Lab Report 3.2

Title: Comparing Different Interpolation Types
Notice: Dr. Bryan Runck
Author: Evening Hade
Date: 3 December 2024

Project Repository: https://github.com/eveningsaria/gis5571/tree/main/Lab3
Google Drive Link: N/A
Time Spent: 15 hours

Abstract
Using parameters with an API query, real-time station data was downloaded from NDAWN. These were used to make four different spatial interpolations, two differential and two stochastic. The varying results of the output rasters provides food for thought.

Problem Statement
The purpose of this exercise is to compare how different interpolation methods produce different results. The focus of this exercise was twofold: first, an ETL for real-time data needed to be made, then, that data needed to be interpolated in different ways.

Table 1. Types of data used.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Station	A weather station in the NDAWN network	Point feature	Station name	NDAWN	Dataset must be cleaned & geolocated
2	Average temperature	.Average temperature over a 30-day period	Attribute information	Average temperature	NDAWN	“”

Input Data
One giant .csv file was downloaded from NDAWN containing the daily average temperature recorded at every NDAWN station over the last 30 days. Each day & location constitute a different row in the table. This data needed to be cleaned to only include the data table before it could be used.

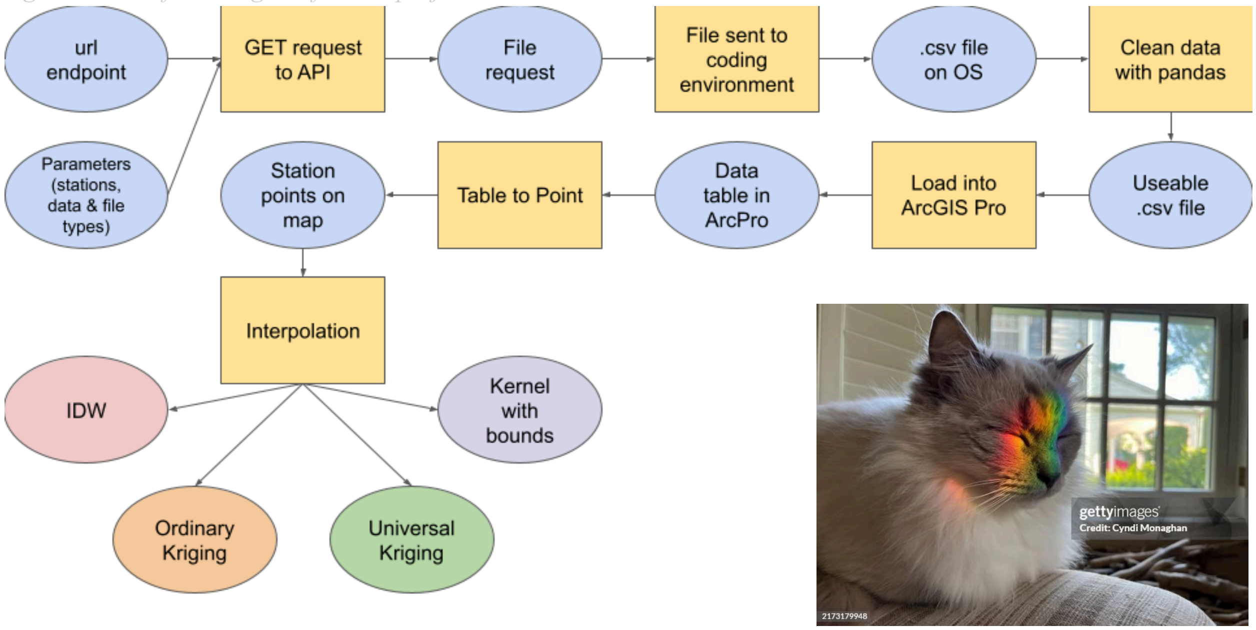
Table 2. Datasets used in this exercise.

#	Title	Purpose in Analysis	Link to Source
1	NDAWN daily data	Weather station location with mean temp. readings	NDAWN

Methods

The .csv table was manipulated first using Python and the modules `requests`, `os`, and `pandas`. To start, the data was accessed by using `requests.get()` on a url endpoint for the data and saved to a file folder in `os`. Next, the data was cleaned using `pandas` to remove everything in the file that was not the data table. Next, using `arcpy` the .csv was added to ArcGIS Pro and turned into a standalone table. This table was then converted to a point feature layer using XY Table to Point. This is the layer used as in input for all the interpolations, four of which were made using four different methods: IDW, kernel with barriers, ordinary spherical kriging, and universal linear drift kriging.

Figure 1. Data flow diagram for this project.



Results

Even while keeping the same input, the rasters created by the different interpolation methods were all different. The IDW created areas of extreme change around the station locations which were usually warmer than the surrounding interpolated area. The kernel method only interpolated within a buffer around the stations, so the empty areas in South Dakota and northern Minnesota were avoided; it was the only interpolation method to do so. While I tried to set the extent of the other rasters, they just didn't want to be constrained! The kernel interpolation resulted in a smooth transition from cold in the northwest to warm in the southeast, and unlike the IDW, did not always match the station readings taken at the location. Next were the two stochastic methods of spatial interpolation, ordinary and universal kriging. I had never done kriging before so I wanted to experience the difference between the two main types. To my surprise there was indeed a clear difference in the results. The universal linear drift kriging interpolation was the only interpolation to make the southeast corner colder. On the other hand, the ordinary spherical kriging interpolation spread the warm values further throughout the interpolated area than any of the other methods.

Results Verification
These results can be replicated using ArcGIS Pro's Jupyter Notebook integration. They will change should the parameters of the interpolations change, i.e. the polynomial of the kernel funcion or the power of the IDW function.

Discussion and Conclusion
With so much variance between the interpolations, it was hard to pick a favorite, but I would say the ordinary spherical kriging method created the most representative interpolation. While the average temperature values did not exactly match those recorded by the weather stations, the values were pretty close. Also, this method made interpolated areas very similar to the weather stations they were next to, unlike the IDW method. While the IDW method preserved the exact average temperature values for each weather station, the data looked "choppy" and the discord made the map difficult to interpret. In conclusion, the vast variance between the results of the interpolations teaches us two lessons. One is that we should endeavor to consider how data has been interpolated when we encounter interpolations in news or research. After all, interpolations are only a representation, and all representations can be skewed to benefit a narrative. The second is that there are many, many representations we can choose for our data and we shouldn't feel bound to one method of interpretation simply because it is a popular or simple form. This exercise has heightened my understanding and made me more curious all at once.

Figure 2. Interpolation of weather station data using an inverse distance weighted method. Stations = circles.

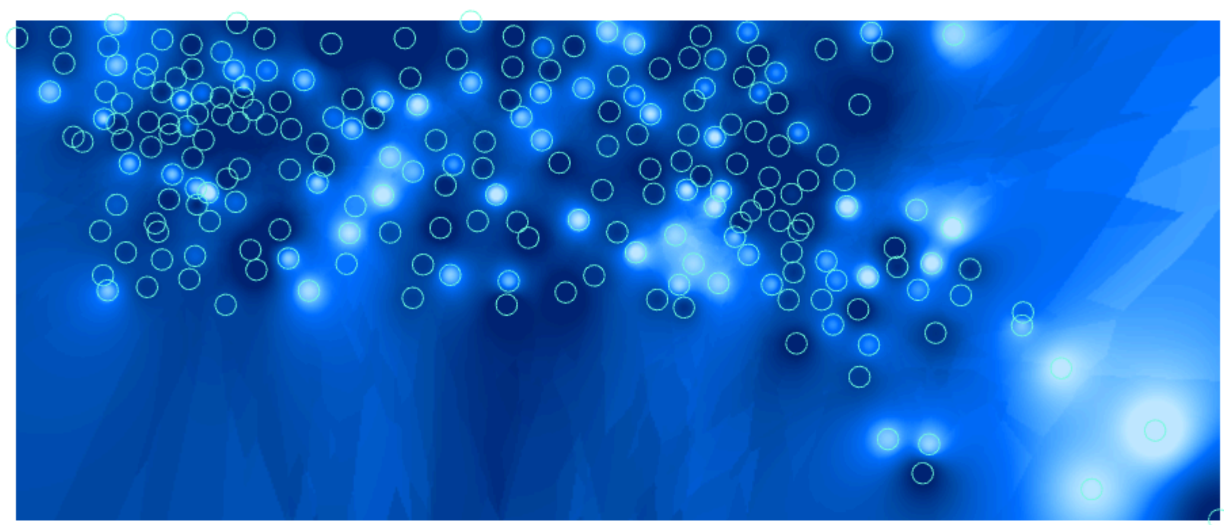


Figure 3. Interpolation of weather station data using a kernel funcion with barriers. Stations = circles.

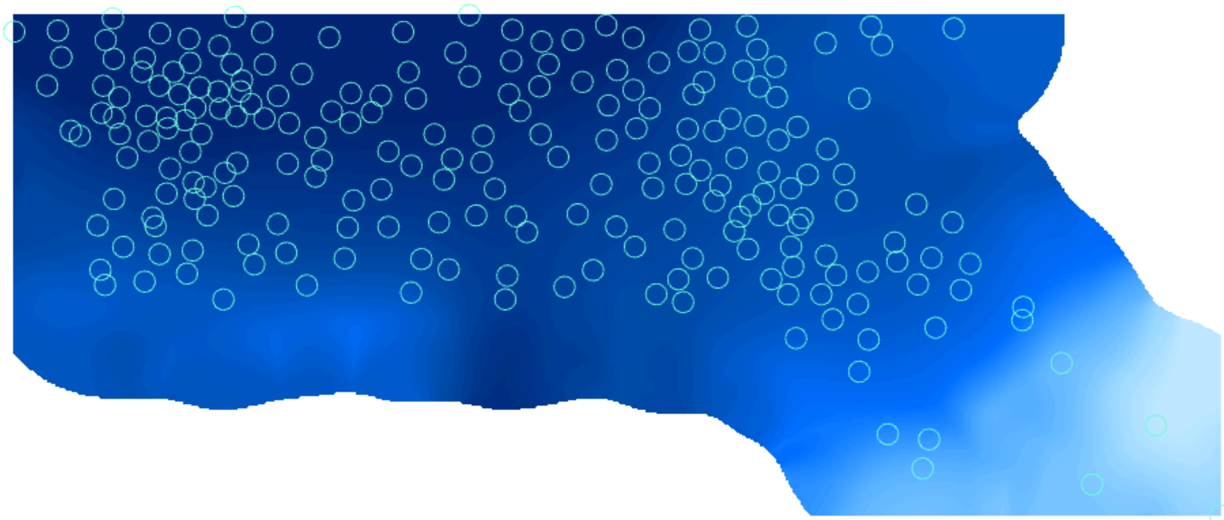


Figure 4. Interpolation of weather station data using an ordinary spherical kriging method. Stations = circles.

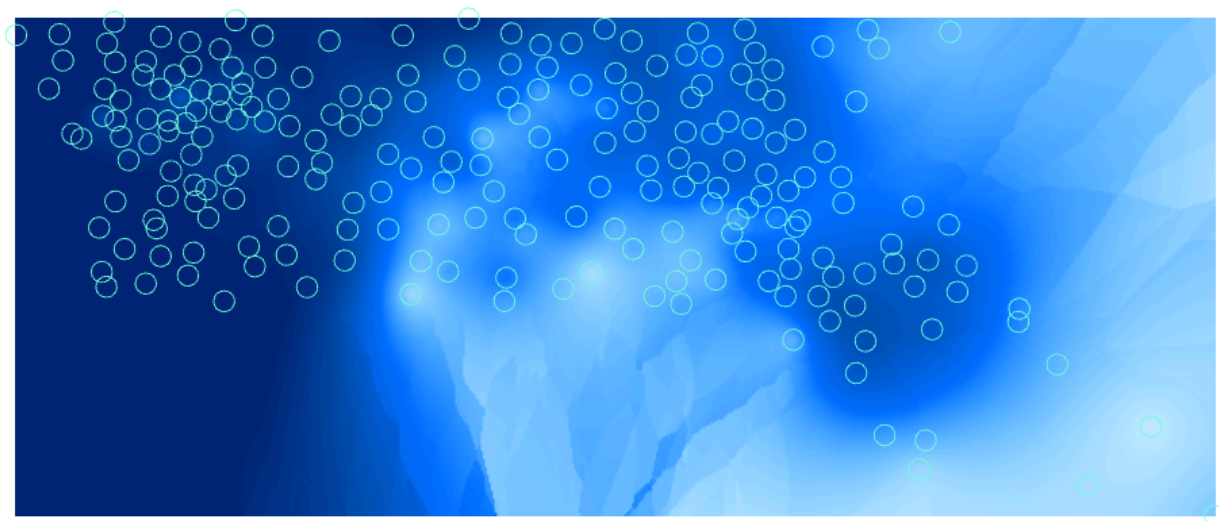
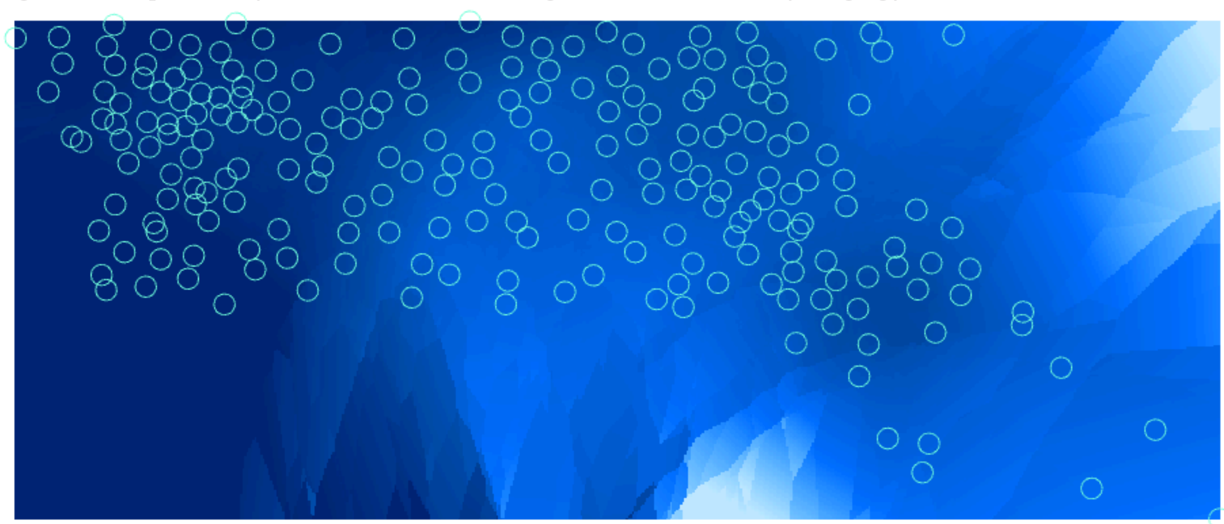


Figure 5. Interpolation of weather station data using a universal linder drift kriging function. Stations = circles.



References

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

North Dakota State University. (2024). *Daily Weather Data*. NDAWN. https://ndawn.ndsu.nodak.edu/weather-data-daily.html

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**