

Lab Report

Title: Lab 4 - Interpolation
Notice: Dr. Bryan Runck
Author: Mohsen Ahmadkhani
Date: 12/05/2021

Project Repository: <https://github.com/mohsen-gis/GIS5571.git>

Google Drive Link: -

Time Spent: 15 hours

Abstract

In this lab assignment, I demonstrated the extract, transform, and load (ETL) of temperature data for the last 30 days from the NDAWN dataset. I also implemented four types of interpolation analysis, namely, inverse distance weighted (IDW), global Kriging, ordinary Kriging, and Polynomial interpolation.

Problem Statement

In this lab, I will demonstrate the implementation of an ETL method for NDAWN temperature data, aggregating them based on the station name by averaging the daily average temperature of each station. As a part of this lab, I will justify the methods I chose and explanation for them shortly.

Table 1. The list of required data sets for the proposed study.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Meteorological data	A set of sites collecting weather data	Point data (Lat/Lng)	Temperature, humidity, etc	NDAWN	Download using REST API as a pandas df and convert it to GeoPandas df

Input Data

To demonstrate the interpolation methods, I downloaded temperature data from NDAWN API to satisfy the requirements of the lab assignment. Table 2 shows this dataset. This dataset is being stored as CSV format as a part of the ETL. The dataset itself contains latitude and longitude so I used them to build a geometry column and make the data spatial.

Table 2. Input data description.

#	Title	Purpose in Analysis	Link to Source
1	Temperature data	For demonstration of interpolation analysis	API

Methods

The analyses in this project are divided in three groups. First, for downloading data I used the `urllib.parse` module to set the parameters for creating the API calls. I also used `requests` module to fetch data from the API provider sites. Second, processing the data and building the interpolation models. To do this, I used `ArcPy` module and the cartographic modules in the ArcGIS Pro GUI environment. The interpolation methods included in this assignment are IDW, universal and ordinary kriging, and local polynomial interpolation methods. Lastly, to visualize the data and results I used `Folium` and `matplotlib` packages. The data flow diagram of this project is depicted in figure 1. It is noteworthy that the justification for selecting the interpolation models is discussed in the discussion section.

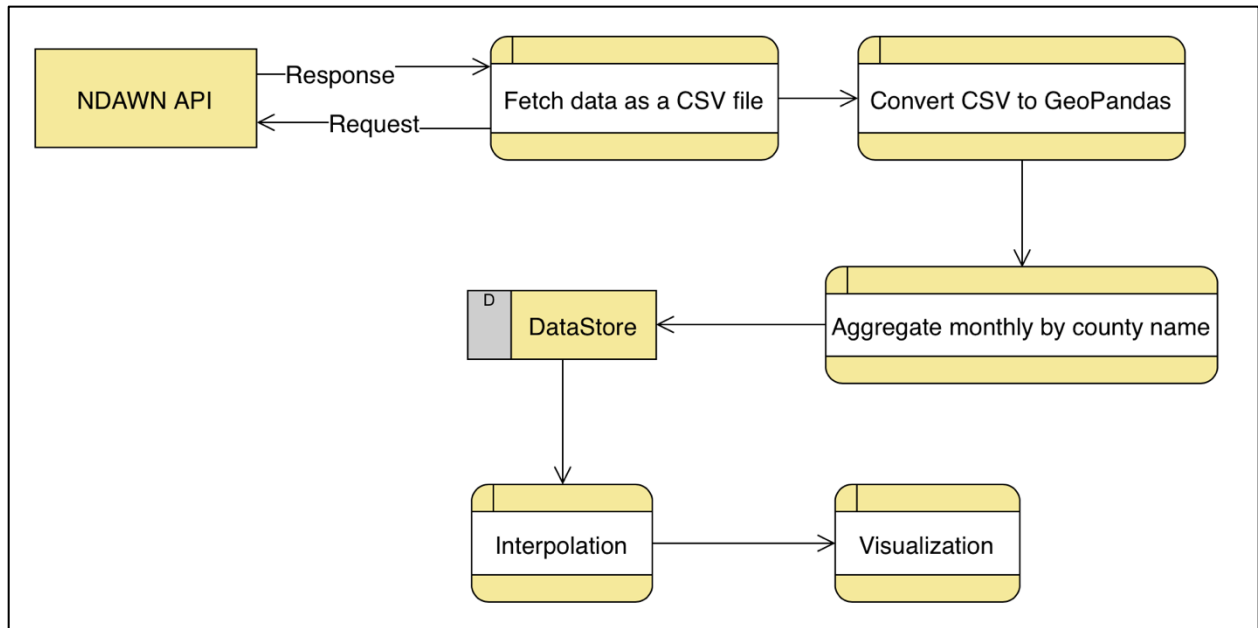


Figure 1. Data flow diagram.

Results

After downloading the data from the NDAWN API [1], and running the selected interpolation methods, the results are depicted in figures 2 to 5. Generally, from the first glance at all figures, it is evident that there is a gradual incremental pattern for the average monthly temperature by decreasing the latitude in the state of North Dakota (and partly Minnesota). This pattern was not unexpected since the higher the value of the latitude is, the closer to the north pole and consequently lower average temperature would be expected. This decremental trend is predicted as a set of roughly parallel layers of temperature that is visible in all generated raster layers. However, the result of IDW method shown in figure 2, is slightly different than the rest of the results and it is due to the implicit assumption of spatial autocorrelation for temperature embedded in the IDW method. Unlike IDW, the result of local polynomial interpolation depicted in figure 5 is quite different than IDW and is having smoother layered pattern. The interpolated raster depicted in figures 3 and 4 are more similar as they are the results of universal and ordinary kriging interpolation methods, respectively.

Last Month Average Temperature (IDW)

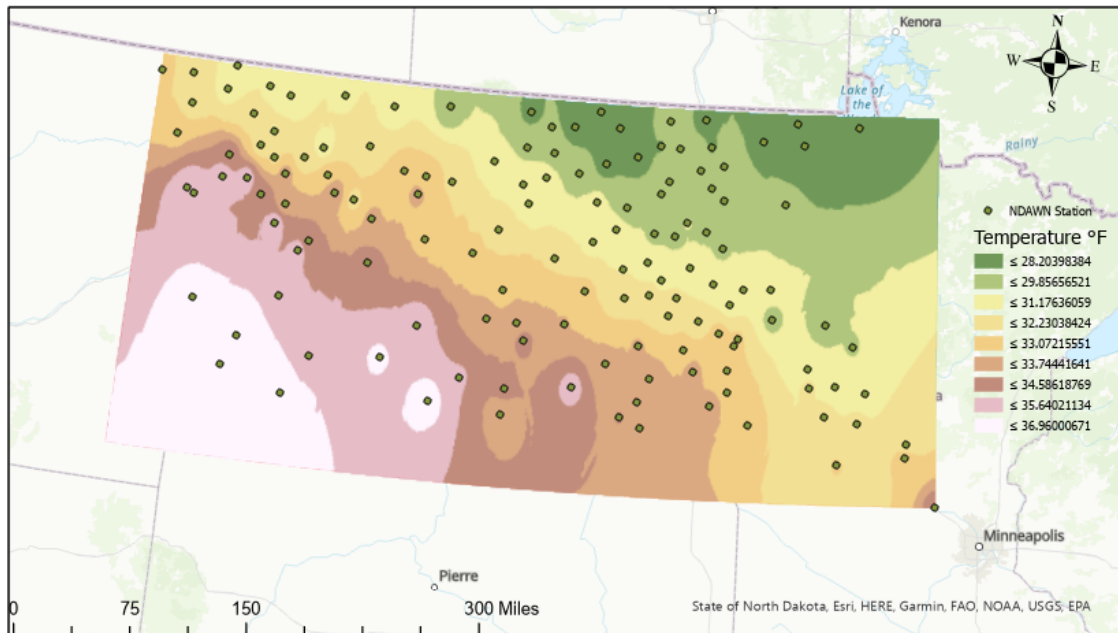


Figure 2: The interpolated raster resulted from IDW interpolation analysis.

Last Month Average Temperature (Global Kriging)

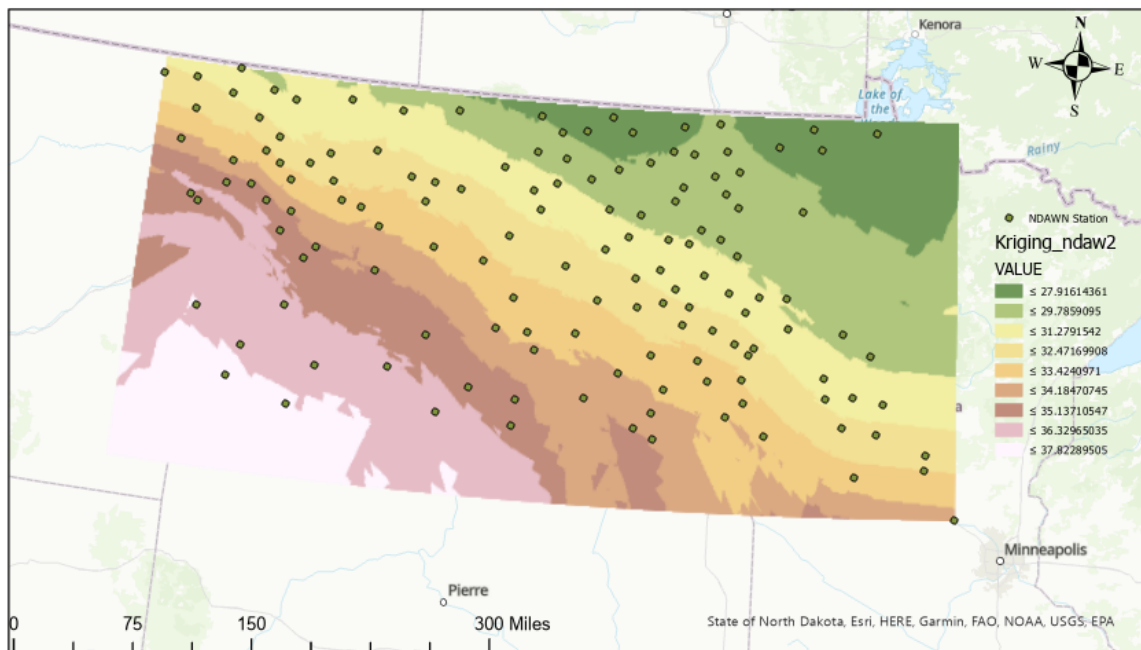


Figure 3: The interpolated raster resulted from universal kriging interpolation analysis.

Last Month Average Temperature (Kriging)

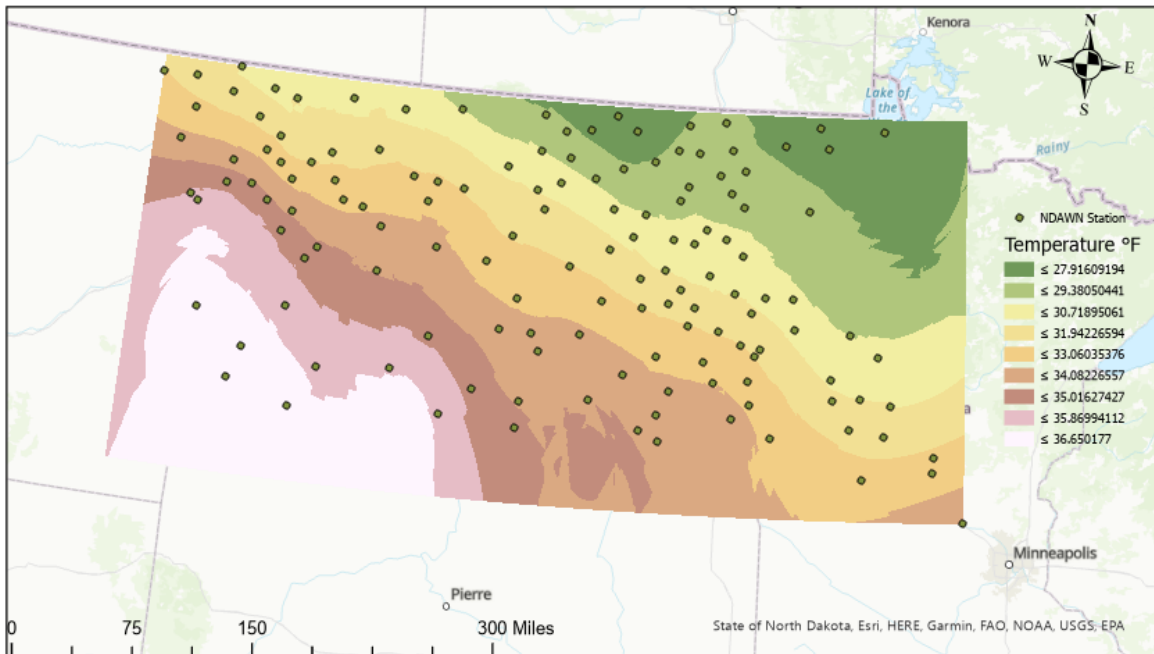


Figure 4: The interpolated raster resulted from ordinary/local kriging interpolation analysis.

Last Month Average Temperature (Polynomial Interpolation)

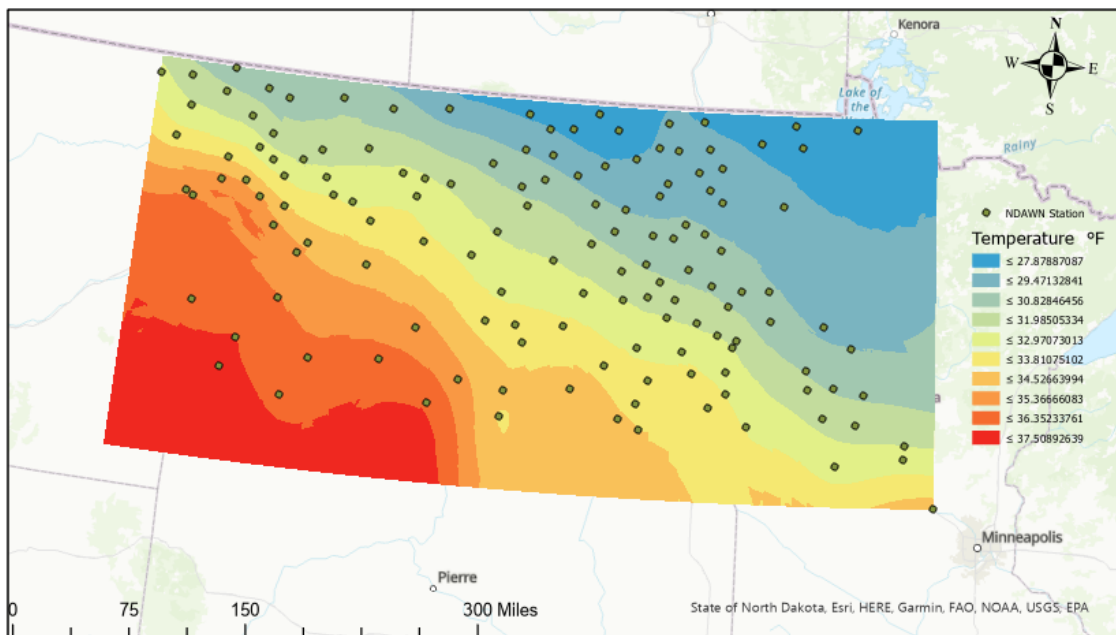


Figure 5: The interpolated raster resulted from local polynomial interpolation analysis.

Results Verification

The evaluation of an interpolation method requires the presence of actual ground truth dataset to check and evaluate the performance of the interpolation methods based on that (i.e., cross-validation technique). However, in this project such a ground truth dataset was missing.

Therefore, to evaluate and verify the accuracy of the created models could be done visually and by comparing the results with the ones provided in the literature. The resulting raster data provided in this report are consistent with those provided in the literature [2], [3]. Moreover, a comparison between the results of the interpolation methods in this study shows the presence of a very similar pattern. Hence, this could also be regarded as another way of result verification.

Discussion and Conclusion

In this project, I selected IDW, ordinary, and universal kriging methods as the interpolation methods that consider a spatial autocorrelation in the data [4]. On the other side, I also wanted to experiment an interpolation method that does not assume a spatial autocorrelation in the data. Hence, I chose local polynomial interpolation. In addition, by choosing IDW, local polynomial, and kriging interpolation methods, I wanted to implement all types of model complexities of low, medium, and high, respectively. Moreover, according to ESRI, the local polynomial interpolation method is most appropriate for normally distributed data [4]. Since the temperature data has an innate normal distribution [5], local polynomial interpolation seemed to be appropriate. Also, the studies conducted on the evaluation of interpolation methods for climate data show that universal and local kriging methods followed by IDW and local interpolation methods are the most appropriate approached for interpolating monthly average temperature data for the state-wide scale [6], [7]. Overall, it is evident that the result of IDW method has considered the spatial autocorrelation more significantly. Therefore, the impact of distance is more counted in this interpolation method. However, in the other methods, there is an overall similar layered pattern.

References

- [1] "NDAWN GitHub repository," 2021. [Online]. Available: https://github.com/gettec/get_us_weather. [Accessed: 12-Jan-2021].
- [2] M. R. Holdaway, "Spatial modeling and interpolation of monthly temperature using kriging," *Clim. Res.*, vol. 6, no. 3, pp. 215–225, 1996.
- [3] T. Wu and Y. Li, "Spatial interpolation of temperature in the United States using residual kriging," *Appl. Geogr.*, vol. 44, pp. 112–120, 2013.
- [4] ESRI, "Classification trees of the interpolation methods offered in Geostatistical Analyst," 2021. [Online]. Available: <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm>. [Accessed: 12-Apr-2021].
- [5] R. D. Harmel, C. W. Richardson, C. L. Hanson, and G. L. Johnson, "Evaluating the adequacy of simulating maximum and minimum daily air temperature with the normal distribution," *J. Appl. Meteorol.*, vol. 41, no. 7, pp. 744–753, 2002.
- [6] N. Hofstra, M. Haylock, M. New, P. Jones, and C. Frei, "Comparison of six methods for the interpolation of daily, European climate data," *J. Geophys. Res. Atmos.*, vol. 113, no. D21, 2008.
- [7] W. D. Walter, J. W. Fischer, S. Baruch-Mordo, and K. C. VerCauteren, "What is the

proper method to delineate home range of an animal using today's advanced GPS telemetry systems: the initial step," in *Modern telemetry*, InTech, 2011.

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	26
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	20
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	94