Lab Report

Title: Lab 3 - 2
Notice: Dr. Bryan Runck
Author: Greg Kohler
Date: November 14th, 2023

Project Repository: https://github.com/greg-kohler/GIS5571/tree/main/Lab3/Part 2
Google Drive Link: N/A
Time Spent: 10

Abstract

In this lab, I interpolated maximum, average, and minimum temperature data for weather stations in the NDAWN network. This report outlines the resources used and the input data needed to perform the analysis. Following, it details the methods used to import and interpolate the data, as well as explains the three interpolation methods chosen and the justification for using them. Results are showcased in the form of various maps that show different interpolation methods for different temperature datasets. This section also compares and contrasts the different methods. After this, the results are verified using both qualitative and quantitative methods. The final section discusses lessons learned and main takeaways.

Problem Statement

The purpose of this lab was to download maximum, average, and temperature station data from all weather stations that report to NDAWN. With this temperature data, the next step was to use three different interpolation methods to estimate the temperature between stations. For this lab, I chose to use IDW, Empirical Bayesian Kriging, and Diffusion Interpolation with Barriers.

Table 1 - Requirements

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	NDAWN Weather Data	Raw CSV file from NDAWN with Temperature	Station Coordinates	Maximum, average and minimum temperatur	NDAWN	CSV had to be cleaned up and converted to Shapefile
2	ArcGIS Pro	Program used to store and visualize spatial data	N/A	N/A	N/A	N/A
3	ArcGIS Pro Notebooks	Jupyter notebooks used to store and run Python code	N/A	N/A	N/A	N/A
4	ArcGIS Pro Toolbox	Geoprocessing tools used for interpolation analysis	N/A	N/A	N/A	N/A

Input Data

The data used for this lab all came from NDAWN, or the North Dakota Agricultural Weather Network. For this analysis, I pulled data for the last 30 days from all stations in North Dakota. The data I pulled was daily average, daily, maximum, and daily minimum temperature. This data is pulled from the website in a CSV format. The columns for this CSV are renamed and the latitudes and longitudes are used to create point geometry.

Table 2 - Input Data

#	Title	Purpose in Analysis	Link to Source
1	NDAWN Daily Average 30 day Temperature	Used to create station points and have average daily temperature.	NDAWN
2	NDAWN Daily Maximum 30 Day Temperature	Used to create station points and have average maximum daily temperature.	NDAWN
3	NDAWN Daily Minimum 30 Day Temperature	Used to create station points and have average minimum temperature.	NDAWN

Methods

The methods used in this lab included two for loops that iterated through a list of three variables. These variables represented the daily average, maximum, and minimum temperatures at weather stations in the NDAWN network. In the first loop, the first step is to pull the CSV from the NDAWN website. This CSV contains 30 days of temperature data from all weather stations in the NDAWN network. Once the CSV is pulled it is saved to a local folder in its raw format. Next, I used pandas to open up the CSV, rename the columns to something usable, and save the newly formatted CSV. After this, the formatted CSV is put into a pandas data frame, and using the groupby function (OpenAI, 2023), the stations ‘ day temperatures are averaged. The averaged temperatures are then saved to a CSV. The final step of the loop is to convert the latitude and longitude into point geometry, open it in a geodata frame, and save the points to a shapefile.

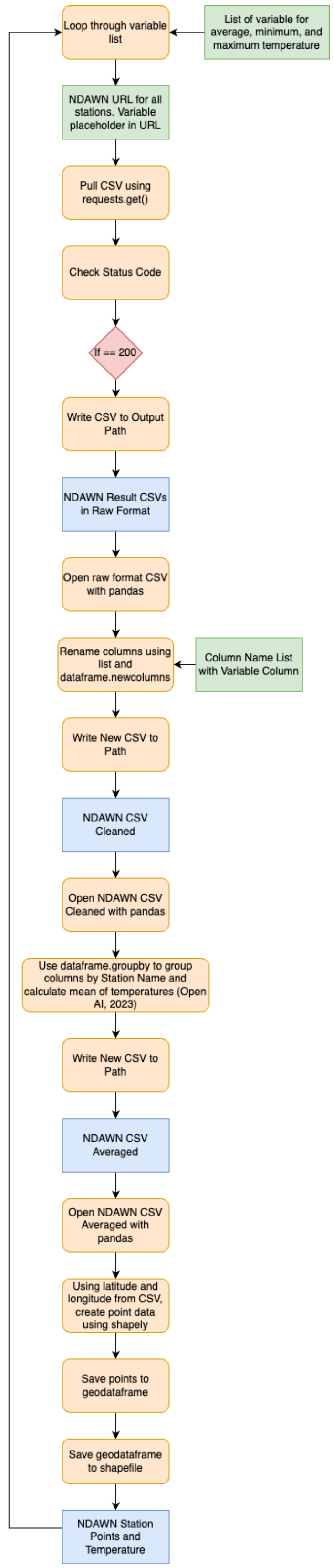


Figure 1 - Average Temperature Loop

After retrieving and converting the weather data into points, the next section of the lab was to create interpolation layers (See Figure 2). For this lab, I chose to use Inverse Distance Weighting (IDW), Empirical Bayesian Kriging, and Diffusion Interpolation with Barriers (DKB). I chose to use Inverse Distance Weighting because it was a requirement. I chose to use this interpolation method with an alpha of 2, as this is the most common alpha. This type of interpolation works when there is one prediction per location, which is the case for the weather stations in this lab. IDW also works well with low complexity, which works for this dataset (*Classification Trees of the Interpolation Methods Offered in Geostatistical Analyst—ArcMap | Documentation*, n.d.). For my second interpolation method, I chose to use Empirical Bayesian Kriging. Kriging interpolation is a good choice because it provides probability, has a fast processing time, and works for having one prediction per location (*Classification Trees of the Interpolation Methods Offered in Geostatistical Analyst—ArcMap | Documentation*, n.d.). I chose to use Empirical Bayesian Kriging because ArcGIS Pro recommended the use of Empirical Bayesian. This form of Kriging processing automatically calculates parameters and is more accurate for smaller datasets than other forms of Kriging. It also uses an intrinsic random function, which differs from other Kriging models. This allows values to have a greater deviation (*What Is Empirical Bayesian Kriging?—ArcMap | Documentation*, n.d.). The third method, Diffusion Interpolation with Barriers, I chose to use because it works well with heat transfer (*How Diffusion Interpolation with Barriers Works—ArcMap | Documentation*, n.d.), which is what is being interpolated in this lab. Similar to Kriging and IDW, DKB works well for analysis with one prediction per location. DKB also provides a prediction for values, rather than a probability and has intermediate complexity (*Classification Trees of the Interpolation Methods Offered in Geostatistical Analyst—ArcMap | Documentation*, n.d.). For all of these interpolation methods, a for loop ran through them plugging in the different variables for minimum, average, and maximum temperature.

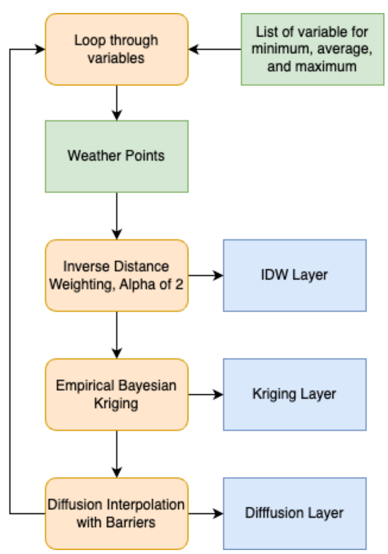


Figure 2 - Interpolation Loop

Results

The first result for this lab is a map of all of the stations in the NDAWN network displaying the average temperature over the last 30 days. Each point represents a station, and the color of the point correlates with its temperature, See Figure 3.

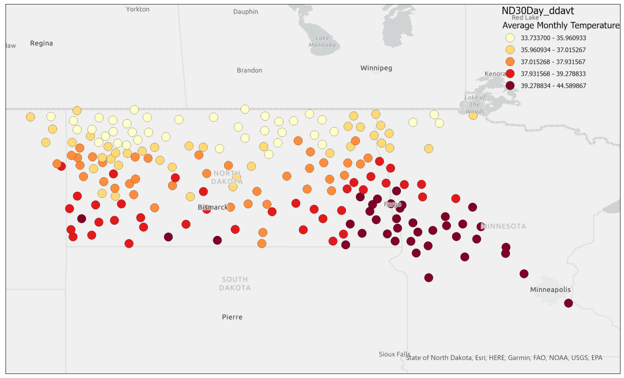


Figure 3 - Average Station Temperatures

The next set of results from this lab are nine interpolation maps (See Figure 4-12). The first set of maps is for the average monthly temperature, see Figures 4-6.

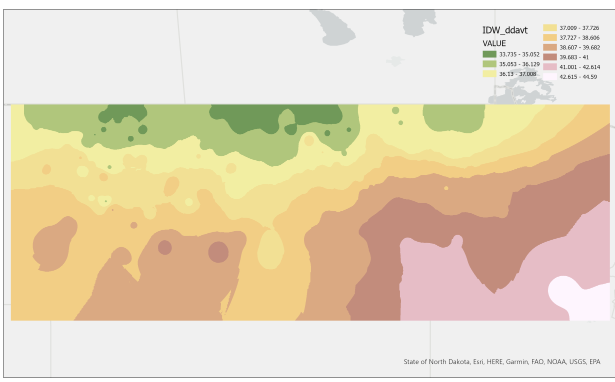


Figure 4 - Average Monthly Temperature IDW

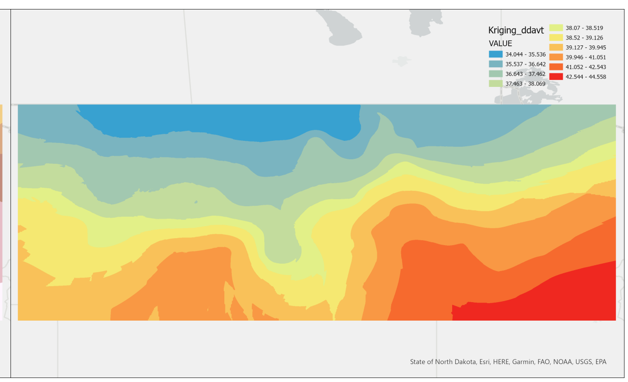


Figure 5 - Average Monthly Temperature Kriging

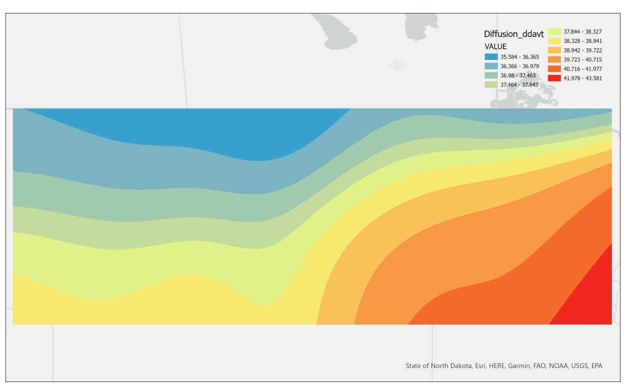
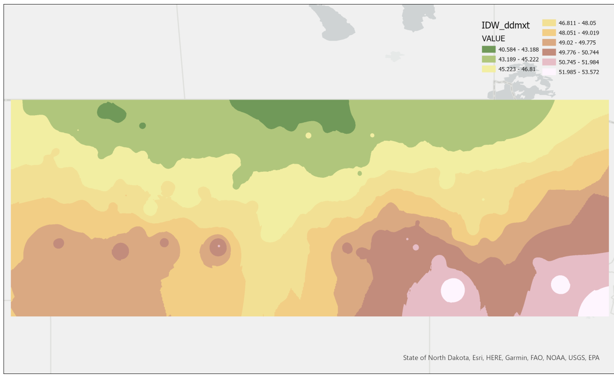


Figure 6 - Average Monthly Temperature Diffusion

The next set of figures represents the interpolation for the average monthly maximum temperature at each station, see Figures 7-9. These figures are similar to the previous maps but with slight variations.



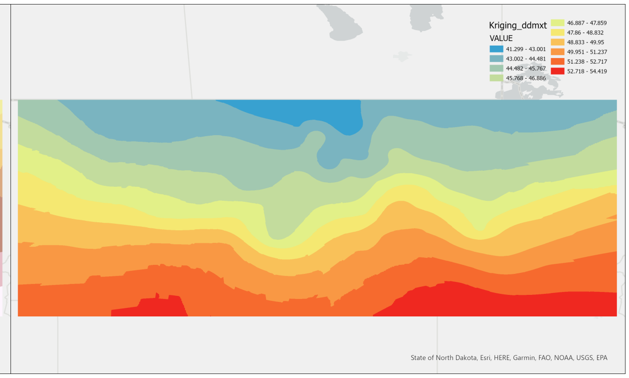


Figure 7 - Maximum Monthly Temperature IDW

Figure 8 - Maximum Monthly Temperature Kriging

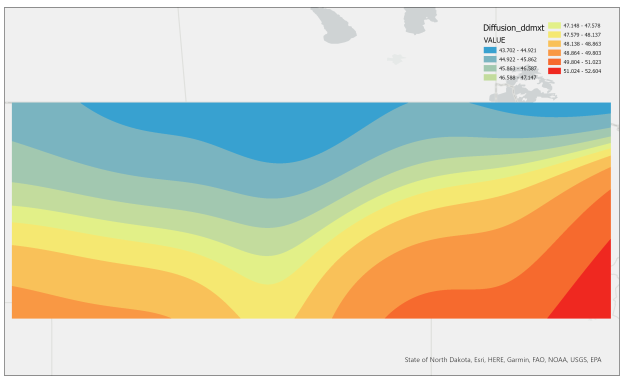


Figure 9 - Maximum Monthly Temperature Diffusion

The final set of figures represents the interpolation for the average lows for each station over the last 30 days, see Figures 10-12. These figures look the least similar at first glance, perhaps because of a greater dispersion of low temperatures.

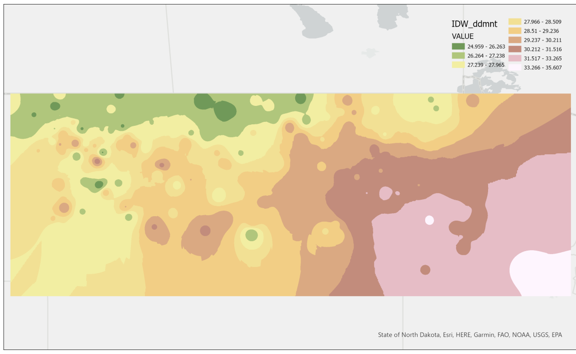


Figure 10 - Minimum Monthly Temperature IDW

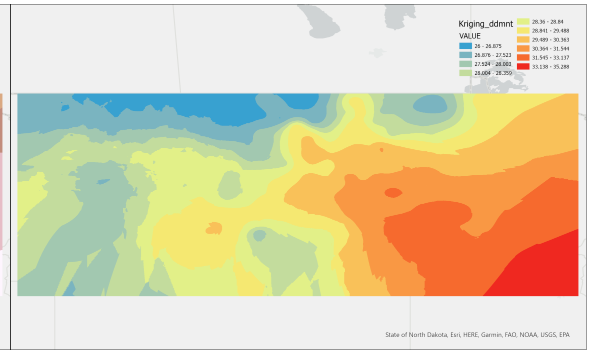


Figure 11 - Minimum Monthly Temperature Kriging

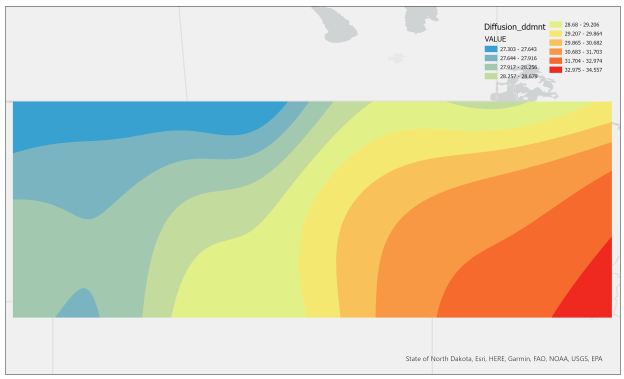


Figure 12 - Minimum Monthly Temperature Diffusion

Now that we have the interpolation results for all three temperature scenarios, we can discuss the interpolation methods. As stated in the methods section of this report, each interpolation method was chosen for a certain reason. Each interpolation method has similarities and differences. The IDW figures have the most variance in prediction. The IDW interpolation figures contain small pockets of different values. This makes sense, as IDW offers more weight to values that are closer together. The weather stations are not evenly dispersed, so there will be pockets with more weight. This aligns with the literature (*IDW (Spatial Analyst)—ArcGIS pro | Documentation,* n.d.). The Empirical Bayesian Kriging method has similar shapes to the IDW method. The Kriging method does not keep the values at station locations, even though these locations provide a known value. This is because the Kriging method offers probability rather than prediction (*What Is Empirical Bayesian Kriging?—ArcMap | Documentation*, n.d.). The final interpolation method was diffusion interpolation with barriers. The figures created with this interpolation method look visually the most like how heat transfers through the air (Yang & Xing, 2021). Similar to the Kriging method, the exact temperature values are not maintained, but rather a broader prediction is used. The diffusion method allows predictions to flow, rather than have harder lines like the other two interpolation methods (*How Diffusion Interpolation with Barriers Works—ArcMap | Documentation*, n.d.).

Results Verification

Several different aspects of this lab needed to be verified. First, I needed to verify the data was being pulled from the last 30 days. For this, I ran the code on multiple different days. For each day I ran the code, the dataset was updated to reflect the most recent day. This confirmed the URL I used for the data pull was updated to the previous 30 days. To verify the temperatures were correctly averaged, I confirmed the code ran successfully. This code used a built-in function to group all of the stations by name and average out the temperature. We can assume this built-in mean function worked correctly. To verify that all of the stations are present, we know that 193 NDAWN weather stations report data. Looking at our average temperature point map, there are 193 points. This confirms we have all of the stations. With this, all the code for the pipeline to convert the NDAWN CSV to average temperature points is verified.

For the verification of the interpolation results, the code used was ArcGIS Pro tools. The code for all three interpolation methods successfully ran and created the interpolation rasters. This verifies that the interpolation was created correctly, given the inputs were also correct.
To verify the interpolation methods used, as discussed in the methods section, I chose to use ESRI's decision trees to decipher why IDW and Empirical Bayesian Kriging would be good fits for this dataset *Classification Trees of the Interpolation Methods Offered in Geostatistical Analyst—ArcMap | Documentation*, n.d.). I also referenced Yang & Xing, 2021 and their discussion about Diffusion Interpolation with Barriers. In this article, the authors discussed the basis of DKB, which is the heat equation. This equation attempts to solve how heat spreads out in similar environments (Yang & Xing, 2021). This article confirms that diffusion is the preferred method used for temperature interpolation.

Discussion and Conclusion

This lab involved creating a pipeline to retrieve temperature data from NDAWN and use different interpolation methods to visualize the data. The process of retrieving the temperature data was straightforward, as we had done in several other labs. My biggest hurdle with this step was deciphering the URL so that it would automatically update to be the most recent 30 days, and give me the temperature variable I needed. After analyzing the URL, I found there to be a variable "quick_pick=30_d." This automatically gave me the previous 30 days, so I could remove other variables that specified a certain date and year. For the temperature variable, I found the "variable=ddavt." I deciphered that this variable represents the daily average, and I could plug in ddmxt and ddmnt to get results for daily maximum and daily minimum. My next challenge was taking all of this data and calculating the 30-day average. After consulting with OpenAI, 2023, I learned about the pandas groupby function. This function allowed me to group all the station data by name and average it out in one command. This helped save me time and erased any possibility of error.

Once I had the temperature points, a big part of this lab was the interpolation methods. It was a requirement to use IDW and a form of Kriging, but that left flexibility for a third interpolation method and the form of Kriging. I consulted the ESRI decision trees for why I needed to use IDW and a form of Kriging. IDW seemed to be a pretty standard form of interpretation and is what we went over in class. For Kriging, I was overwhelmed by all of the options, but luckily ArcGIS Pro toolbox recommended the use of Empirical Bayesian. This method has greater functionality and seemed the easiest to use for this lab. Finally, I chose to use diffusion interpolation with barriers after referencing the ESRI article and the Yang & Xing, 2021 article. Both of these articles discuss the use of diffusion interpolation for heat diffusion, which is exactly what is being modeled in this lab.

Overall, this lab allowed me to build my confidence in creating an ETL pipeline. I was able to create a for loop that created all of the points and performed all of the interpolations without much help needed, which is not something I could've done for the first lab. This lab also gave me the chance to explore interpolation methods in a more hands-on approach. It is interesting how different models can be used for different datasets. There is not one right answer for which interpolation method to use, but there are some that are better than others.

References

References

Classification trees of the interpolation methods offered in Geostatistical Analyst—ArcMap | Documentation. (n.d.). ArcGIS Desktop; ESRI. Retrieved November 13, 2023, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

draw.io. (n.d.). Www.drawio.com. Retrieved October 9, 2023, from http://www.drawio.com

How Diffusion Interpolation With Barriers works—ArcMap | Documentation. (n.d.). ArcGIS Desktop; ESRI. Retrieved November 13, 2023, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-diffusion-interpolation-with-barriers-works.htm

IDW (Spatial Analyst)—ArcGIS Pro | Documentation. (n.d.). ESRI; ESRI. Retrieved November 14, 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/idw.htm

OpenAI. (2023). *ChatGPT*. Chat.openai.com; OpenAI. https://chat.openai.com/

What is empirical Bayesian kriging?—ArcMap | Documentation. (n.d.). ArcGIS Desktop; ESRI. Retrieved November 13, 2023, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Yang, R., & Xing, B. (2021). A Comparison of the Performance of Different Interpolation Methods in Replicating Rainfall Magnitudes under Different Climatic Conditions in Chongqing Province (China). *Atmosphere*, *12*(10), 1318. MDPI. https://doi.org/10.3390/atmos12101318

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	**18**
		100	**98**