Lab Report

Title: Lab 3 - Part 2

Notice: Dr. Bryan Runck

Author: Mattie Gisselbeck

Date: December 2, 2022

Repository: https://github.com/mgisselbeck/GIS5571

Time Spent: 15 hours

Abstract

The main objective of this lab is to build an ETL that scrapes the last 30 days of temperature data for all the NDAWN stations and compare three different interpolation methods (IDW, GPI, Kriging) using the NDAWN weather data. The data was sourced from Minnesota Geospatial Commons and was scraped through an ETL in ArcGIS Pro via a Python notebook.
The results are shown in the figures below (see Figure 3 through Figure 11). The data flow diagram below (Figure 1 & Figure 2) shows all the variables and commands I applied to compare the different interpolation methods. The results could be qualitatively verified with the findings of Cao, W., Hu, J., & Yu, X. (2009) when interpolating temperature from 327 weather stations. In this lab, I was able to learn about different interpolation methods and what method would be best suited for future project needs. The objectives of this lab helped me to gain practical applications of how I would conduct an IDW, GPI, and Kriging through ArcPy or an open-source package.

Problem Statement

The main objective of this lab is to build an ETL that scrapes the last 30 days of temperature data for all the NDAWN stations and compare three different interpolation methods (IDW, GPI, Kriging) using the NDAWN weather data (Runck, 2022). The analysis workflow should be able to collect real-time data and create an interpolated temperature map for the highs and lows of the last 30 days from NDAWN in real-time.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Weather Data	Average Temperature - Last 30 Days	XY Coordinates	Temperature (F)	NDAWN Center	ETL

Table 1. Required Data.

Input Data

The table below is a collection of data from the North Dakota Agricultural Weather Network (NDAWN). Data was scraped through an ETL in ArcGIS Pro via a Python notebook. All the data described below will be used to analyze the average daily temperature recorded at all the NDAWN stations in the last 30 days.

#	Title	Purpose in Analysis	Link to Source
1	NDAWN Weather (Average Temperature for the Last 30 Days)	The raw input dataset will be extracted into a .csv file to analyze the average daily temperature recorded at all the NDAWN stations in the last 30 days.	NDAWN Center

Table 2. Input Data.

Methods

The table below was designed with the intention to compare the different interpolation methods and find the most suitable method for interpolating temperature data. The information in the table was sourced from Ersi's documentation on classification trees of the interpolation methods offered in Geostatistical Analyst.

Category		IDW	GPI	Kriging
Type of Information	One Prediction per Location	✓	✓	✓
	Quantile Value			✓
	Many Predictions per Location			
	Predicted Values and Errors			✓
Measurement of Spatial Autocorrelation	Yes			✓
	No		✓	
	Implicit	✓		
Output Type	Prediction	✓	✓	
	Prediction Error			✓
	Probability			✓
Level of Assumptions	Few	✓	✓	
	Intermediate			
	Many			✓
Type of Interpolation	Exact	✓		
	Inexact		✓	✓
Smoothness	Smooth		✓	
	Intermediate			✓
	Not Smooth	✓		
Uncertainty of Predicted Values	Yes			✓
	No	✓	✓	
Processing Speed	Slow			
	Intermediate			
	Fast	✓	✓	✓

Table 3. Comparing IDW, GPI, and Kriging (Ersi, 2021).

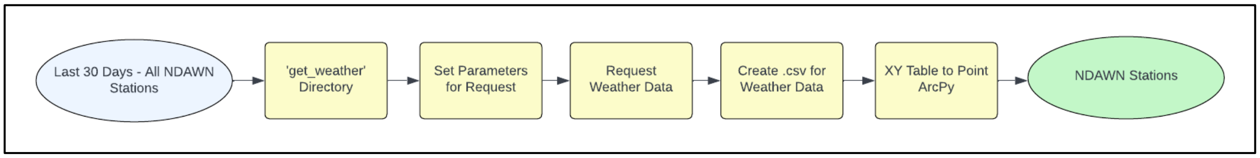

Figure 1. Extracting Weather Data from NDAWN.

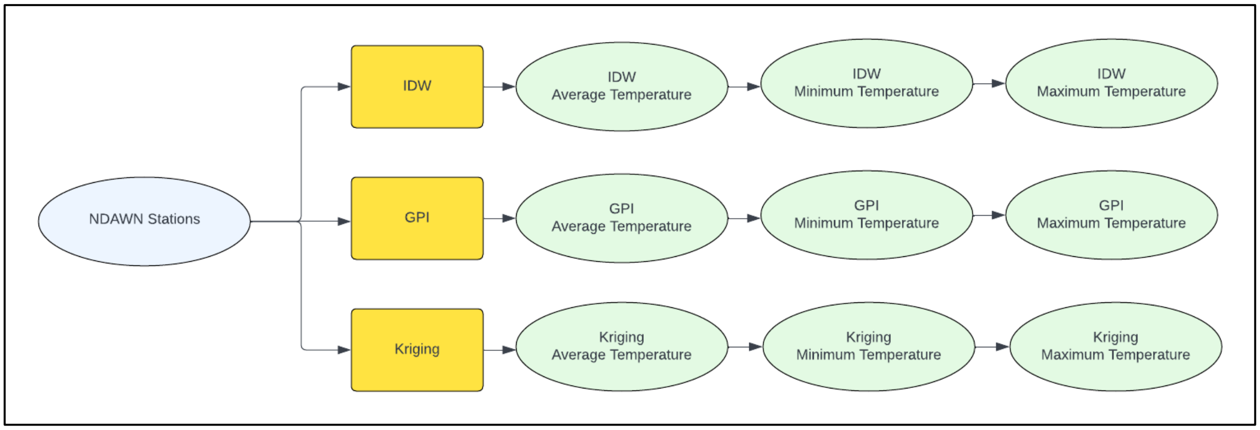

Figure 2. Comparing Interpolation Methods (IDW, GPI, and Kriging).

Part 2.1: ETL Pipeline for Extracting Weather Data from NDAWN

To extract data from NDWN, I set the downloadable .csv link on NDAWN as the 'url' variable (For the complete 'url' link, see Lab 3 – Part 2 Notebook). Then I requested the data through a 'get' function from importing the panadas package into the notebook. Following that I extracted and read the data from the CSV into a data frame. I skipped the first four rows since they were just information headers that weren't necessary for the analysis. After that, I created variable columns, cleaned the data frame, and aggregated the data. To export the cleaned data, I exported the data frame as a csv so it could be converted to a feature class with the ArcPy tool 'XY Table to Point'.

```
# Set Weather Data - Last 30 Days
url = "https://ndawn.ndsu.nodak.edu/table.csv?station
```

```
# Request Weather Data as a .csv
response = requests.get(url)


# Extract Weather Data into a .csv
file_name = r'C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\last_30_days.csv'
csv = open(file_name,'w')
csv.write(response.text)
csv.close()

# Extract CSV and Read into Data Frame
raw_df = pd.read_csv(file_name, header=3, skiprows=[4])

raw_df.rename(columns={'Unnamed: 0':'Station Name', 'deg':'Lat', 'deg.1':'Lon', 'Degrees F':'Max',
'Degrees F.1':'Min', 'Degrees F.2':'Avg'}, inplace=True)
raw_df.head()

# Create Copy of Data with Relevant Columns
columns = ['Station Name', 'Lat', 'Lon', 'Max', 'Min', 'Avg']

cleaned_df = raw_df[columns].copy()

# Aggregate Data
agg_functions = {'Lat':'first', 'Lon':'first', 'Max':'mean', 'Min':'mean', 'Avg':'mean'}
agg_df = cleaned_df.groupby(cleaned_df['Station Name']).aggregate(agg_functions)

agg_df.head()

# Export Aggregated Data Frame to .csv

agg_df.to_csv(r'C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\aggregated_temps.csv')

# Converting Weather Data to a Feature Class

csv_path = r'C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\aggregated_temps.csv'

temperature_features = arcpy.management.XYTableToPoint(csv_path, 'station_temperatures', 'Lon',
'Lat')
```

Part 2.2: Comparing Interpolation Methods (IDW, GPI, and Kriging)

To run the interpolation methods, I used their associated ArcPy tool. For each of the methods, I used the same number of variables, 141, due to number of weather stations equaling that sum. For the Kriging, I selected ordinary as the method and a circular as the semi-variogram. Results are shown in Figure 3 through 11.

```
# Inverse Distance Weighting (IDW)

arcpy.ddd.Idw("station_temperatures", "Avg",
r"C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\Lab3.gdb\IDW", 0.0144790399999999, 2,
"VARIABLE 141", None)

# Ordinary Kriging (Circular)

arcpy.sa.Kriging("station_temperatures", "Avg", "Circular 0.014479 # # #", 0.0144790399999999,
"VARIABLE 141", None, r"C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\Lab3.gdb\kriging")

# Global Polynomial Interpolation (GPI)

arcpy.ga.GlobalPolynomialInterpolation("station_temperatures", "Avg", "gpi_layer",
r"C:\Users\gisse015\Documents\ArcGIS\Projects\Lab3\Lab3.gdb\gpi_raster", 0.0144790399999999, 1,
None)
```

Results

The results are shown in the figures below (see Figure 3 through Figure 11). While running the interpolation tools in ArcGIS Pro, it was clear that setting the number of variables to how many points there were helped to smooth out the interpolation raster. Also, I noticed that the higher the Alpha value was, the rougher the IDW outputted as.

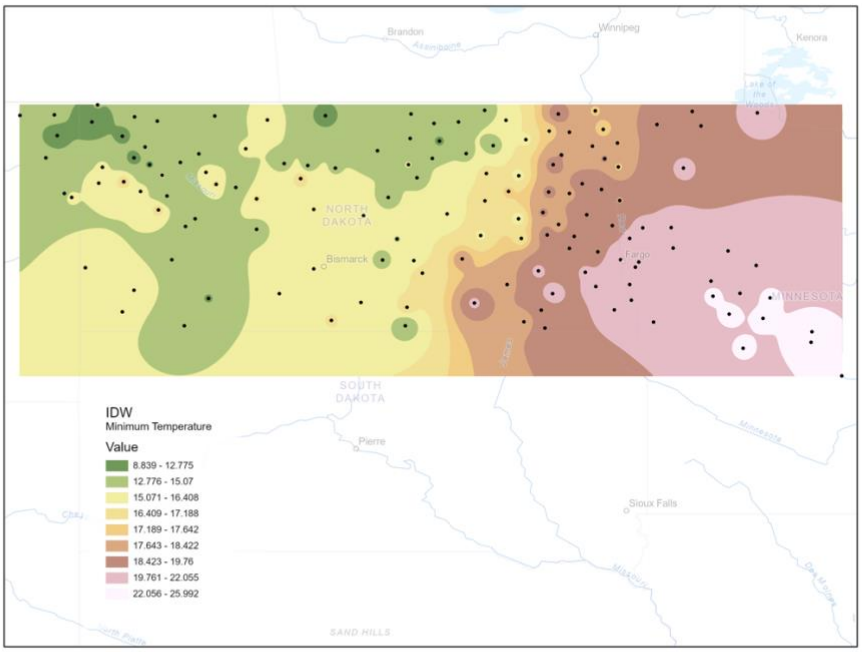

Figure 3. Inverse Distance Weighting (IDW) – Minimum Temperature.

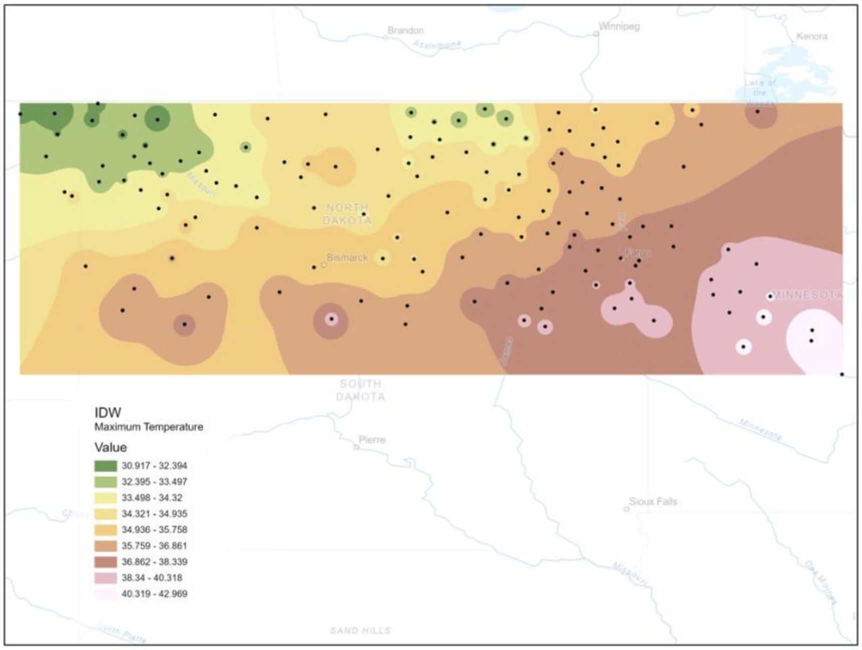

Figure 4. Inverse Distance Weighting (IDW) – Maximum Temperature.

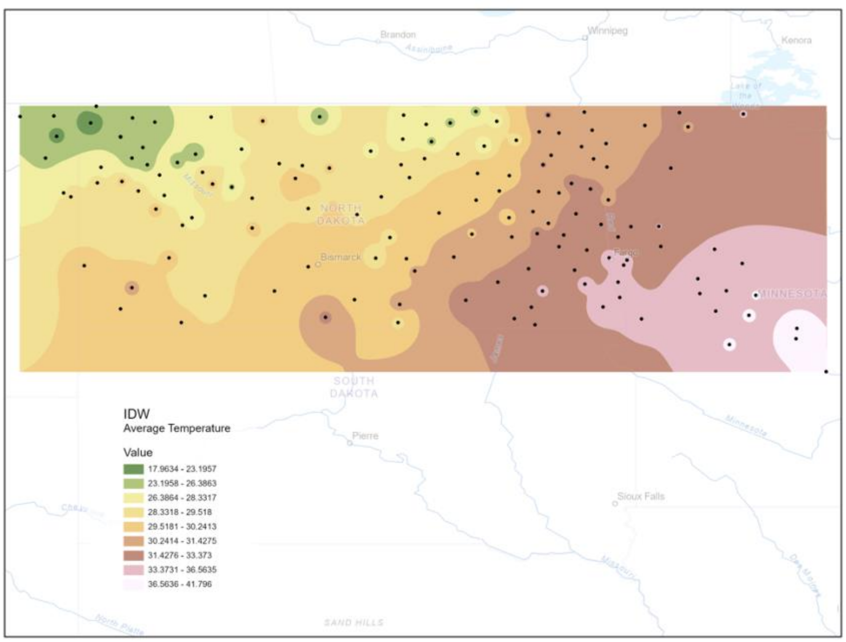

Figure 5. Inverse Distance Weighting (IDW) – Average Temperature.

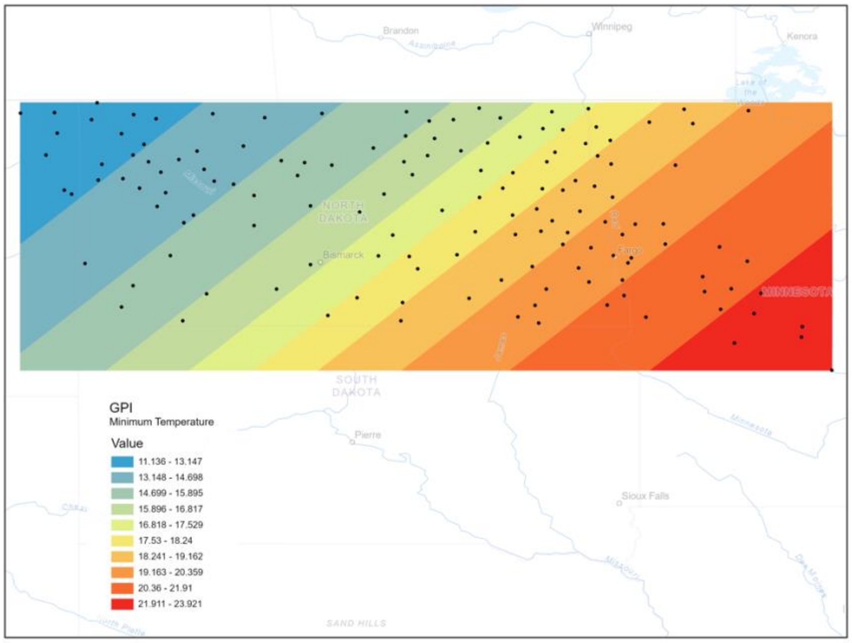

Figure 6. Global Polynomial Interpolation (GPI) – Minimum Temperature.

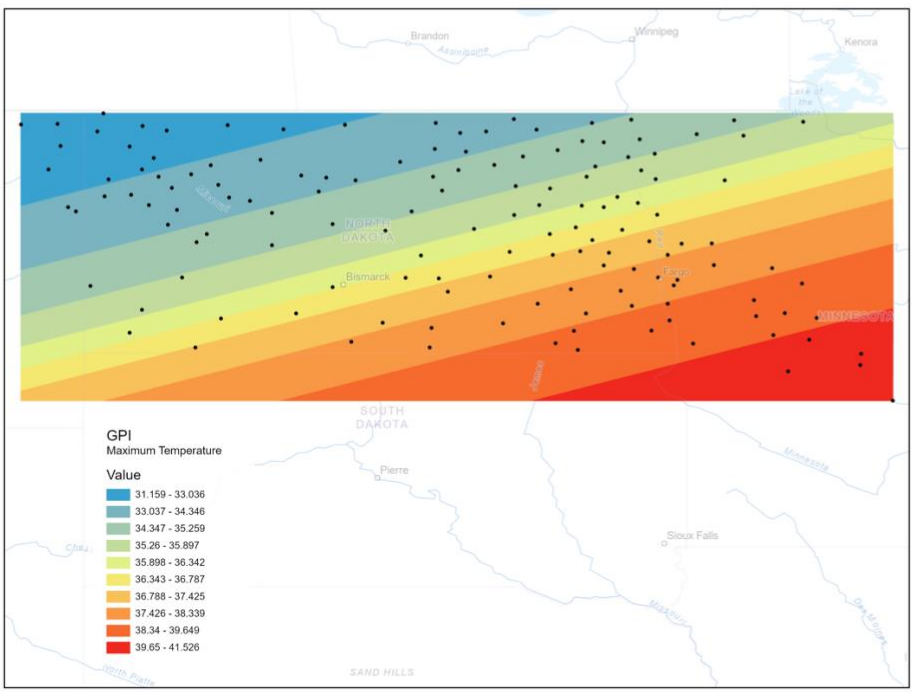

Figure 7. Global Polynomial Interpolation (GPI – Maximum Temperature.

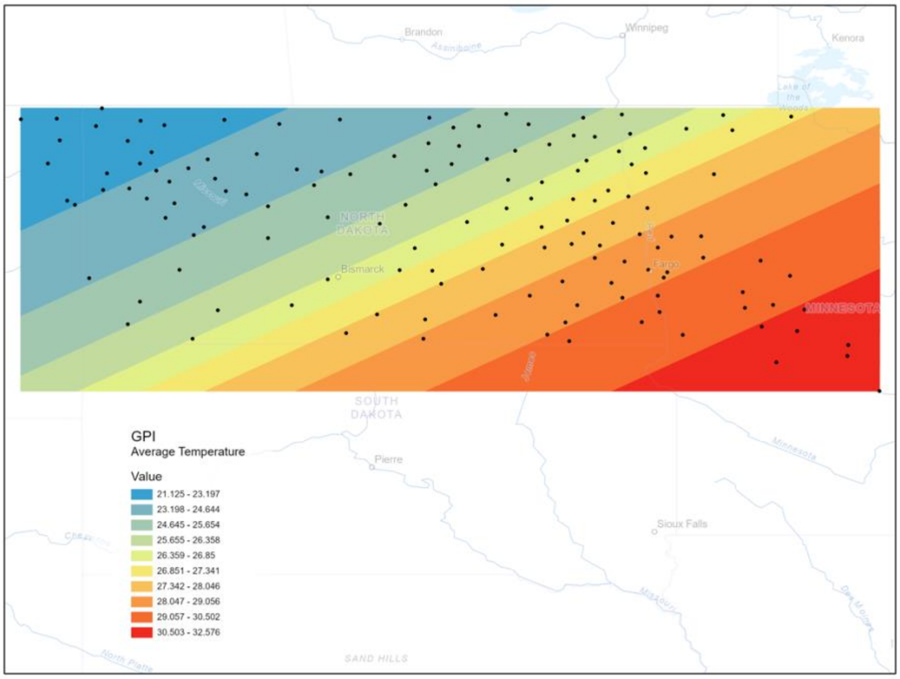

Figure 8. Global Polynomial Interpolation (GPI) – Average Temperature.

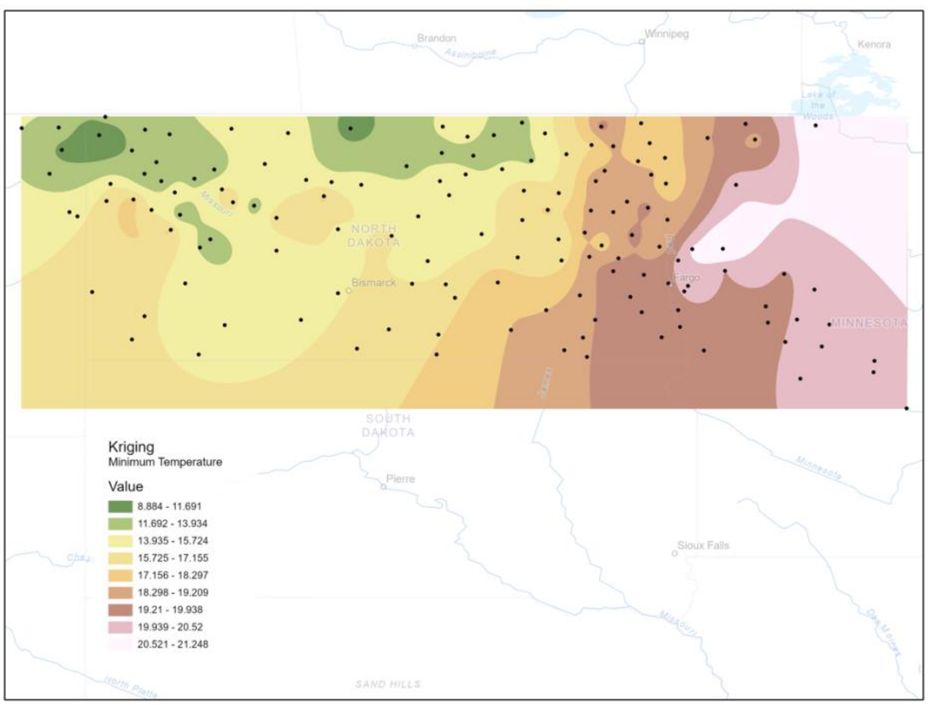

Figure 9. Kriging – Minimum Temperature.

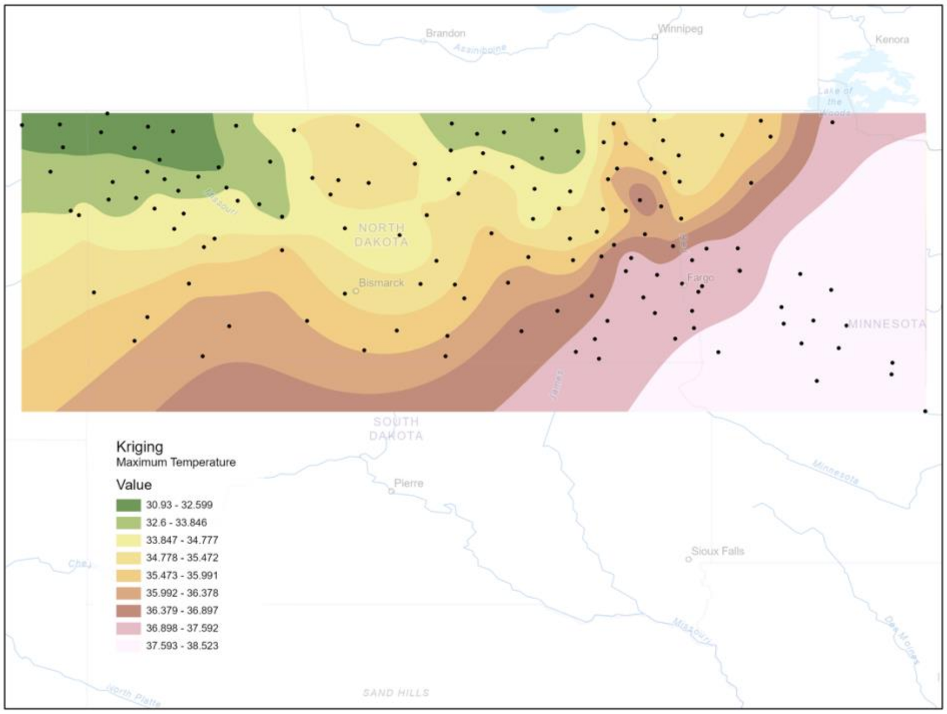

Figure 10. Kriging – Maximum Temperature.

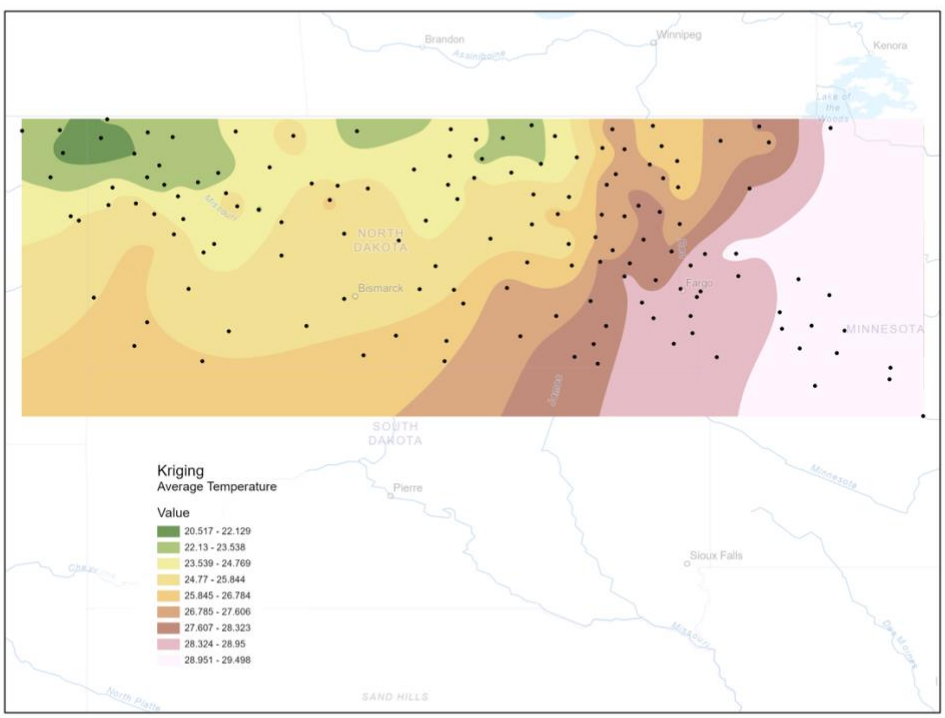

Figure 11. Kriging – Average Temperature.

Results Verification

The results could be qualitatively verified with the findings of Cao, W., Hu, J., & Yu, X. (2009) when interpolating temperature from 327 weather stations. Their results showed that Kriging Spherical and Exponential have the highest accuracy while IDW is less accurate in comparison. They also found that Gaussian Kriging is the least accurate.

Discussion and Conclusion

The Kriging Ordinary Method with the semi-variograms (Spherical, Circular, Exponential and Linear) proved to yield the best model when interpolating temperature data. The other interpolation methods failed to output an accurate visualization (i.e., distorted values, roughness). In this lab, I was able to learn about different interpolation methods and what method would be best suited for future project needs. The objectives of this lab helped me to gain practical applications of how I would conduct an IDW, GPI, and Kriging through ArcPy or an open-source package. Fortunately, I didn't run into any roadblocks while completing the lab.

References

Cao, W., Hu, J., & Yu, X. (2009). A Study on Temperature Interpolation Methods Based on GIS. 17th International Conference on Geoinformatics. https://ieeexplore.ieee.org/document/5293422

Ersi. (2021). Classification Trees of the Interpolation Methods Offered in Geospatial Analyst. ArcGIS Desktop. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

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	**27**
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	**98**