Lab Report

Title: Lab 3 Part 2
Notice: Dr. Bryan Runck
Author: Eric Gibson
Date: 11/28/2023

Project Repository: https://github.com/gibso632/GIS5571/tree/main/Lab3
Google Drive Link:
Time Spent: 16

Abstract

This lab focuses on various interpolation techniques and how they compare with each other regarding the structure of the output and their accuracy. The three techniques used within this analysis included inverse distance weighting (IDW), splining, and ordinary kriging. The data used was pulled from the North Dakota Agricultural Weather Network (NDAWN) from the last thirty days depicting each weather station location as points for performing this interpolation analysis. Overall, it was determined that kriging was the most accurate interpolation technique, as the output raster dataset depicted accurate temperatures without favoring areas near each point. The least accurate interpolation technique was determined to be splining, as the output dataset was too smooth along the surface, causing various temperatures to cover much broader ranges than the IDW or kriging techniques.

Problem Statement

Table 1. Requirements and deliverables for this lab.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Weather station data	Point data taken from an NDAWN API spreadsheet.	Point feature class	Maximum, minimum, and average temperatures	NDAWN 30-day temperature data	Set up spreadsheet and convert coordinates to points
2	IDW, Splining, and Kriging raster dataset	Raster determining continuous temperature throughout North Dakota	Raster dataset	Mean maximum, average, and minimum daily temperatures over the last month	Originated from point data presented in above cell	Dissolve points and average temperature data, then interpolating

Input Data

The data used within this lab is temperature data taken from NDAWN weather stations throughout North Dakota and portions of Minnesota. This data was brought in through a CSV file via an API call from the NDAWN website and converted to point data within ArcGIS Pro. The temporal range of the temperature data spans one month from, in this lab's case, October 21, 2023 to November 21, 2023. The temperatures include daily highs, averages, and lows, which were separated to determine the means of each over the past month. The points were then used to compare three different interpolation techniques with the three separate temperature averages monitored.

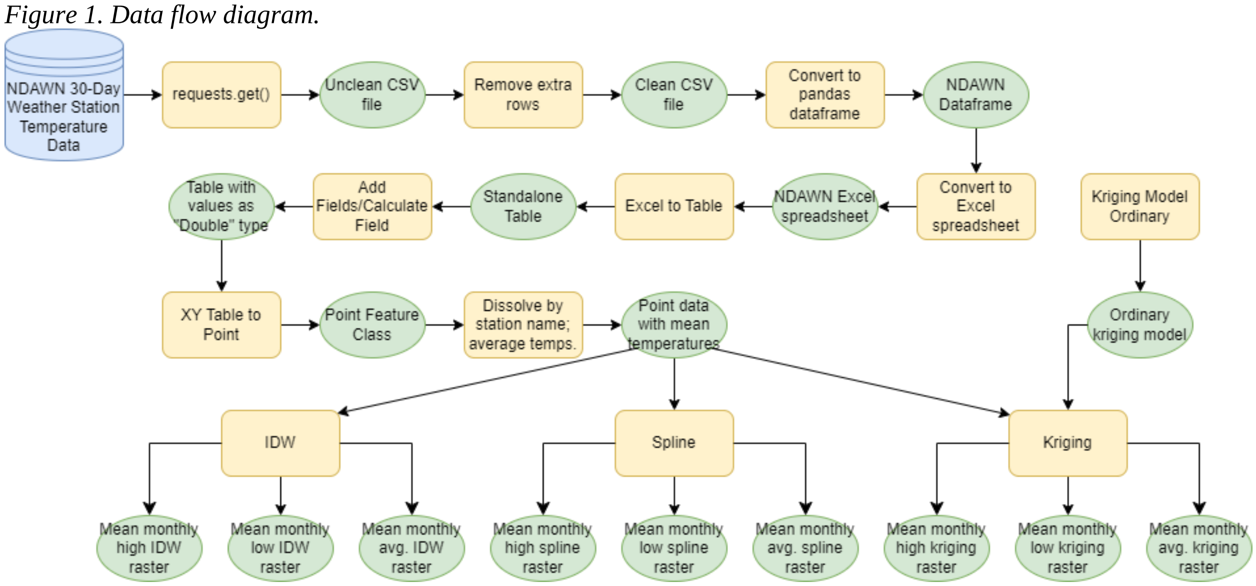
Although not a dataset used for analysis within this lab, it is also worth noting two journal articles and other documentation was used for determination of which interpolation technique works best regarding continuous temperature data. The first article describes the utilization of five different interpolation techniques for determining temperature data throughout China, providing favorable insight to interpolation method accuracy in this lab. This article found kriging (specifically spherical and exponential kriging) to be most accurate. The second article determines precipitation values over the province of Ontario, Canada. While not an exact match with this lab, as precipitation values are unique from temperature values, the article still provides helpful insight regarding interpolation technique accuracy. It was found local polynomial interpolation was most accurate with ordinary and universal kriging also described as valid interpolation methods.

Table 2. Dataset used within the lab for analysis.

#	Title	Purpose in Analysis	Link to Source
1	NDAWN Temperature Dataset	Creating points to perform various types of interpolation.	[NDAWN 30-day temperature data]

Methods

Figure 1. Data flow diagram.



As visualized in the data flow diagram above, the primary portion of this lab was creating the data pipeline and cleaning the data, as much of it could not be used for an interpolation analysis. Initially, the data was gathered via a CSV file, meaning it would need to be converted to a standalone table within ArcGIS Pro to create the point feature class first. The data was also not the correct type for calculating averages through the **Dissolve** tool, so additional fields were created and the **Calculate Field** tool through arcpy was utilized to provide the same temperature and coordinate values as a “Double” type instead of a “Text” type. Once that was completed, the coordinate data within the standalone table was used to create points via the **XY Table to Point** tool via arcpy and the point feature class was created. A **Dissolve** was then performed to match all of the multipoints to one individual point based on the weather station the point referenced. The mean value of the daily temperatures over the past month were found through the **Dissolve** as well. Finally, the mean high, low, and average daily temperatures over the last month were interpolated through IDW, splining, and kriging performed by the **IDW**, **Spline**, and **Kriging** tools through arcpy.

Results

Due to three separate interpolation techniques being used along with three separate averages of temperature values (highs, lows, and averages), a total of nine different raster datasets were created. All temperatures are daily averages taken from October 21, 2023 to November 21, 2023. These rasters are present below and labeled in the caption of each screenshot.

Figure 2. IDW of mean daily average temperatures.

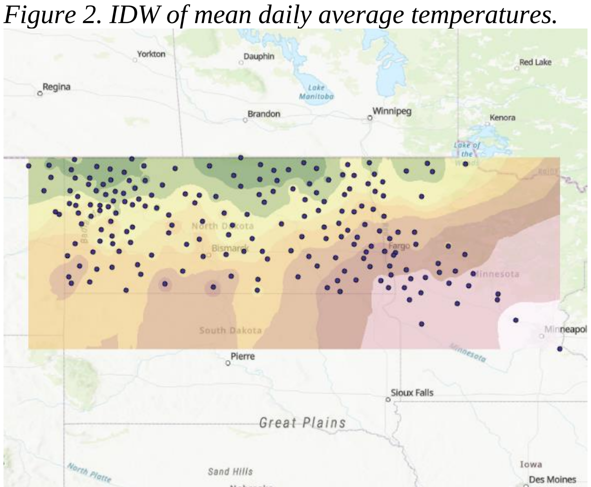


Figure 3. Spline of mean daily average temperatures.

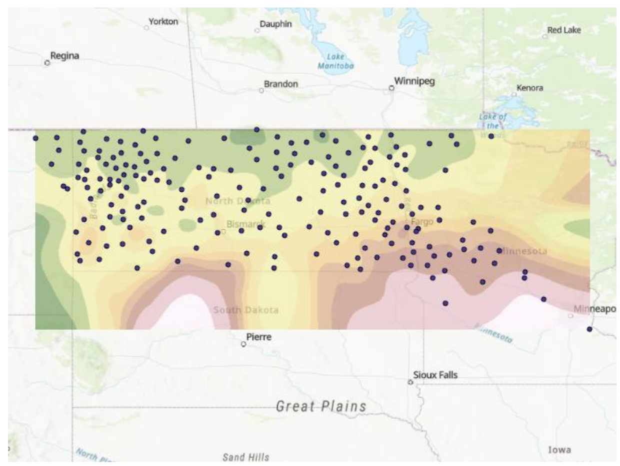


Figure 4. Krige of mean daily average temperatures.

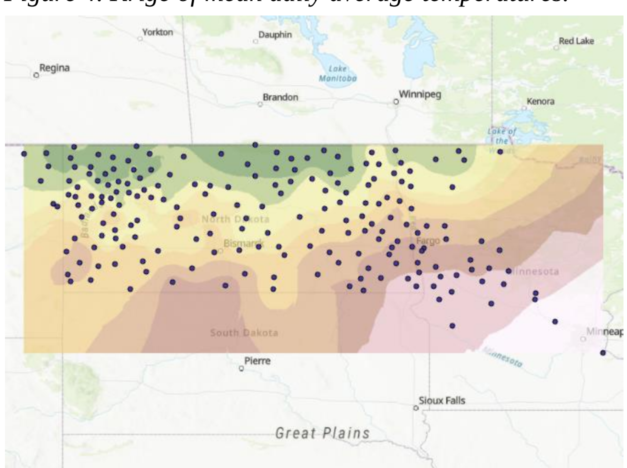


Figure 5. IDW of mean daily high temperatures.

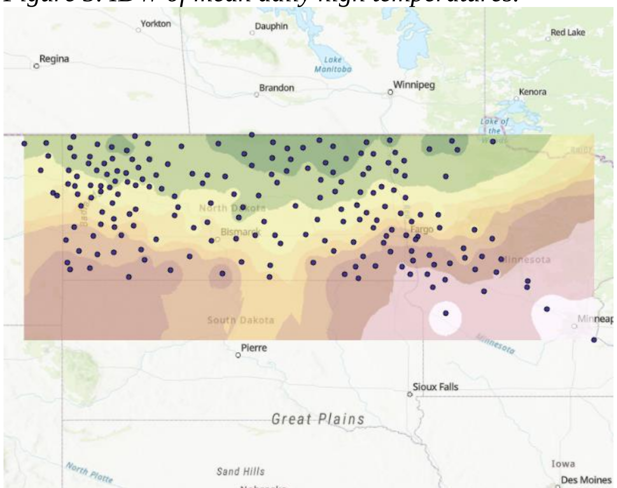


Figure 6. Spline of mean daily high temperatures.

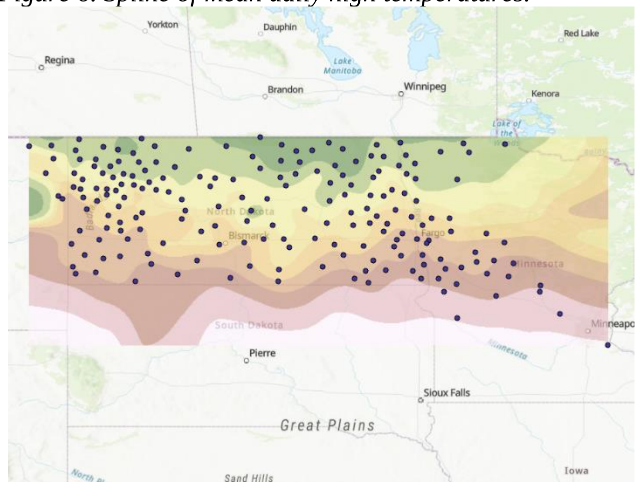


Figure 7. Krige of mean daily high temperatures.

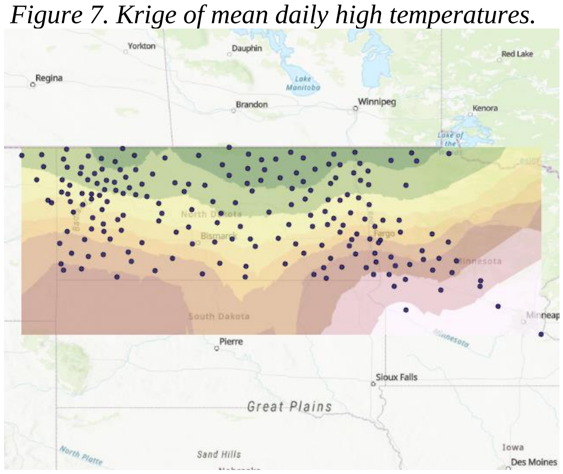


Figure 8. IDW of mean daily low temperatures.

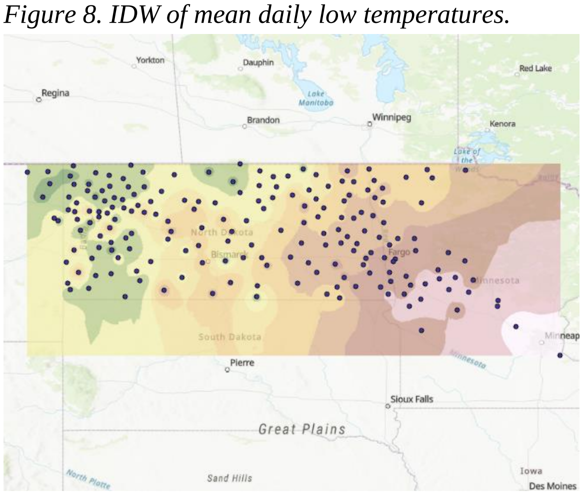


Figure 9. Spline of mean daily low temperatures.

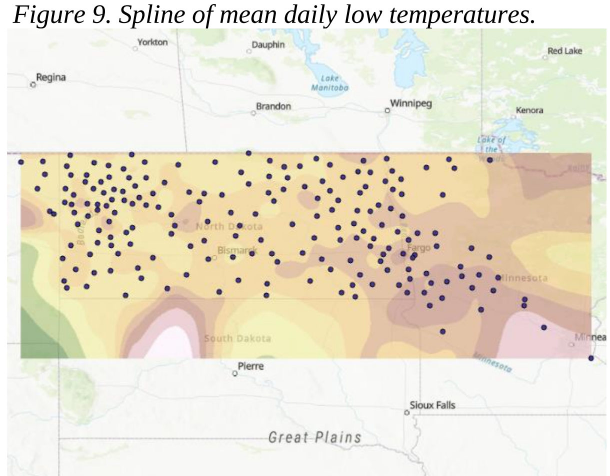
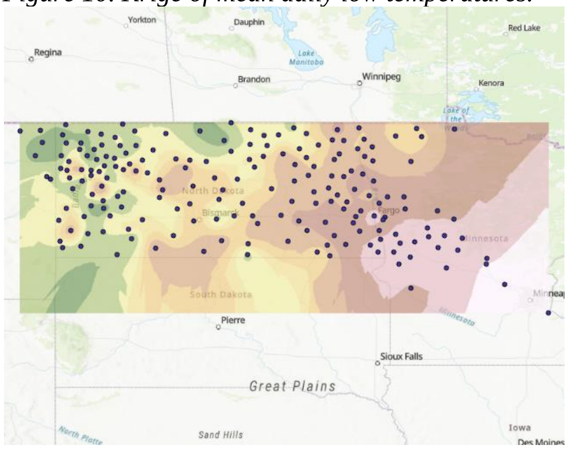


Figure 10. Krige of mean daily low temperatures.



Each raster has the same symbology, though the values each color references is different depending on the ranges of the highs, lows, and averages. Still, the reds and whites correspond to warmer temperatures and the yellows and greens correspond to lower temperatures. Each raster was also grouped together based on the data they are referencing, so **Figures 2-4** only depict daily averages, **Figures 5-7** only depict daily highs, and **Figures 8-10** only depict daily lows. Each of these groups reference the same data to make it simpler to visualize the difference between each interpolation method. It is also worth noting that the spread of each point is not consistent, as North Dakota contains points with relative spatial consistency throughout the state, but stations in Minnesota are much fewer and further between. This manipulates the raster by illustrating there to be much less change of temperature over space within Minnesota.

Results Verification

Through visualization of the output rasters, it is clear the correct interpolation algorithms were used to display continuous temperature data from the daily mean highs, lows, and averages over the past month. IDW uses a deterministic interpolation technique meaning the input values from the points and the power for dictating how much distance is weighted are set. There are no random functions used to estimate a more gradual change over distance and area. This creates a raster which changes dramatically around each point and contains more even values the further areas are from a point. This is incredibly obvious when observing the IDW rasters, as they tend to create “circles” of the same or similar values around points. Due to this issue with IDW interpolation, it is likely not the greatest option for creating a raster dataset depicting changes of temperature over a surface.

Splining “smooths” out the data by forming a raster which “flows” through each point using a polynomial which references each datapoint. This leads to a more gradual change from one point to another, like a parabola for example. When viewing cross-sections between points, the nature of the splined raster would closer to a series of parabolas due to the polynomial equation it utilizes. This is more pleasant to view and avoids the issue of weighting points too harshly like IDW interpolation, however it is also inaccurate because the real world does not change so gradually. Many areas in spline interpolation cut through the actual values in favor of a more smooth surface, so it is also likely not an acceptable option for showing continuous temperature data.

Finally, kriging was used and is mostly likely the most accurate due to its middle-ground between weighting points and smoothing areas between points too much. Kriging uses semivariograms to determine the spatial autocorrelation and variance between points to find values in areas surrounding the data points. This is fairly visible within the result screenshots, as the rasters created via kriging seem to show some areas of smoothing between points, while other areas show a drastic change in temperature surrounding a weather station. This is much more authentic to temperature data on Earth. Unlike IDW and splining, kriging determines the variance between the values in the dataset, providing an extra layer of accuracy to the output raster dataset. This combination of distance relationships between points and variance in their values provides an extremely accurate visualization of temperature data.

Discussion and Conclusion

One article referenced for this lab compares five different interpolation techniques: IDW, splining, and three different types of kriging (exponential, spherical, and Gaussian). The researchers even determined the accuracy of interpolation techniques based on mean lows, highs, and averages like what was performed within this lab, however the data was only taken over ten days at the start of January, April, July, and October. Both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to determine interpolation accuracy. While IDW did provide relatively accurate data regarding temperatures throughout China, two methods of kriging were determined to be most accurate in their analysis: exponential and spherical kriging. Splining was determined to be the least accurate in general with IDW behind exponential and spherical kriging (Cao 2009).

Another article compares six different interpolation techniques: IDW, global polynomial interpolation (GPI), local polynomial interpolation (LPI), radial basis functions (RBF), ordinary kriging (the kriging used within this lab), and universal kriging. It is worth noting ordinary and universal kriging output the same dataset in this context, so they will both be referred to as simply "kriging". Overall, in this study LPI was found to be the most accurate interpolation technique depicting precipitation totals through Ontario, Canada. Interestingly, LPI is fairly similar to splining in that it uses polynomial functions, however the output dataset does not move through each datapoint. LPI interpolation is inexact and, thus, the output surface raster does not equal the value of each datapoint at that point's location. This likely allows for less dramatic changes through smoothing and led to precipitation totals being most accurate when measured via LPI. Kriging was also determined to be accurate with IDW and RBF showing the most inaccuracy. Also note the data used within this study was precipitation data, which is not a perfect comparison to the temperature data interpolation within this lab (Wang 2014).

Ultimately, this lab provided fantastic information regarding interpolation techniques and how and when to choose them based on the input data. Different forms of kriging do seem to provide the most accuracy depending on the situation due to the utilization of variograms along with weighted distances, however IDW and splining do have benefits as well. IDW could work well with data containing many points over an areas where they are all relatively close to each other. Since the output is so heavily weighted towards the actual point itself, the output should be accurate in this regard. Additionally, splining works well when the viewer wants an output which is pleasant to view and models relatively accurate elevation data for smoother slopes. Each interpolation method has its purpose and should be determined based on the factors outlined within this lab.

References

North Dakota Agricultural Weather Network. “NDAWN - North Dakota Agricultural Weather Network.” *NDAWN Center*, ndawn.ndsu.nodak.edu/.

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.

Wang, S., Huang, G.H., Lin, Q.G., Li, Z., Zhang, H. and Fan, Y.R. (2014), Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada. Int. J. Climatol., 34: 3745-3751. https://doi.org/10.1002/joc.3941

ESRI. “ArcGIS Pro Geoprocessing Tool Reference.” *ArcGIS Pro Geoprocessing Tool Reference-ArcGIS Pro | Documentation*, 2023, pro.arcgis.com/en/pro-app/latest/tool-reference/main/arcgis-pro-tool-reference.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	**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	**22**
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	**17**
		100	**87**