Lab3 Part 2 Report

Title: Comparing Interpolation Techniques for Predicting Temperature Trends
Notice: Dr. Bryan Runck
Author: Tzu Yu Ma
Date: Nov. 28, 2023

Project Repository: Link
Time Spent: 8

Abstract

In this lab, we're examining different methods to predict temperature. We're using three approaches - Inverse Distance Weighting (IDW), Kriging, and Global Polynomial Interpolation (GPI) - to forecast temperatures from the DNAWN site over the last 30 days. After analyzing the results, we'll explain which interpolation method is the most reliable for temperature predictions. To gather and clean our data, we'll be setting up an ETL process.

Problem Statement

According to the lab instructor's guidance, establish an ETL process to retrieve the temperature data from the North Dakota Agricultural Weather Network (NDAWN) site for the past 30 days. Visualize all station points, including their average monthly temperatures. Apply three interpolation methods to visualize the maximum and minimum temperatures of the data from the last 30 days. Conduct a comparative analysis of these three methods to identify the most effective one for interpolating temperature data.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Average temperature data- last 30 days	Average daily temperature data (point)	Coordinates	Station location, daily	NDAWN	ETL

				temperature (F)		

Table 1. Required Data

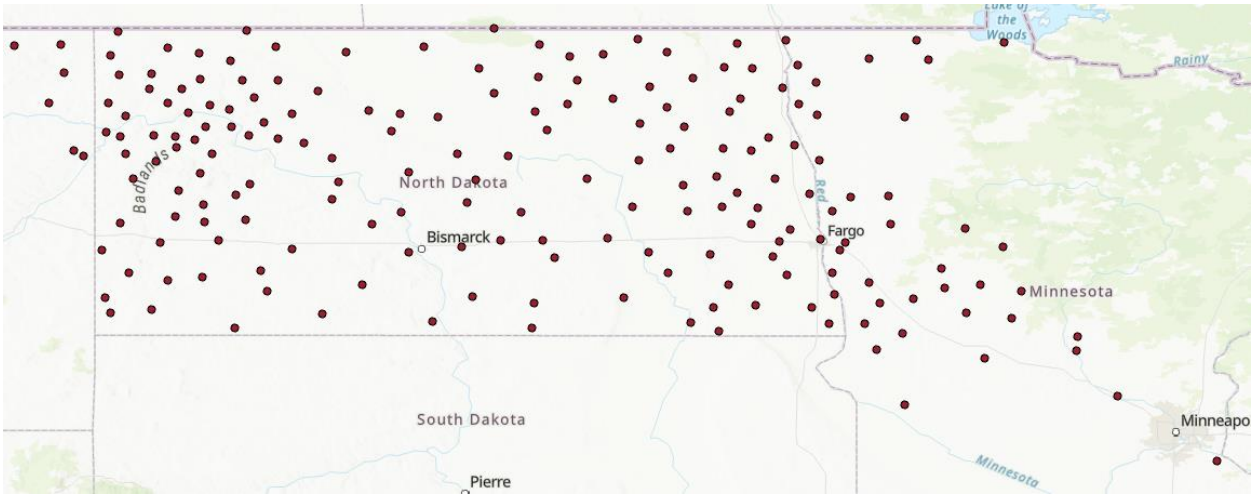


Figure 1. NDAWN Station Locations

Input Data

The data originated from the NDAWN. For this lab, the dataset spans from October 23rd, 2023, to November 23rd, 2023, providing one-month average temperature data. After extracting the information from NDAWN, the next steps involve creating a CSV file for data storage and using the coordinates to generate a point feature. Subsequently, the analysis phase can commence.

#	Title	Purpose in Analysis	Link to Source
1	Average temperature data-last 30 days	The data will be used in interpolation methods (IDW, Kriging, GPI), then clean, calculate, and visually present the maximum, minimum, and average temperatures for each station.	NDAWN

Table 2. Input data

Methods

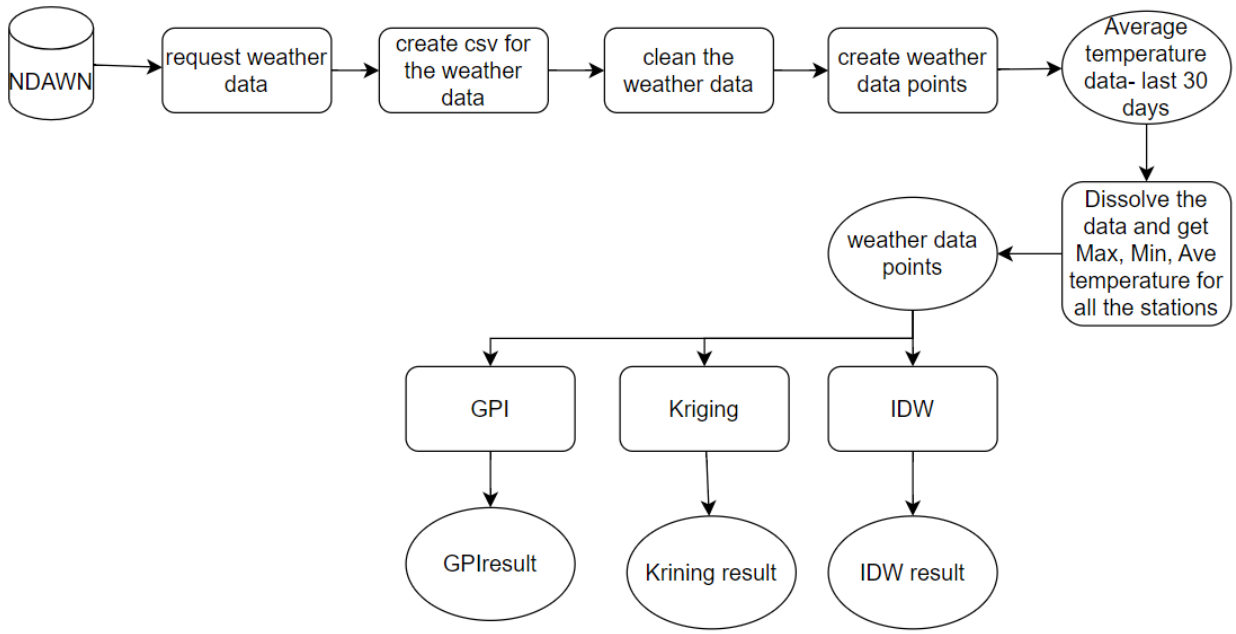
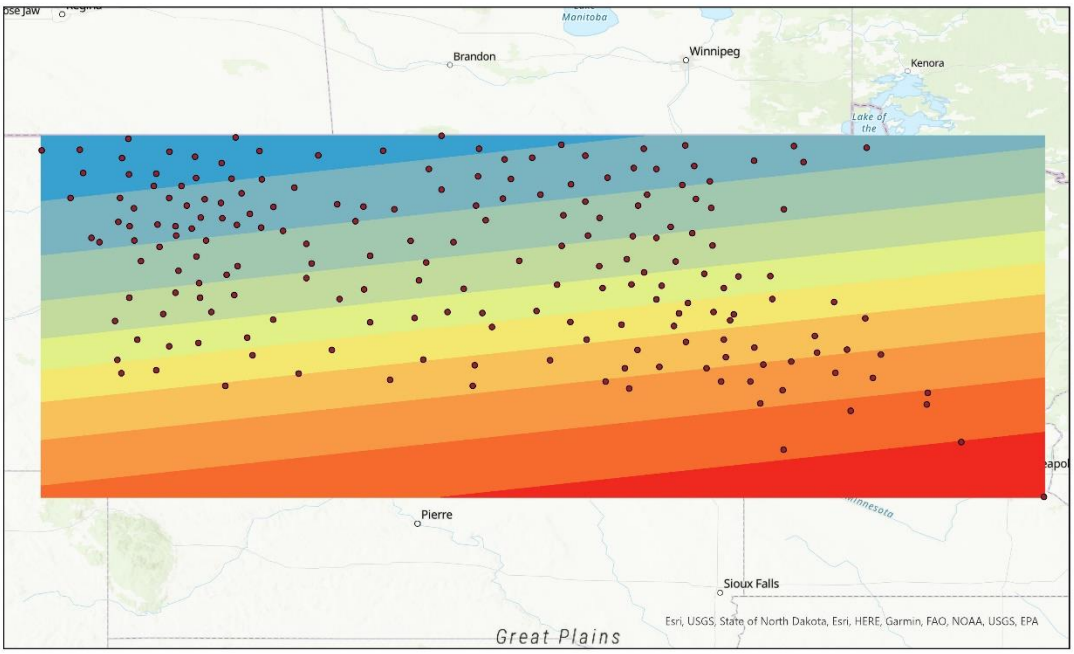


Figure 2. Data flow

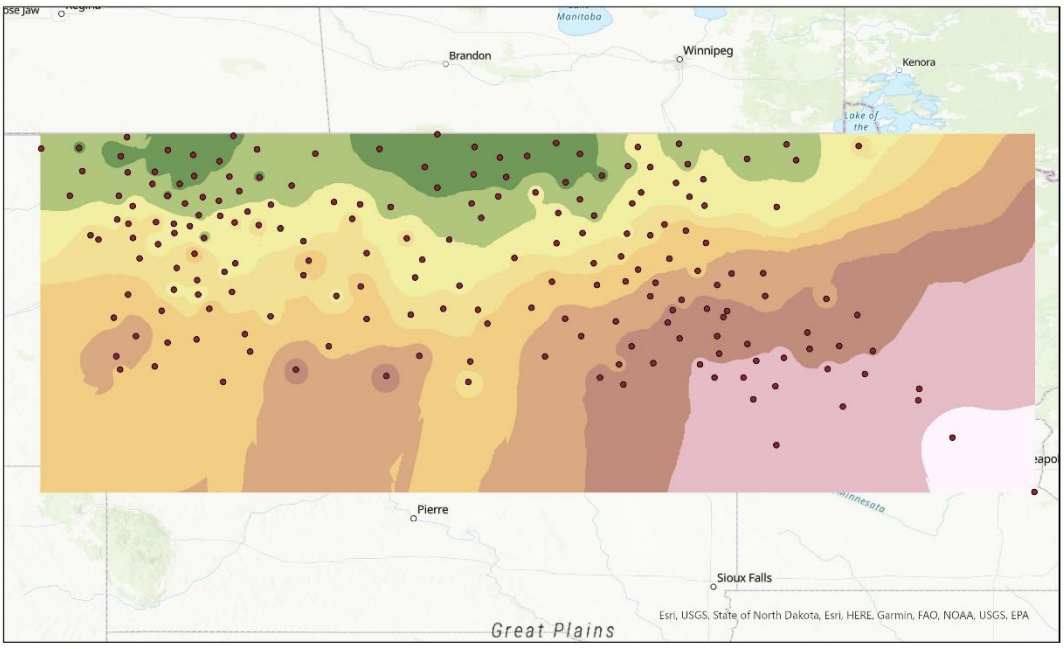
Utilize an API to extract data from NDAWN, followed by the creation of a CSV file to store the information. Given that the data contains additional rows at the top, it need to be cleaned these before constructing the data frame. Subsequently, use the cleaned data and coordinates to establish a point feature. As the dataset encompasses information from all 193 stations each day, employ a dissolve operation to organize the data by the same station name, extracting the highest and lowest temperatures. Finally, initiate the interpolation process for temperature data, employing three methods - IDW, Kriging, and GPI - to calculate mean, maximum, and minimum temperatures for each station.

Results



GPI Mean Temperature

TMP_stations	29.449 - 30.843	32.983 - 33.796	35.584 - 36.748
27.778 - 29.448	30.844 - 32.009	33.797 - 34.609	36.749 - 38.144
	32.01 - 32.982	34.61 - 35.583	38.145 - 39.814



IDW Mean Temperature

TMP_stations	29.361 - 30.654	32.29 - 33.234	36.302 - 38.731
27.587 - 29.36	30.655 - 31.599	33.235 - 34.528	38.732 - 42.059
	31.6 - 32.289	34.529 - 36.301	

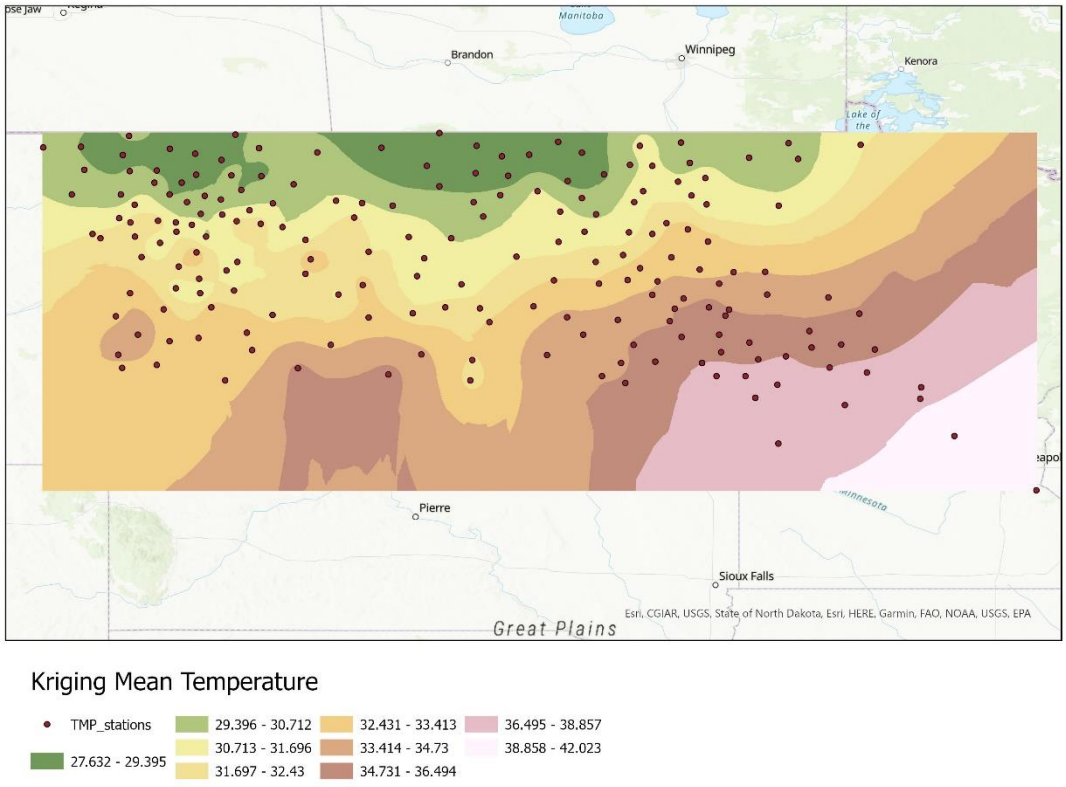
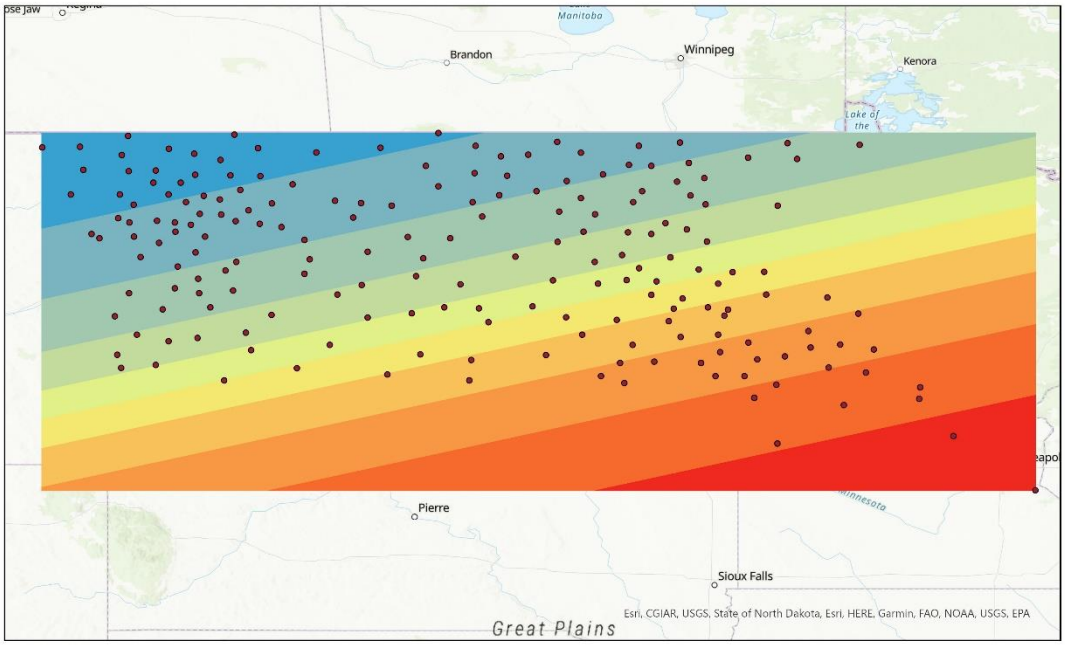
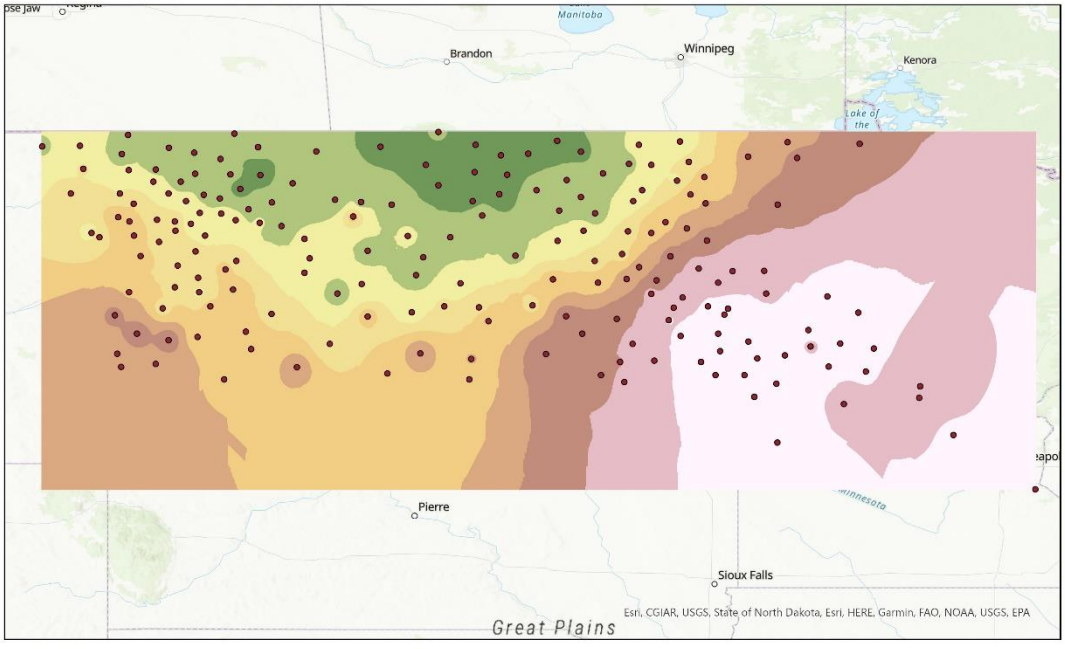


Figure 3. Mean temperature



GPI Maximum Temperature

TMP_stations	42.017 - 44.718	48.204 - 49.302	51.885 - 53.886
38.37 - 42.016	44.719 - 46.72	49.303 - 50.401	53.887 - 56.587
	46.721 - 48.203	50.402 - 51.884	56.588 - 60.234



IDW Maximum Temperature

TMP_stations	41.179 - 43.044	45.557 - 46.981	51.292 - 54.492
38.735 - 41.178	43.045 - 44.468	46.982 - 48.847	54.493 - 58.685
	44.469 - 45.556	48.848 - 51.291	

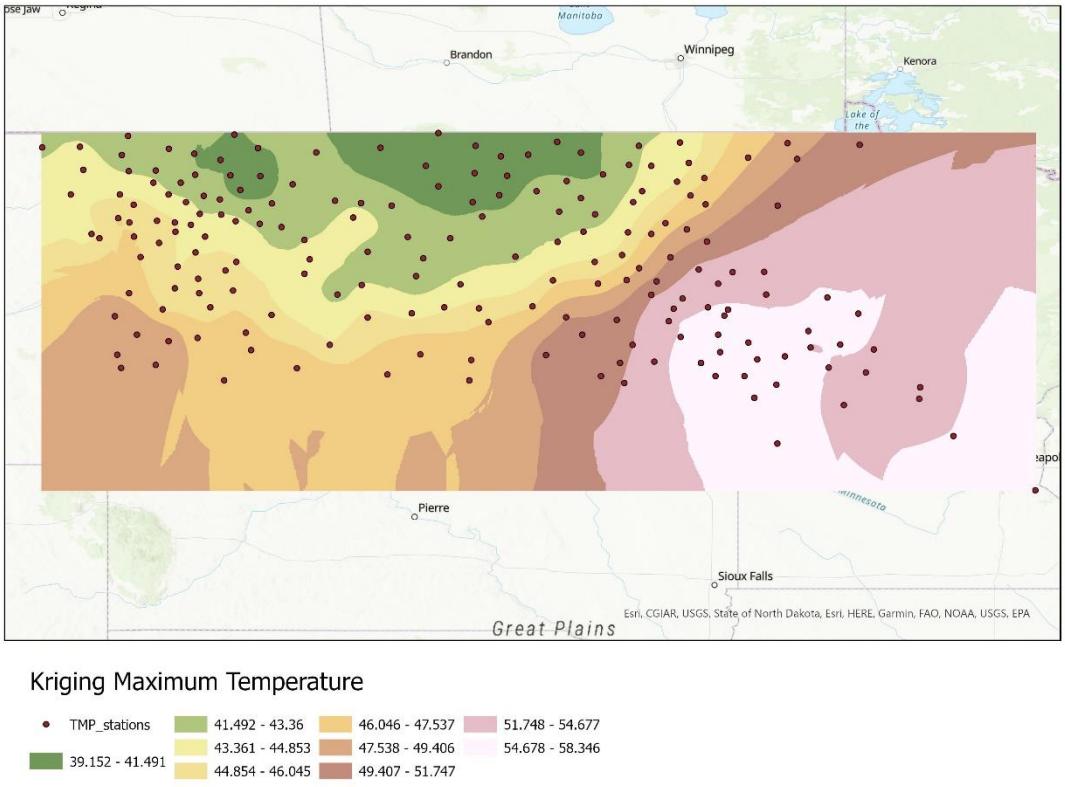
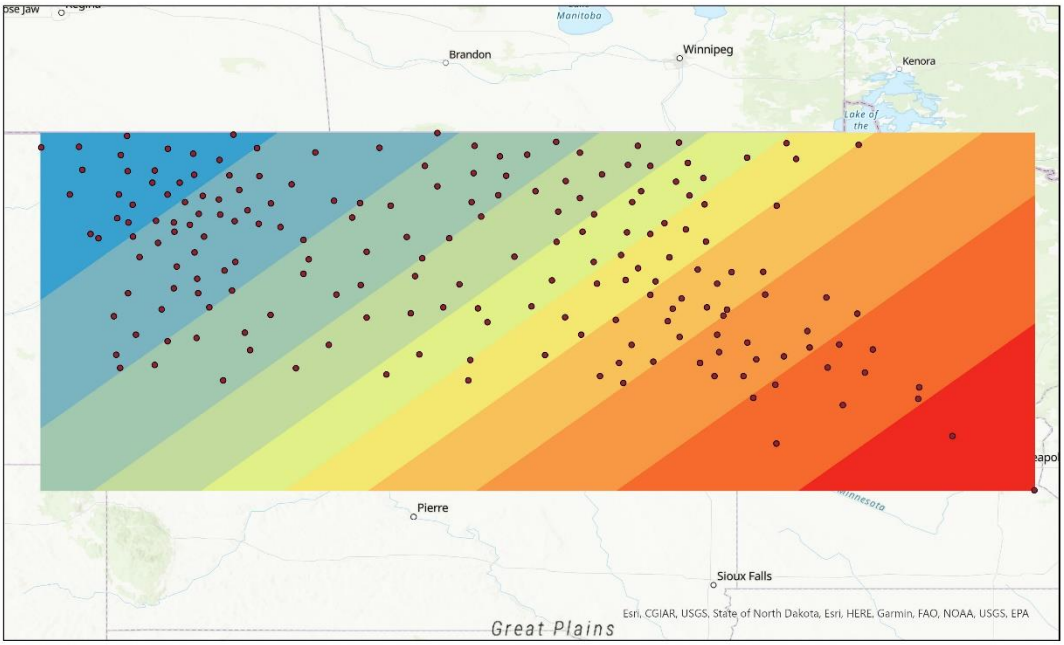
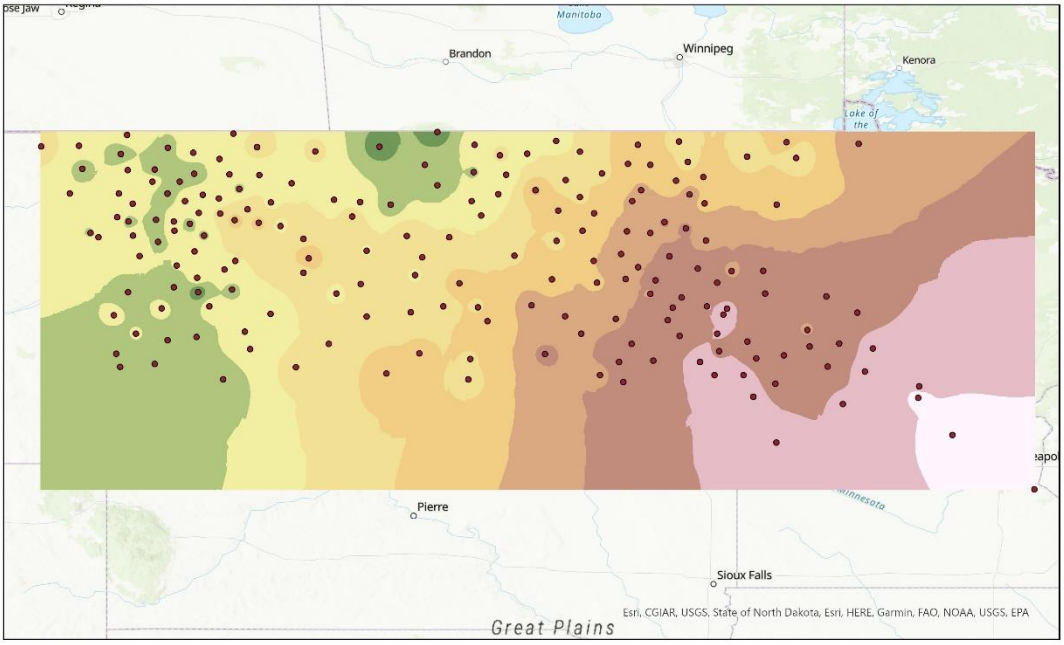


Figure 4. Maximum temperature



GPI Minimum Temperature

TMP_stations	9.649 - 12.207	15.704 - 16.876	19.573 - 21.545
6.33 - 9.648	12.208 - 14.181	16.877 - 18.05	21.546 - 24.104
	14.182 - 15.703	18.051 - 19.572	24.105 - 27.423



IDW Minimum Temperature

TMP_stations	8.064 - 10.186	14.207 - 16.582	22.216 - 25.544
5.688 - 8.063	10.187 - 12.083	16.583 - 19.241	25.545 - 29.269
	12.084 - 14.206	19.242 - 22.215	

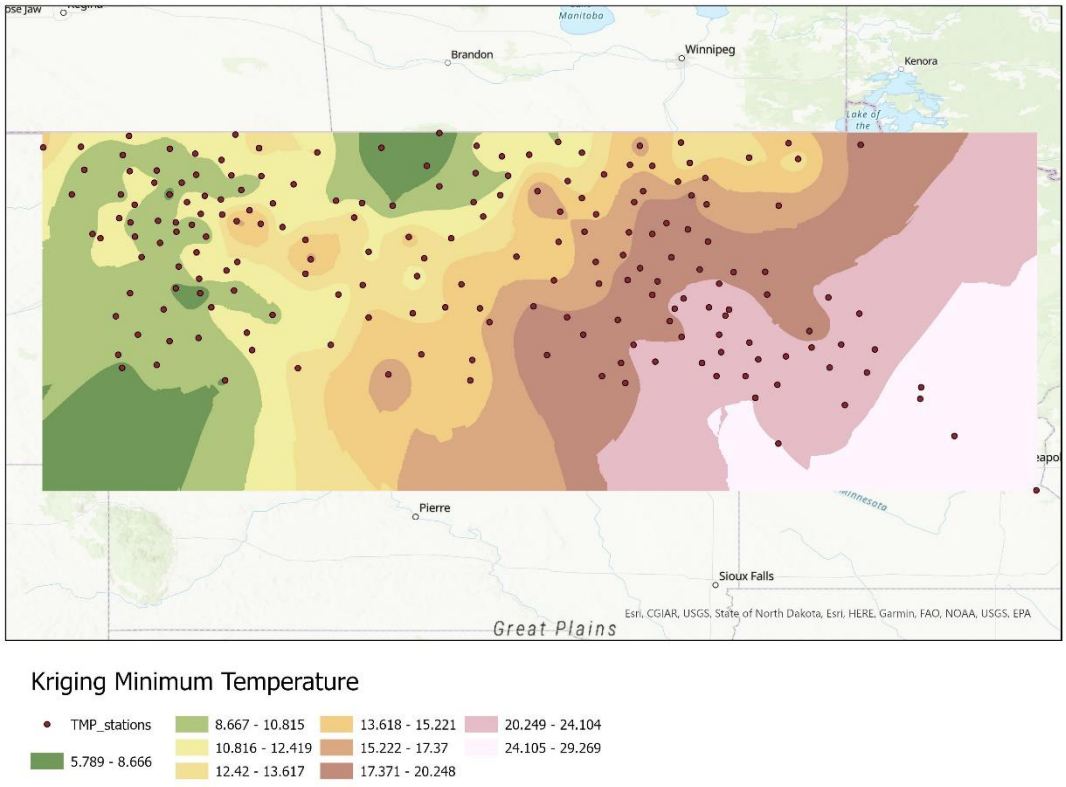


Figure 5. Minimum temperature

Results Verification

I have verified the results with the paper by Wenjing Cao, JinXing Hu, and Xiaomin Yu (2009). In their research, Kriging demonstrated the best interpolation precision for temperature data. Additionally, in the study conducted by Tingting Wu and Yingru Li (2013), it was mentioned that "Over the past several decades, kriging has become an essential tool in the field of geostatistics. Previous research has demonstrated that kriging is a superior interpolation method with high accuracy and low bias compared to other methods." Based on Kriging, they developed an improved Kriging model for temperature data. Overall, the researches indicates that Kriging is the most common and accurate method for temperature data.

Discussion and Conclusion

A weather station represents just a single point, and utilizing the information collected from these points to predict the temperature for an entire area can be achieved through interpolation. This process aims to estimate values between known data points. Before this laboratory experiment, I had never contemplated the origin of daily temperature readings. While I regularly checked the weather to plan my attire, it didn't occur to me that weather stations are strategically situated, and their limited locations somehow manage to provide information that covers an entire area.

Furthermore, the realm of interpolated methods presents numerous options, making it challenging to choose the most suitable one. I am appreciated that ESRI's classification tree simplifies the selection process, guiding us to the most appropriate method. In this lab, I experimented with four methods (excluding Spline, which is not included in the report). However, there are still other methods to explore, each dependent on various inputs and considerations. Moving forward, I am eager to explore additional interpolation methods to enhance the accuracy and reliability of temperature predictions.

References

ESRI. (n.d.). Classification trees of the interpolation methods offered in Geostatistical Analyst. ArcMap | Documentation. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

Tingting Wu, Yingru Li, "Spatial interpolation of temperature in the United States using residual kriging, "Applied Geography,Volume 44,2013, Pages 112-120, ISSN 0143-6228, https://doi.org/10.1016/j.apgeog.2013.07.012

Wenjing Cao, JinXing Hu and Xiaomin Yu, "A study on temperature interpolation methods based on GIS," 2009 17th International Conference on Geoinformatics, Fairfax, VA, USA, 2009, pp. 1-5, doi: 10.1109/GEOINFORMATICS.2009.5293422.

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	23
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	26
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	19
		100	94