Lab Report

Title: Lab 3, Part 3 Repot
Notice: Dr. Bryan Runck
Author: Kyle Olson
Date: 11/26/2024

Project Repository: https://github.com/kolson5581/GIS5571/tree/main/Lab%203
Google Drive Link:
Time Spent: 3 hours

Abstract
This lab report and associated Jupyter Notebook contain methods to create a pipeline for data from the NDAWN weather station network, and create interpolated temperature surfaces with the data. Three different methods of interpolation were used in this lab: Inverse distance weighting, Local polynomial interpolation, and ordinary Kriging. Maps using each of these methods display information about the past 30 days and the temperature averages, minimums, and maximums within that time frame. Discussion of the characteristics of these interpolation methods concludes the report.

Problem Statement
This lab requires the creation of an ETL pipeline for NDAWN data, which can then be used to test different interpolation methods for visualization of temperatures across the entire region where NDAWN has sensors.

Table 1. Requirements

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	ETL Pipeline	Pipeline to bring in past 30 days of NDAWN Temperature data	Locations of weather stations	Temperatures	NDAWN	Consolidate 30 days of data into each point
2	Map of weather stations	Average temperature data for each weather station	Locations of weather stations	Average Temperature	NDAWN (after processing)	Label with 30 day average
3	Interpolate temperature data between weather stations	Use three different methods of interpolation to predict temperature values between stations			NDAWN	Select interpolation methods
4	IDW and Kriging interpoalted surfaced	Must use these two methods for two of the three surfaces generated	Uses points gathered from NDAWN	Temperature averages, mins, maxs	NDAWN	Use arcpy commands

5	Additional interpolated surface	Use another interpolation method using Esri's decision guide	Uses points gathered from NDAWN	Temperature averages, mins, maxs	NDAWN	Use arcpy commands

Input Data
All data for this section of Lab 3 is drawn from NDAWN. The data is from the last 30 days, and covers all of the NDAWN stations available. The request url for this data uses a "quick pick" tag to ensure that the last 30 days of data are always called. To meet the requirements of the lab assignment, the data that is included is the average daily temperature, the minimum daily temperature, and the maximum daily temperature. The data is called as a csv file.

Table 2. Input Data

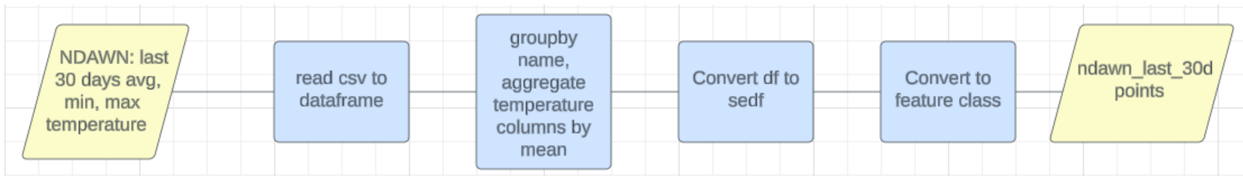
#	**Title**	**Purpose in Analysis**	**Link to Source**
1	NDAWN Temperature Data	Provides daily maximum, daily minimum, and daily average temperature data for all NDAWN weather stations	NDAWN last 30 days (Web version)

Methods
Figure 1, below, shows the data flow diagram for the data cleaning portion of the notebook. First, I used the NDAWN interface to create a table that provides the daily average temperature, daily minimum temperature, and daily maximum temperature for each of the 218 weather stations in the NDAWN network. Below is the url I used, with the list of stations shortened. Highlighted in blue is the key term that tells us we are requesting a .csv file, and highlighted in orange is the key term that says the frequency of data (daily), and that I want the last 30 days.

"https://ndawn.ndsu.nodak.edu/table.csv?station=78&...&station=110&variable=ddmxt&variable=ddmnt&variable=ddavt&year=2024&ttype=daily&quick_pick=30_d"

From there, I used pandas to read the csv directly into a dataframe, with no need to save it to disk. I then use the pandas groupby and aggregate functions to group each weather stations rows together, while aggregating the temperature values by the mean. This gives us the mean of the average temperatures for all 30 days, which I use to create the first map, and then the mean of minimum and maximum temperatures over the past 30 days. From there, I converted the dataframe into a spatially enabled dataframe, using the longitude and latitude fields to spatialize the points. Lastly I convert this sedf into a feature class which contains all the points, each of which has the required data I need for the rest of the lab.

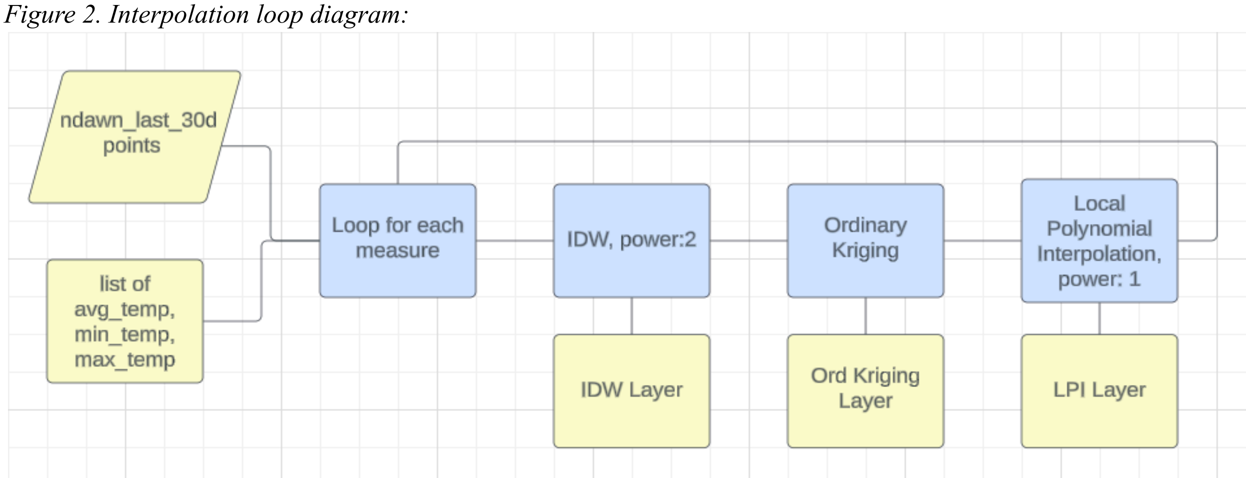
Figure 1. Data cleaning data flow diagram:



From here, Figure 2 shows the process by which I used these points to create the interpolated surfaces. I created a list of the field names which will be used to create the different versions of the maps. Then I use this name, along with the points which have those field names, to create three surfaces: and Inverse distance weighted layer, an ordinary kriging layer, and a Local Polynomial Interpolation layer.

I chose Local Polynomial Interpolation as my third method based on the Esri decision trees provided in the lab instructions. LPI provides one prediction per location, it doesn't require a spatial autocorrelation figure (which I didn't have), and is an intermediate level of complexity and smoothness (Classification trees of the interpolation methods offered in Geostatistical Analyst, Esri). I wanted to try a method I was unfamiliar with, as well. I tried the method with a power of 2 and 1, but decided to use 1 for the final version of the analysis because the result was generally more similar to the other methods of interpolation. The power of 2 version looked like it was less able to handle the irregular spacing and caused errors in the areas with no weather station points.

Figure 2. Interpolation loop diagram:

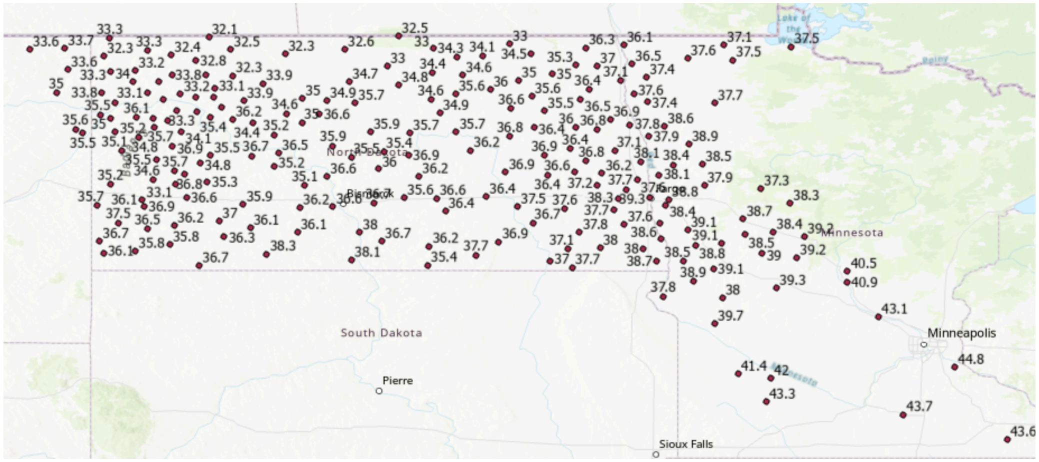


Lastly the notebook ends with a quick for loop to get all of the interpolation methods to use a consistent classification and color scheme within each category of measurement (eg. all the minimum temperature surfaces used the same breaks and color schemes).

Results

Figure 3 shows the first output - weather stations with the average monthly temperature. These values were determined by taking the average of all the averaged daily temperatures for the last 30 days.

Figure 3: Weather station locations and average temperatures in the past 30 days



The next section will include the interpolation results for average temperatures, minimum temperatures, and maximum temperatures. For these maps I have left the weather station points on the map so spatial patterns can still be observed, but removed the label with the average temperature to de-clutter the maps.

Figure 4: Average temperature using IDW

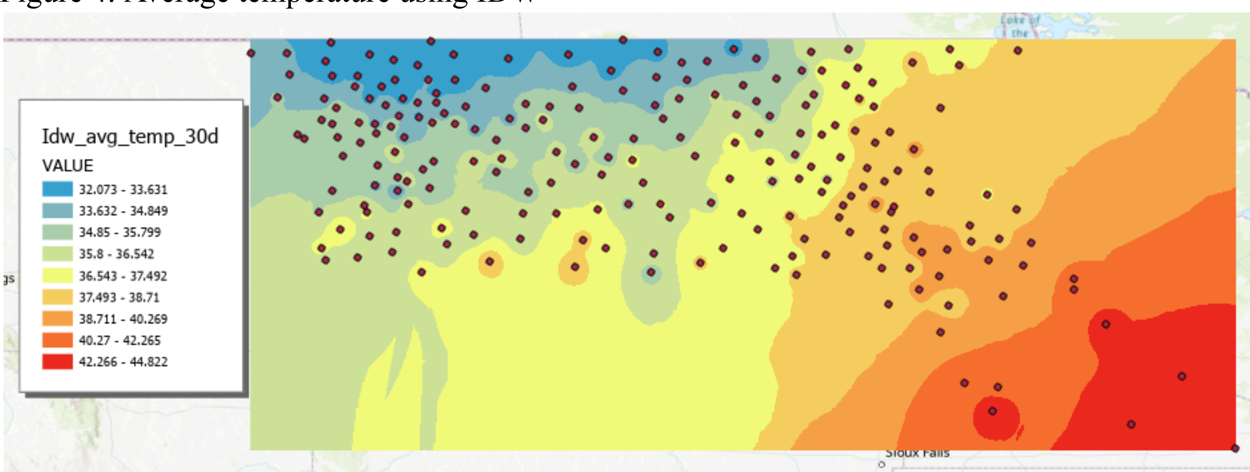


Figure 5: Average temperature using Ordinary Kriging

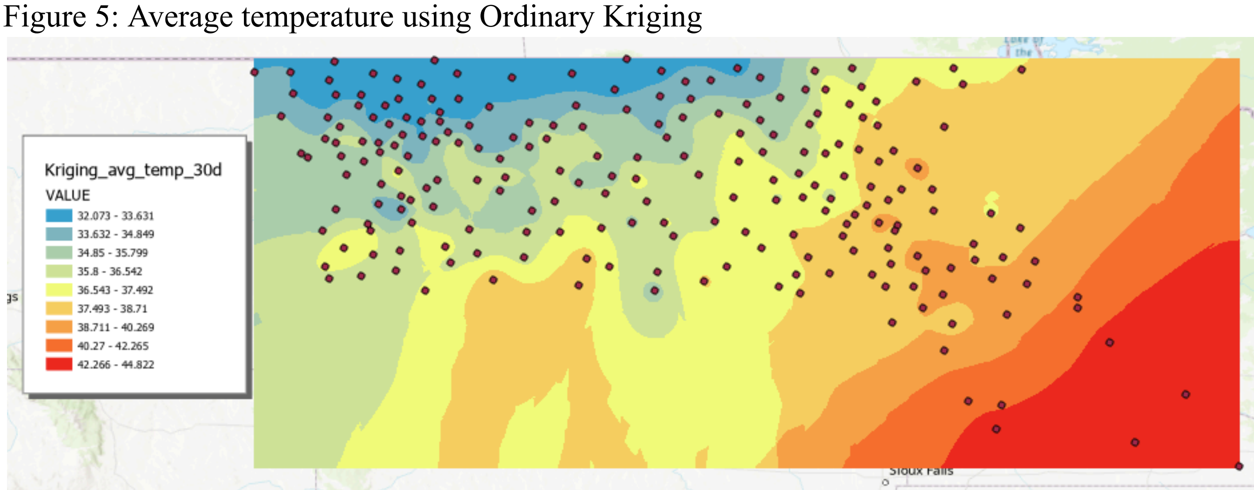


Figure 6: Average temperature using LPI - power 1

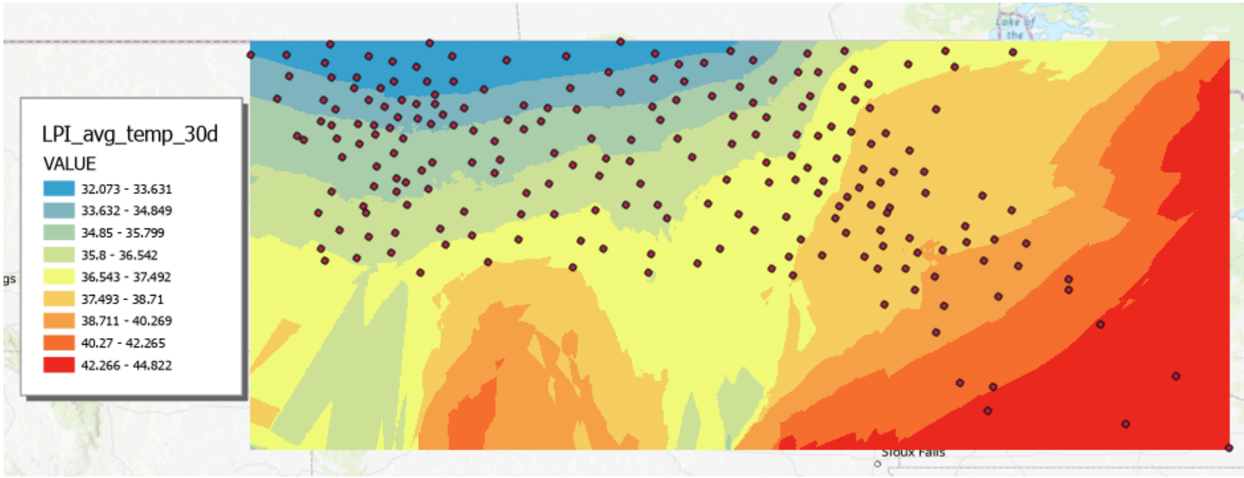


Figure 7: Minimum temperature using IDW

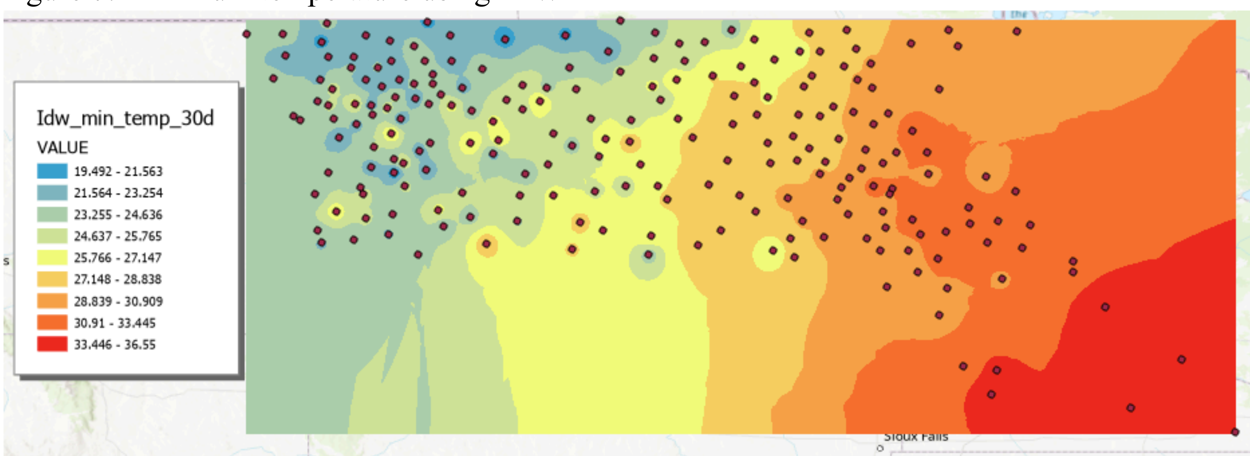


Figure 8: Minimum temperature using Ordinary Kriging

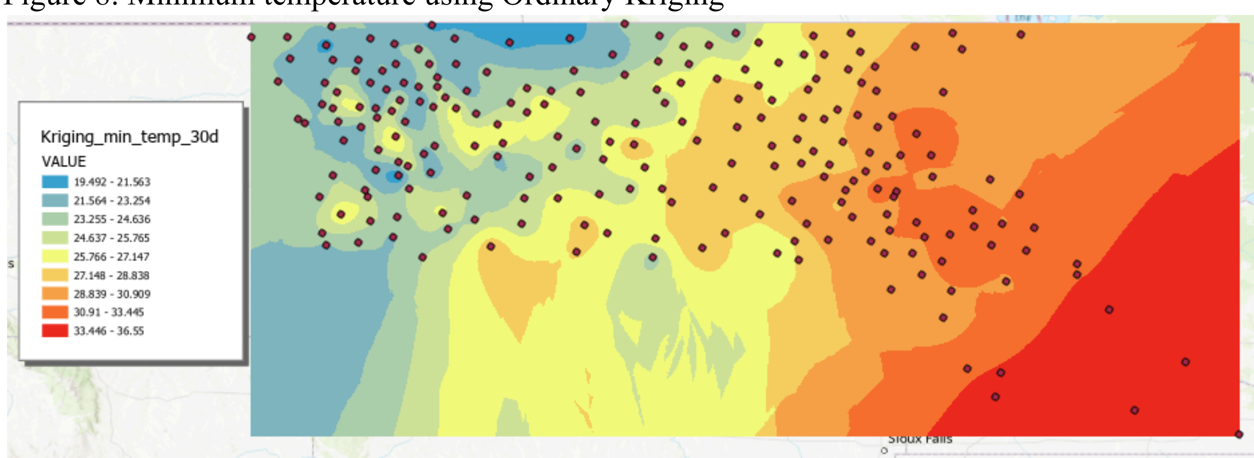
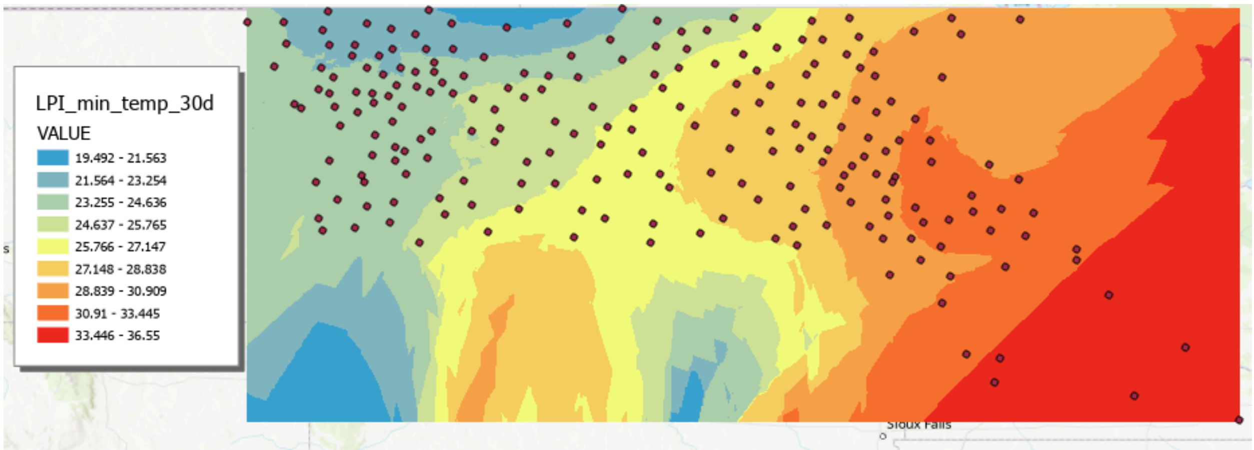


Figure 9: Minimum temperature using LPI - power 1



In these first two sets of maps, we see that generally the only times we see weather station points completely circled by one color is in the IDW and the Kriging methods, while the LPI surfaces are generally more smooth. This is likely because the IDW and Kriging are considering the distance to the nearest point more than the LPI method. The LPI method, meanwhile, seems to fill in the wide open spaces with no points with more differing values than IDW, in particular.

Figure 10: Maximum temperature using IDW

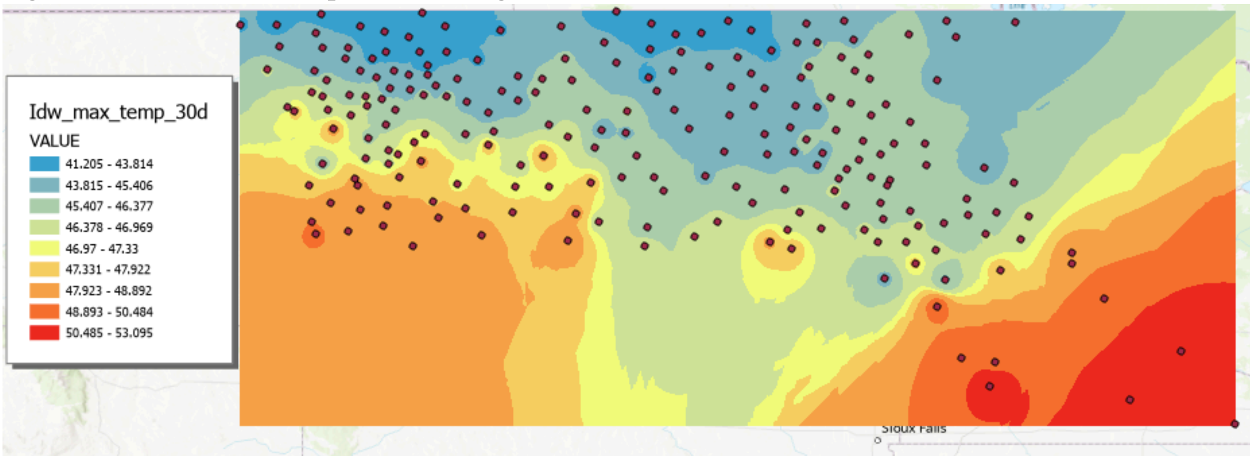


Figure 11: Maximum temperature using Ordinary Kriging

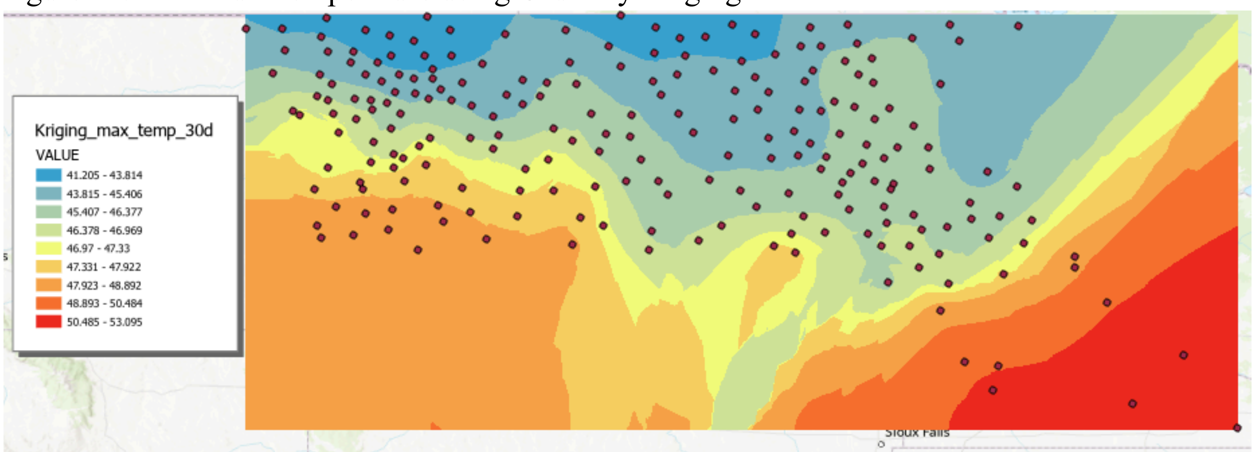
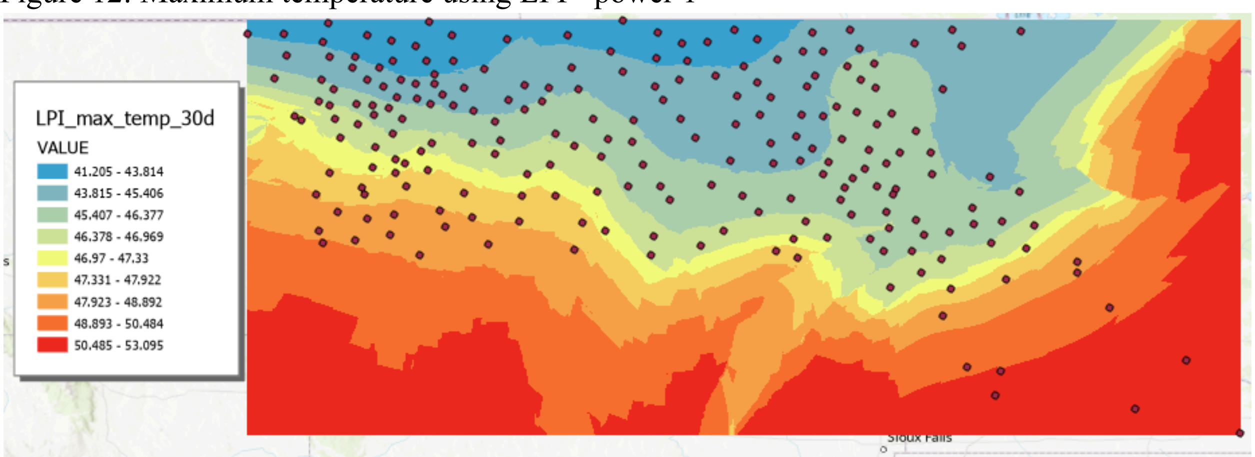


Figure 12: Maximum temperature using LPI - power 1



Here too we see situations in the IDW surface where a single point (or small group of points) is surrounded by just one color, while the Kriging and LPI surfaces tend to be more smooth. Again, this is likely due in part to the power figure used for the IDW. The LPI and Kriging functions are generally expected to be more smooth (Classification trees of the interpolation methods offered in Geostatistical Analyst, Esri).

However, while there are differences across the three methods, all three of these maps show broadly similar patterns within each type of measurement. The edges of the LPI maps are all more jagged, and we see the LPI maps react differently to the open spaces in the southwest corner of the map - trying to create more variance than the other methods.

Results Verification

The results for these figures are all based on one source of data, and each interpolation method is using the same input data (though they have different parameters). Thus, we can be certain that the input data is at least consistent, so each of these operations should be correct relative to one another.

Additionally, I conducted quality assurance on the calculated average temperatures by spot checking several of the weather station average values. I can be certain that the results of that basic data wrangling operation is correct. Lastly, I confirmed that the data in the csv and my dataframe matches the data in the table on the NDAWN website when checked manually.

Discussion and Conclusion

One of the core questions of this lab is the impact of the different interpolation methods. Ultimately, the maps all are using the same input data, so it is unlikely we would see vastly different maps. Nonetheless, there are differences to be observed. IDW schemes rely on distance as the primary driver of the values - we can see this quite clearly in the maps we generated, as all three IDW maps have areas where a single point is contained within one color band. Kriging and LPI both rely on more complex fitting methods to create a generally more smooth surface - this is born out in observations of all three maps.

The other major difference is how the various methods handle the open space in the Southwest and Northeast of the map area, where there are fewer stations feeding in measurements. LPI is the method with the most variation within these open spaces, while Kriging varies more than IDW, but still has less jagged barriers between classes.

The literature about what type of interpolation is “best” for temperature readings seems to vary based on factors like spatial scale, temporal resolution, and the space between weather stations. However, I found several studies which show that some variation of a Kriging method is generally the best for accuracy of temperature. Cao, Hu & Yu (2009) used GIS tools and weather station data from China to compare the results of different interpolation methods and found that two variations of Kriging: spherical and exponential, are the highest accuracy methods of interpolation for temperature. Brunetti et al (2014) compared three methods of temperature interpolation in Italy and found that regression Kriging, while not the highest performing model they tested, performed nearly as well as their preferred method , LWLR. Hofsta et al (2008) also found that Global Kriging performed best among six methods they tested across Europe.

It seems that there is consensus that Kriging methods are generally accepted as the most useful for temperature interpolation. While this is hardly an exhaustive review of the literature, it is consistent with the results I found. Additionally, it is important to note that there are a wide variety of Kriging methods, and each method has its own parameters which impact the performance. This should be thought of as a group of methods, not a single method.

References

Brunetti, M., Maugeri, M., Nanni, T., Simolo, C., & Spinoni, J. (2014). High‐resolution temperature climatology for Italy: interpolation method intercomparison. *International Journal of Climatology*, *34*(4), 1278–1296. https://doi.org/10.1002/joc.3764

Cao, W., Hu, J. X., & Yu, X. (2009). A study on temperature interpolation methods based on GIS. *2009 17th International Conference on Geoinformatics*, 1–5. https://doi.org/10.1109/GEOINFORMATICS.2009.5293422

Esri. (n.d.). *Classification trees of the interpolation methods offered in Geostatistical Analyst*. ArcGIS Pro. Accessed 11/24/2024. https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

Esri. (n.d.). *How local polynomial interpolation works*. ArcGIS Pro. Accessed 11/24/2024. https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-local-polynomial-interpolation-works.htm

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research - Atmospheres*, *113*(D21), D21110-n/a. https://doi.org/10.1029/2008JD010100

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	**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	**95**