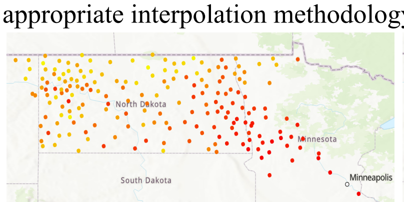
Lab Report

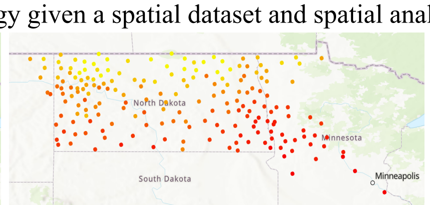
Title: Spatial Analysis of 30-Day Temperature Trends at NDAWN Stations: An ETL Approach and Interpolation Comparison Using IDW and Kriging Methods
Notice: Dr. Bryan Runck
Author: Laure Briol
Date: 11/10/2023
Project Repository: https://github.com/Briol009/GIS-5571
Google Drive Link:
https://docs.google.com/document/d/1sb7AX6bW3Zl7s2_NE1MC1OtwEvpiFbqA/edit?usp=sharing&ouid=109074632626385117400&rtpof=true&sd=true
Time Spent: *20*

Abstract
The purpose of this lab was to build a functional real-time data pipeline for the purposes of visualization and analysis of the data. The North Dakota Agricultural Weather Network's (NDAWN) data was extracted into the pipeline to examine the last 30 days of temperature data to create a final map product. The data was station points of the average monthly temperature, maximum temperatures, and minimum temperatures from the last 30 days. The data used three different methods of interpolation: Inverse Distance Weighting, Kriging, and Diffusion with Barriers. This resulted in a total of nine different interpolation outputs. Understanding and interpreting data is critical and must be informed by ESRI's decision guide and other literature within the field. In the discipline of geography, the practical application and analysis of spatial data outputs play a crucial role in deepening our insights into spatial patterns and relationships. Concepts such as Tobler's First Law, spatial autocorrelation, and the alpha parameter in interpolation methods contribute significantly to our understanding and other's understanding into the rationale and importance of this type of spatial analysis.

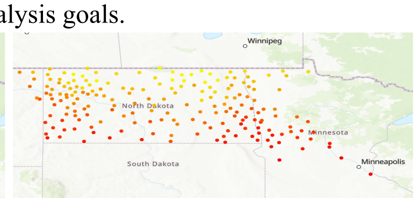
Problem Statement
The purpose of this section of the lab is to gain understanding into how spatial interpolation allows us to understand how nearby or neighboring values are more similar than distance values to create more meaningful predictions across space using real time visualization data. The goal was to use Inverse Distance Weighting, Kriging, and Diffusion interpolation methods to access data distribution across an area in space for better data driven decisions. Tobler's First Law of Geography 'everything is related to everything else, but near things are more related than distant things' uses the concept of spatial autocorrelation to apply these principles at a practical level. With all this in mind, these interpolation methods are powerful spatial analyst tools to help discern spatial patterns in space with no predicted or actual available data. Still, what needs to be understood and is most important is the justification into which interpolation methodology should and should not be used and why. This lab seeks to understand the appropriate interpolation methodology given a spatial dataset and spatial analysis goals.



Monthly Minimum Temperature



Monthly Average Temperature



Monthly Maximum Temperature

Exploring Temperature Interpolation Techniques with NDAWN CSV Datasets
*All requirements were done within the same notebook

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	NDAWN Data Download	Used CSV file to extract average, minimum, and maximum temperature data	CSV, SHP	Point feature (temperature)	NDAWN Weather Stations	Created a variable list to iterate through the NDAWN URL, created for

		points from all the stations				loop that: inserts variable into the NDAWN URL string (replaced r with fr), retrieves csv data, saves data to local file system, cleans the data column, renames the columns, groups data by location, calculates average temperature at each location, saves CSV, creates point features of weather station
2	Interpolation Notebook	Used arcpy: Diffusion Interpolation with Barriers, IDW, and Kriging to visualize the estimated value (temperature) between weather stations.	SHP, TIFF	Point Feature to Raster Surface	NDAWN Weather Stations	Used Geoprocessing tool to user: Diffusion, IDW, and Kriging setting the alpha to 2 (default parameters for the rest of the dialogue boxes).

Input Data
The data is from the North Dakota Agricultural Weather Network and represents for the last 30 days the variables: average, maximum, and minimum air temperature for all the stations. The data was produced by North Dakota State University within the School of Natural Resource Sciences within a csv dataset file.

NDAWN 30 Day Temperature Data

#	Title	Purpose in Analysis	Link to Source
1	Average Air Temperature	The purpose of this dataset is to create a map of the average monthly temperature and use the dataset to compare and contrast three different types of interpolation methods (IDW, Kriging, Diffusion).	NDAWN
2	Minimum Air Temperature	The purpose of this dataset is to create a map of the minimum monthly temperature and use the dataset to compare and contrast three different types of interpolation methods (IDW, Kriging, Diffusion).	NDAWN
3	Maximum Air Temperature	The purpose of this dataset is to create a map of the maximum monthly temperature and use the dataset to compare and contrast three different types of interpolation methods (IDW, Kriging, Diffusion).	NDAWN

Methods

Figure 1: Interpolated Surfaces

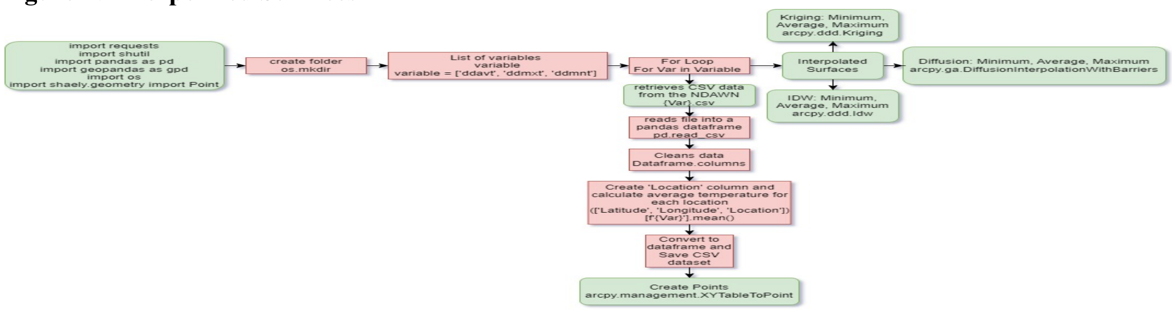


Figure 1. was created in ArcGIS Pro using Jupyter Notebooks. The input data was CSV data from NDAWN weather locations within the past 30 days (minimum, average, maximum) and a for loop used a list of three variables placed within the CSV API URL link, which then: read the input data into a pandas dataframe, cleaned the columns, renamed the columns, grouped the location columns together, saved the dataset, and created points using an arcpy function. Once the CSV was converted to point data, three interpolated raster surfaces were created: Kriging, Inverse Distance Weighted, and Diffusion Interpolation with Barriers to understand and estimate temperature between weather stations.

Results

Figure 2: Inverse Distance Weighted Interpolated Surfaces

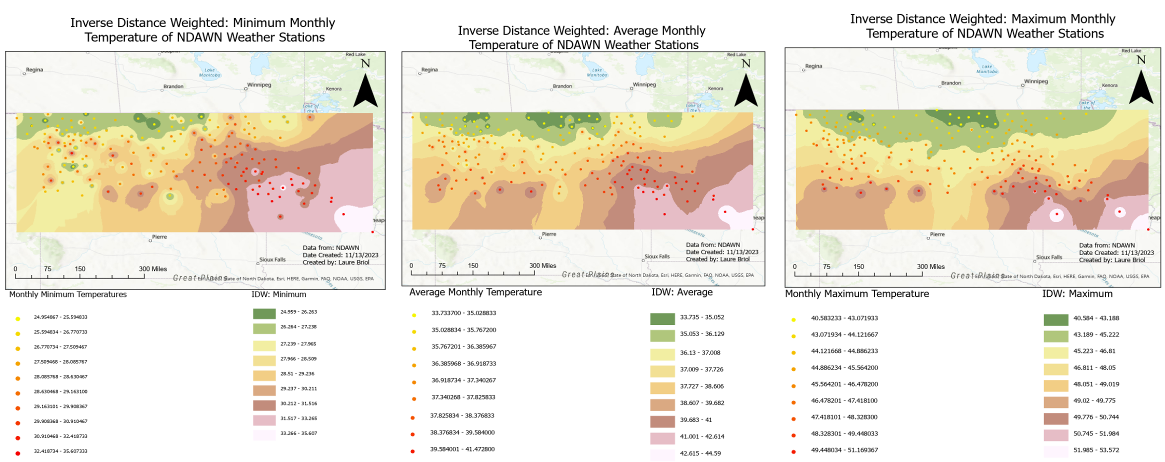


Figure 3: Diffusion Interpolation with Barriers Interpolated Surfaces

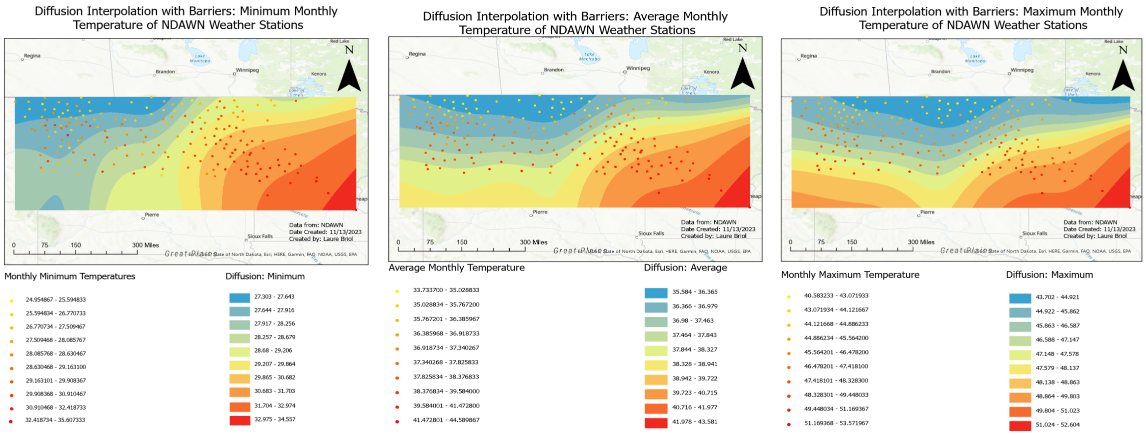


Figure 4: Kriging Interpolated Surfaces

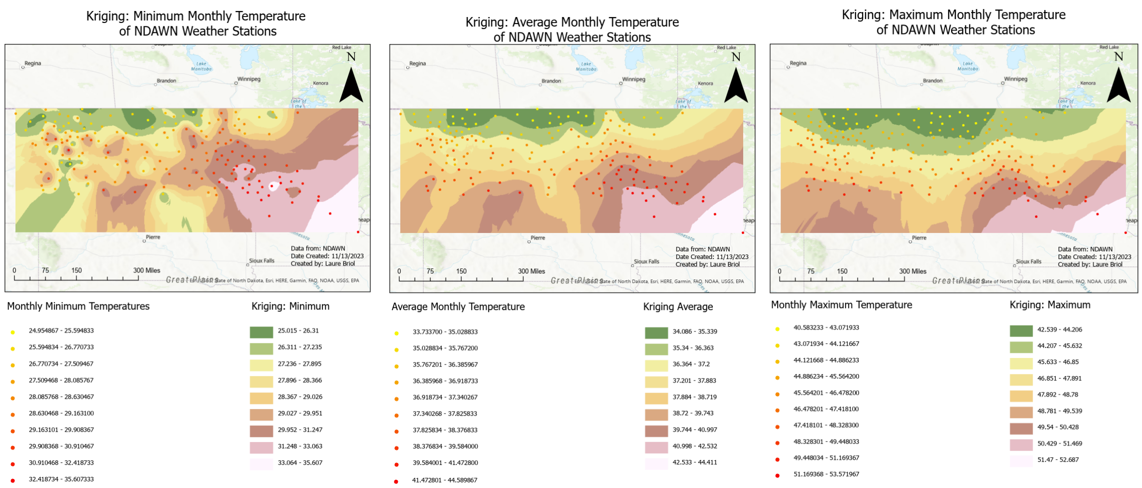
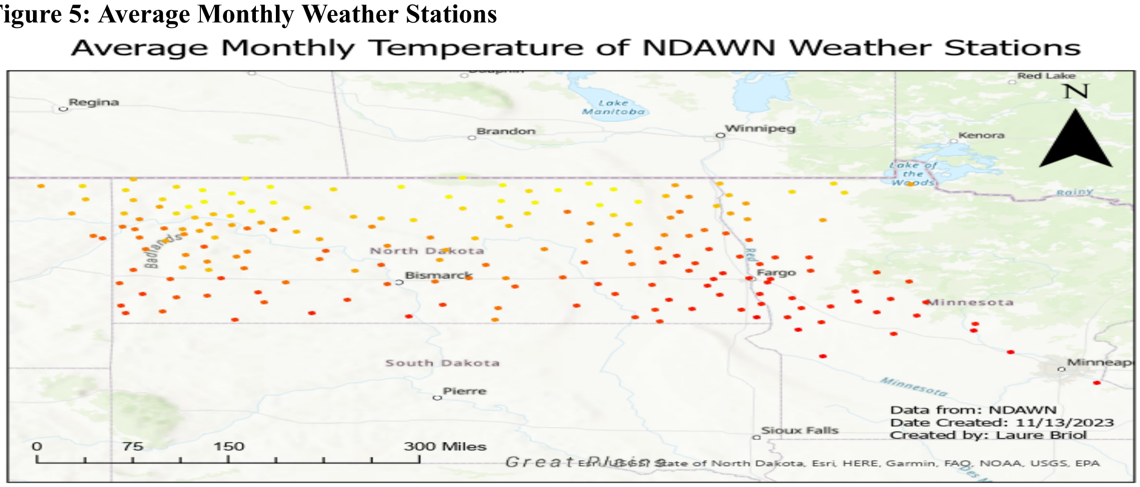


Figure 5: Average Monthly Weather Stations



The results were nine interpolated surfaces created from on-the-fly NDAWN built within an ETL pipeline. The three surface types created were Inverse Distance Weighted, Kriging, and Diffusion Interpolation with Barriers with each being created from the minimum, average, and maximum temperature from all the weather stations from the last 30 days in real-time (Figure 2-4). To understand the above methodologies there needs to be a comparison between Inverse Distance Weighting, Kriging, and Diffusion with Barriers. In terms of similarities, Kriging, like the other interpolation tools, is a geoprocessing tool that allows you to evaluate and estimate values based on the provided weather station point data and considers spatial correlation (how variables are connected via space) and creates uncertainty estimates (ESRI, 2023). Kriging models spatial variance and is useful for areas of uncertainty and where spatial patterns remain constant. Like the others, Inverse Distance Weighted calculates values at unmeasured locations, but IDW assigns weights to measured points and inverse distances (closer points have higher influence) and does not take into account spatial variability (ESRI, 2023). Diffusion with Barriers is a geoprocessing tool that models the spread of the data across space and adds barriers (this lab did not) to represent physical constraints (ESRI, 2023). According to the ESRI classification tree IDW is preferred when there is one prediction per location (weather stations) and low complexity (ESRI, 2023). Further, the ESRI classification tree has charts that show that Kriging and Diffusion provides probability and one prediction per location (ESRI, 2023). Figure 5 is the average monthly temperature map and uses a graduated color scheme to show differences in temperature at each station.

Results Verification

The first verification method used was changing the alpha scores during the data creation phase of the lab. Alpha is the power of decay and is how fast the variable being measured will decay. A higher alpha signifies more decay, resulting in reduced influence of distant variables on the dataset, with the inverse being true as well. The alpha used (2) (standard alpha Example: distance squared weighted interpolation), means that there is exponential decay and distant objects affect the dataset less, while close objects affect the dataset more. Alpha 2 decays faster than alpha 1 (logically linear) and the change in alpha during interpolation of the NDAWN weather stations reinforced how to prove interpolation is scientifically based and meaningful results are created. Further, during the creation of the interpolation surfaces the wrong variable was used and the interpolated surfaces looked like copies of each other. Given the alpha, correct data columns used, and the features were created I would assume the results to be true. For the past week (5 days) the ETL pipeline was run to ensure the newest data was being loaded in form the edited URL.

The ESRI publications make the assumption that ‘things’ (temperature) nearer together are more related that ‘things’ (temperature) far apart and would be a correct choice to use (ESRI, 2023). IDW uses measured values to predict between distances and nearer values are given more weight (local influence taken into account). Ordinary Kriging is used for data with trends, creates errors, semivariograms, and is useful for temperature and elevation data sets (continuous data) (ESRI, 2023). Diffusion Interpolation with Barriers is being used within the academic sphere to measure heat diffusion and other heat variables. Diffusion as explained within the academic literature describes how heat diffuses over time and allows the distance between points to be adjusted (Yang, 2021). The use of the kernel within the Diffusion geoprocessing tool is important, as temperature is not linear nor uniform. This study focused on precipitation data and uses the geoprocessing tool to calculate the average mean daily temperature (similar to this lab) and is excellent for modeling temperature distribution.

Discussion and Conclusion

The main lesson learned from this lab experience is that the nine interpolated surface deliverables are important, but understanding and justification for the methods used to create the final deliverables is just as important--if not more so. The main issue experienced was implementing the for loop to literature through the URL for the three desired variables (minimum, average, maximum) and was not feasible for me to use or create for the actual interpolation portion of this lab (used arcpy). This lab requires Inverse Distance Weighted, Kriging, and Diffusion with Barriers to be compared and contrasted to really at a deep level understand the functionality of these geoprocessing tools. The use of on-the-fly data and justification of methods using the literature made this lab the most important as this lab went beyond geospatial analysis and gave a taste into how deliverables will need to be updated, maintained, and communicated to peers, colleagues, and future stakeholders.

References

JGraph Ltd, Artisans. (2023). Drawio.https://www.drawio.com

OpenAI's ChatGPT. (2023, September 24). Conversation on AI Ethics with OpenAI's ChatGPT.

ESRI. (2023). How Diffusion Interpolation With Barriers works. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-diffusion-interpolation-with-barriers-works.htm

ESRI. (2023). How inverse distance weighted interpolation works. https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm

ESRI. (2023). Understanding ordinary kriging https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm

Yang R, Xing B. A Comparison of the Performance of Different Interpolation Methods in Replicating Rainfall Magnitudes under Different Climatic Conditions in Chongqing Province (China). *Atmosphere*. 2021; 12(10):1318. 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	**20**
		100	**100**