector-type	Maximum-neighbors	Minimum-Neighbours	Major semi-axis	Minor semi-axis	Mean Error	Root Mean Square Error	Root Mean Square Standardised Error
4 with 45^0offset	5	2	0.5447	0.5447	0.00093	0.0965	1.0144
1	10	5	600000	600000	-0.00152	0.0953	1.0075
4	15	5	600000	600000	0.00016	0.0929	0.9945
4 with 45^0offset	5	2	600000	600000	0.00043	0.0942	1.0012

Table 8 showing cross-validation trials for neighborhood definitions and prediction error statistics sm-SMR Variables.

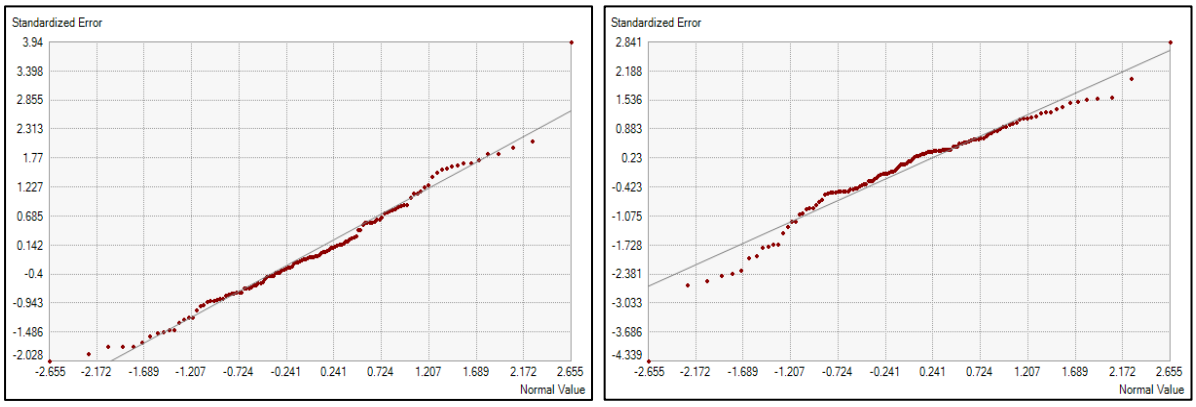

Figure 14 showing cross-validation results with normal QQ plot selected for the sm-SPG on left and sm-SMR on right.

The Normal QQ plot was used to give a graphical indication of the measured values versus the predicted values. Figure 14 shows that there are a few values marginally falling below or above the line, but most points are seen falling close to the straight line informing that the prediction errors for sm-SPG and sm-SMR are near to being normally distributed. Final cross-validation results were successively used to derive prediction surfaces and raster versions produced as well as shown in figure15 below;

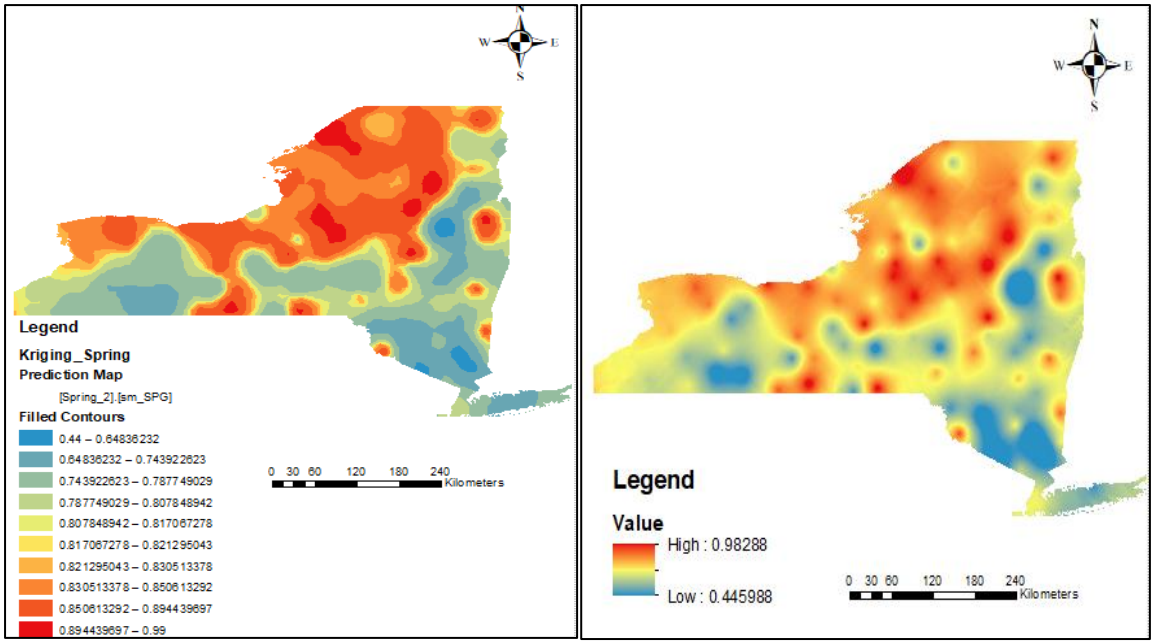

Figure 15 showing ordinary kriging prediction map for the Sm-SPG variable on left with the raster version on the right

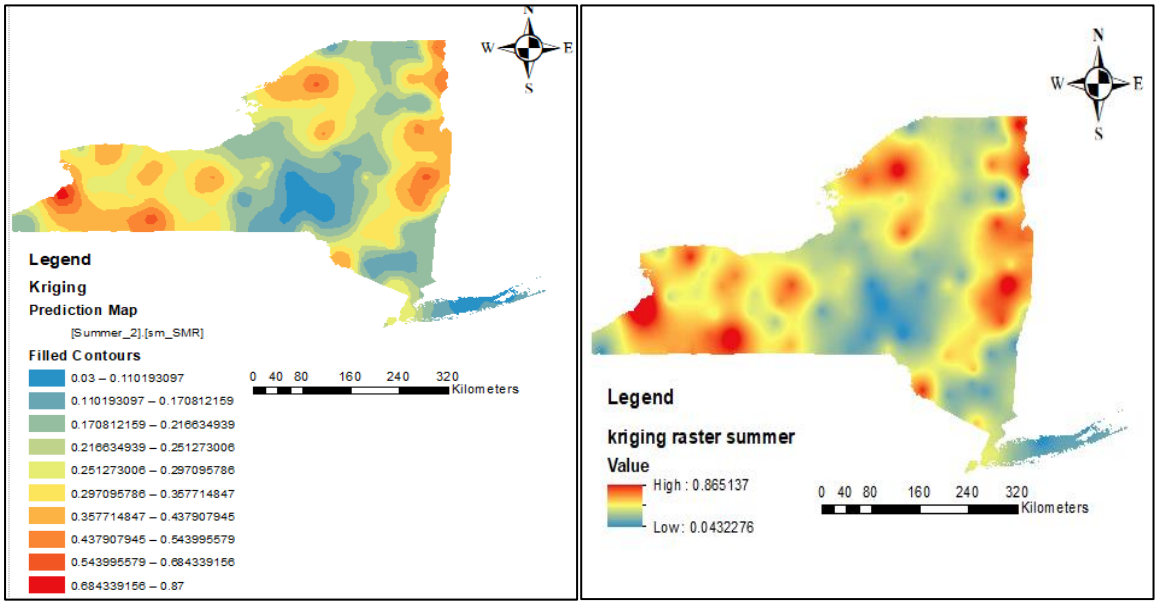

Figure 16 showing ordinary kriging prediction map for the Sm-SMR variable on left with the raster version on the right

The interpolated prediction surfaces created from the kriging method are closely similar to the prediction surfaces obtained from the inverse distance weighting interpolation method. As analyzed earlier, it can be seen from figure 15 that a larger part of the northern region in New York exhibits the highest Sm-SPG variable in spring having a few other locations. The summer season in figure 16 shows that New York experiences its highest sm-SMR values in the west, east and some few parts of the north. Generally, both seasons have their lowest soil moisture volume values in the southern part of the study area. The ordinary kriging prediction maps were used to derive prediction standard error maps.

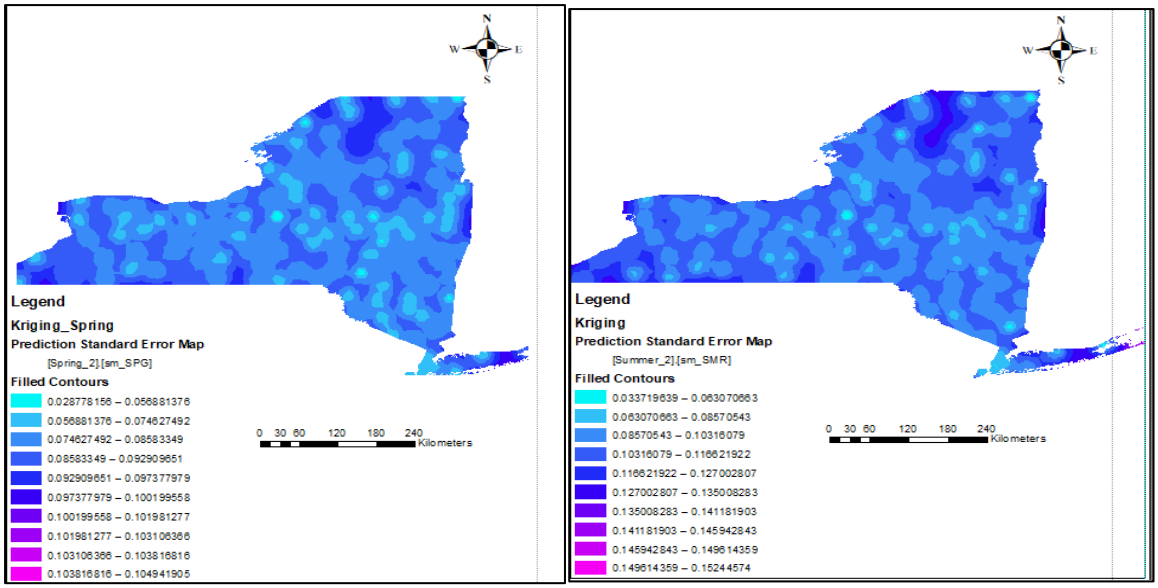

Figure 16 showing ordinary kriging prediction map for the Sm-SMR variable on left with the raster version on the right

The prediction standard error maps give measures of errors of locations depending on their distance from known sm-SPG and sm-SMR values. Locations near sm-SPG and sm-SMR sample points have lower errors compared to those far away. In both seasons, areas having high prediction standard errors are in the southern part especially the tail since it had fewer sample values

5. Conclusion

Understanding the spatial distribution and detailed mapping of soil properties such as soil moisture at a large scale is very valuable for several soil conservation and ecological modeling efforts. Geostatistical approaches were applied to both analyze the spatial distribution of the 5cm soil moisture percentile and produce estimated prediction surfaces for the variable for both summer and spring seasons. The inverse distance weighting and ordinary kriging interpolation surfaces provided closely related results that enabled complete visualization of the 5cm soil moisture percentile (sm-SPG and sm-SMR) in the two seasons that enabled deriving characteristics of the variable. In as much as the interpolation results using the two approaches were closely similar, cross-validation approaches of semivariogram fitting techniques obtained by Ordinary kriging seemed to provide a more accurate spatial extrapolation of the variable than the IDW method that assumes the mean of the known values at any unobserved location. The outcomes of this work provide an understanding and visualization of soil moisture dynamics across different seasonal temporal scales and approaches used under this study can be used to improve soil-water monitoring plans and other modeling activities.

6. References

1. Vinit Sehgala, Venkataramana Sridhara (2018) Watershed-Scale Retrospective Drought Analysis and Seasonal Forecasting Using Multi-Layer, High-Resolution Simulated Soil Moisture for Southeastern U.S
2. Célia Regina Grego,Sidney Rosa Vieira, Aline Maria Antonio,Simone Cristina Della Rosa (2006), Geostatistical Analysis For Soil Moisture Content Under The No Tillage Cropping System
3. Mehmet Zeki İmamoğlu And Elif Sertel (2016) Analysis of Different Interpolation Methods for Soil Moisture Mapping Using Field Measurements and Remotely Sensed Data
4. Gouri Sankar Bhuniaa,Pravat Kumar Shitb,Rabindranath Chattopadhyay (2018) Assessment Of Spatial Variability Of Soil Properties Using Geostatistical Approach Of Lateritic Soil (West Bengal, India).