Lab Report

Title: <Lab 4 Report >
Notice: Dr. Bryan Runck
Author: Rongxuan Zhu
Date: 11.30.2021

Project Repository: < *https://github.com/zhux0474/GIS5571/upload/main/Lab4* >
Google Drive Link: *<N/A>*
Time Spent: *<24 hours >*

Abstract
Lab 4 instructs me to build an ETL in Jupyter notebook to extract the temperature data of all the stations in Minnesota from the NDAWN website for the last 30 days, to map all the station points with their average monthly temperature in ArcGIS Pro, and to perform three different interpolation methods to create interpolated temperature maps for the highs and lows of the last 30 days from NDAWN in real-time.

Problem Statement
Lab 4 wants to compare and contrast three interpolation methods and create three maps for the average monthly temperature of the past 30 days for all the station's points in Minnesota using the data extracted from the NDAWN website in ETL.
Table 1 shows the broken-down elements of the problem statement.

Table 1. < Table of Elements Break Down from Problem Statement >

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	IDW	Inverse Distance Weighted	N/A	N/A	N/A	N/A
2	Kriging	Empirical Bayesian kriging	N/A	N/A	N/A	N/A
3	LPI	Local Polynomial Interpolation	N/A	N/A	N/A	N/A
4	Jupyter Notebook	Tool to build ETL pipeline	N/A	N/A	N/A	N/A
5	NDAWN Dataset	Average monthly temperature dataset for interpolation			NDAWN	

Input Data
The input data is extracted from NDAWN website in the ETL built in Jupyter Notebook
Table 2. <Table of Input Data Downloaded from Websites with Links>

#	Title	Purpose in Analysis	Link to Source
1	NDAWN	Average monthly temperature dataset for interpolation	NDAWN

Methods

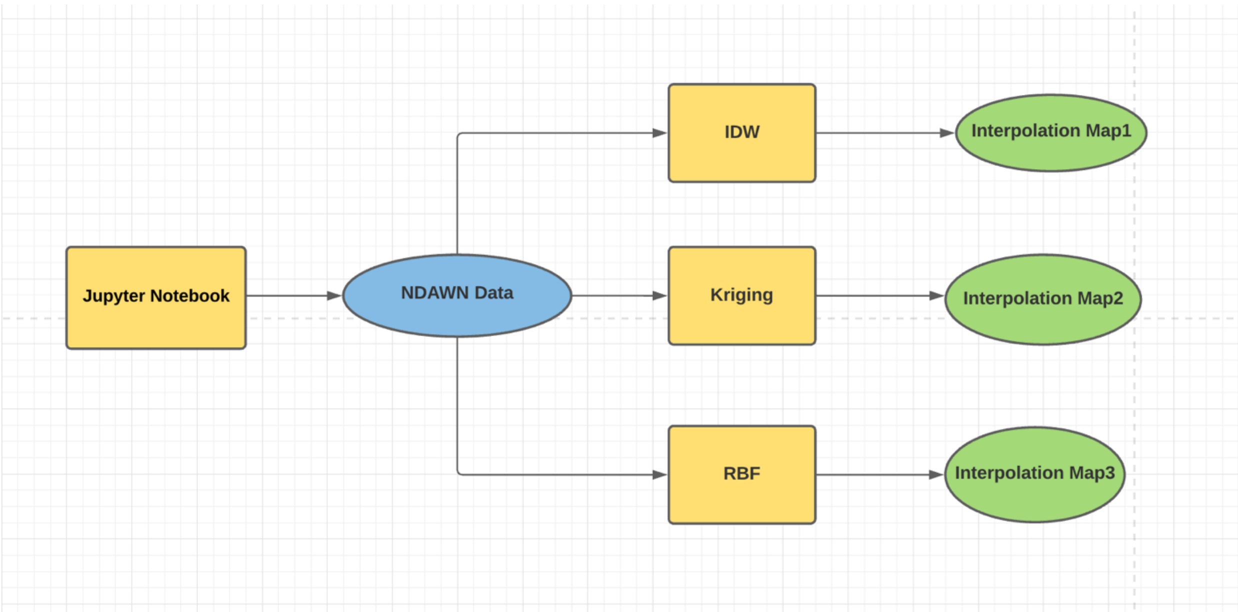

Figure 1 – Data Flow Diagram

The general methods are summarized into Data Flow Diagram(Figure 1) and the code can be found under the notebook folder on GitHub.

Python code in Jupyter Notebook is used to extract average air temperature for the last 30 days in real-time for all the stations in Minnesota from the NDAWN website. The "groupedby" and "agg" function is used to calculate the average monthly air temperature for each station so that one station eventually has one average air temperature in the data file. The datetime library is used to get the date of today and the date of 30 days ago to request the real time data for the last 30 days. The CSV file is generated and imported into ArcGIS Pro for interpolation. The CSV file is first converted to a table in ArcPro and then use the display X, Y point function to transform the data in CSV into point features on the map. Interpolation methods of IDW (Inverse Distance Weighted), Kriging, RBF (radial basis functions) are applied separately to the NDAWN data with average air temperature is set to z value and the rest of the parameters are kept as the default setting to create three interpolation maps. The three interpolation methods and maps are compared and contrasted in the discussion section.

Results

Figures 2,3, and 4 are the interpolation maps generated in ArcGIS Pro.

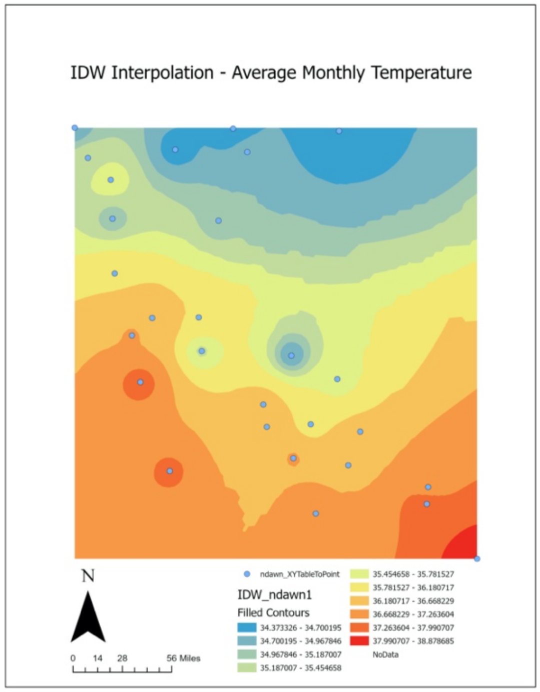

Figure 2 – IDW (Inverse Distance Weighted) Interpolation Map of Average Monthly Temperature

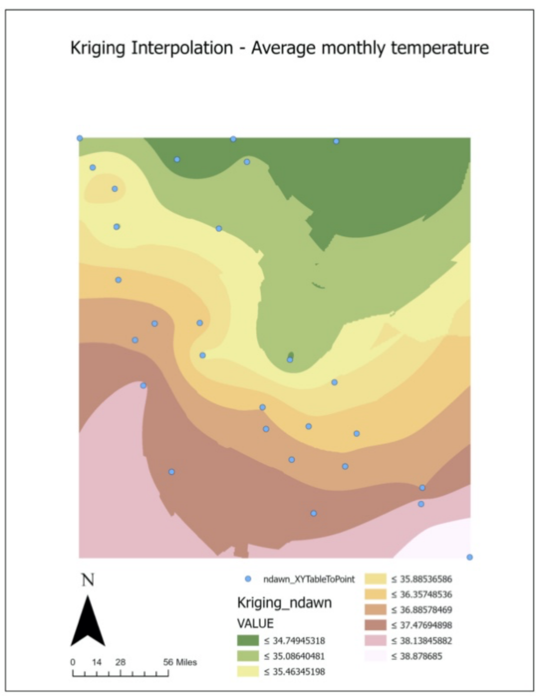

Figure 3 – Kriging Interpolation Map of Average Monthly Temperature

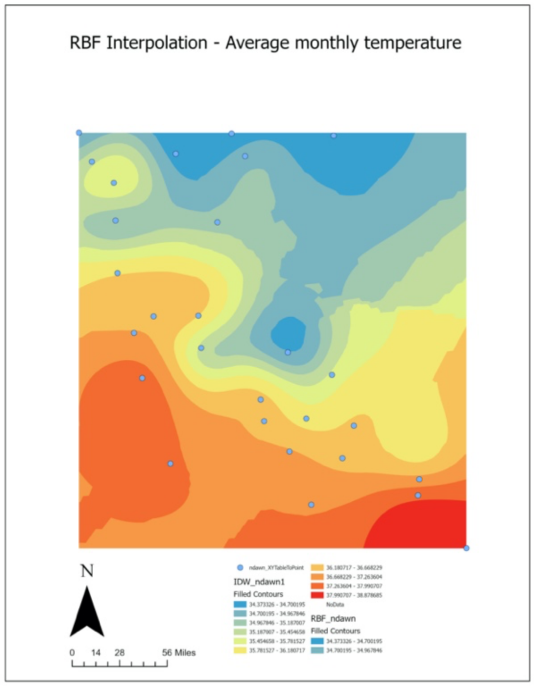

Figure 4 – RBF (radial basis functions) Interpolation Map of Average Monthly Temperature

Results Verification

The results are verified by visual comparison between the three interpolation maps. These three maps show different but similar trends in average monthly air temperature in each station in Minnesota. The highs are located at the south while the lows are located at the north portion of the map with some highs and lows interacting in the middle portion.

Discussion and Conclusion

Interpolation methods generate an estimated surface from a scattered set of points with z values (in this case is average monthly temperature).

IDW(Figure 2), a deterministic method(it is directly based on the surrounding measured value), assumes that points that are closer to one another are more similar than those that are farther away from each other. IDW uses the surrounded measured values to predict a value for the unmeasured location. The measured values that are the closest to the unmeasured location have more influence on the predicted value than those values that are farther away by giving greater weights to the points that are the closest and the weights decrease as a function of distance.

Kriging (Figure 3), a stochastic method, is based on statistical models which include autocorrelation of statistical relationships among the measured points so that it can produce a prediction surface with some measure of certainty or accuracy of the predictions. It assumes that the distance or direction between sample points contains spatial correlation, and it uses a mathematical function to determine the output value of each location. This method is the most appropriate interpolation method when you know there is a spatially correlated distance in the data which is used often in soil science and geology.

RBF (Figure 4), another deterministic method, is a function whose value depends only on the distance from the origin. This method is similar to fitting a rubber membrane through the sample values and it minimizes the total curvature of the surface. The basis function determines how the rubber membrane fits between the values. RBF can predict values that are above the maximum measure value and below the minimum measured values in the sample data (while IDW cannot). RBF utilize global neighborhood to predict value which makes it suitable for the prediction of a continuous smooth surface (Kriging uses a local neighborhood)

Figure 4 shows more continuous and detailly layered values than figures 2 and 3 where the points with surrounding values look slightly more fragmented. The literature recommends radial basis functions (RBF) for best performing interpolating temperature data at the annual scale (in terms of accuracy and unbiasedness) which provides a general mathematical tool that can identify a continuous surface representing the variable behavior based on the comparisons of nine different methods (include IDW and Kriging) for spatial interpolation of temperature over a Mediterranean island Sicily. This study shows that IDW has the worst performances while the kriging has intermediate skills between IDW and RBF. (Piazza et al., 2015). Another study also shows that kriging-exponential and kriging-spherical interpolation methods have higher accuracy when compared to IDW interpolation methods. (Cao et al., 2009)

References

ArcMap. How inverse distance weighted interpolation works-ArcMap | Documentation. (n.d.). Retrieved November 20, 2021, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm.

ArcMap. How radial basis functions work-ArcMap | Documentation. (n.d.). Retrieved November 20, 2021, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-radial-basis-functions-work.htm.

Cao, W., Hu, J. X., & Yu, X. (2009). A study on temperature interpolation methods based on GIS. *2009 17th International Conference on Geoinformatics*. https://doi.org/10.1109/geoinformatics.2009.5293422

Kriging (spatial analyst). Kriging (Spatial Analyst)-ArcGIS Pro | Documentation. (n.d.). Retrieved November 20, 2021, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/kriging.htm.

Piazza, A., Conti, F., Viola, F., Eccel, E., & Noto, L. (2015). Comparative analysis of spatial interpolation methods in the Mediterranean area: Application to temperature in Sicily. *Water*, *7*(12), 1866–1888. https://doi.org/10.3390/w7051866

Self-score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	28
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	24
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	28
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	20
		100	100