Lab Report

Title: Lab 4: Interpolation
Notice: Dr. Bryan Runck
Author: Cole Anderson
Date: 4/20/21

Project Repository: https://github.com/and04671/GIS5572/tree/main/Lab4

Abstract

The problem to solve in this lab was to compare several ways to interpolate and display live monthly average temperature data across the NDAWN study area. The monthly data was retrieved from the NDAWN datasite and was the only required data. The overall process was to create a request function that would retrieve data in the correct time frame from the correct station sites, using a URL. This code was already almost completed as a lab booster. The data was spit out into a saved table where it was interpolated via IDW, Kriging, and Natural Neighbor methods. The output maps were compared to each other and another reference map. The method outputs were found to be similar and relatively accurate. The conclusion discusses the most preferred method for interpolating weather data, which isn't entirely agreed upon.

Problem Statement

The problem to solve is as follows: find several ways to interpolate and display live monthly average temperature data across the NDAWN study area.

Table 1: Requirements

#	Requirement	Defined As	Spatial Data	Attribute Data	Dataset	Preparation
1	Retrieve last 30 days of data from NDAWN	Average monthly temperature at each stations	Station Locations	Avg. Temp.	Mn GeoSpatial Commons	Integrate pre-made ETL script
3	IDW Interpolation	**arcpy.sa.Idw**(inPointFeatures, zField, cellSize, power)	Station Location	Station average temp.	AADT Data	Aggregated average temperature table
4	Kriging Interpolation	**arcpy.sa.Kriging**(inFeatures, field, kModelOrdinary, cellSize)	Station Location	Station average temp.		Aggregated average temperature table
5	Nearest Neighbor Interpolation	**arcpy.NaturalNeighbor_3d**(inPntFeat, zField, outRaster, cellSize)	Station Location	Station average temp.		Aggregated average temperature table

Input Data

Only aggregated NDAWN data is required for this lab.

Table 2. Input Data

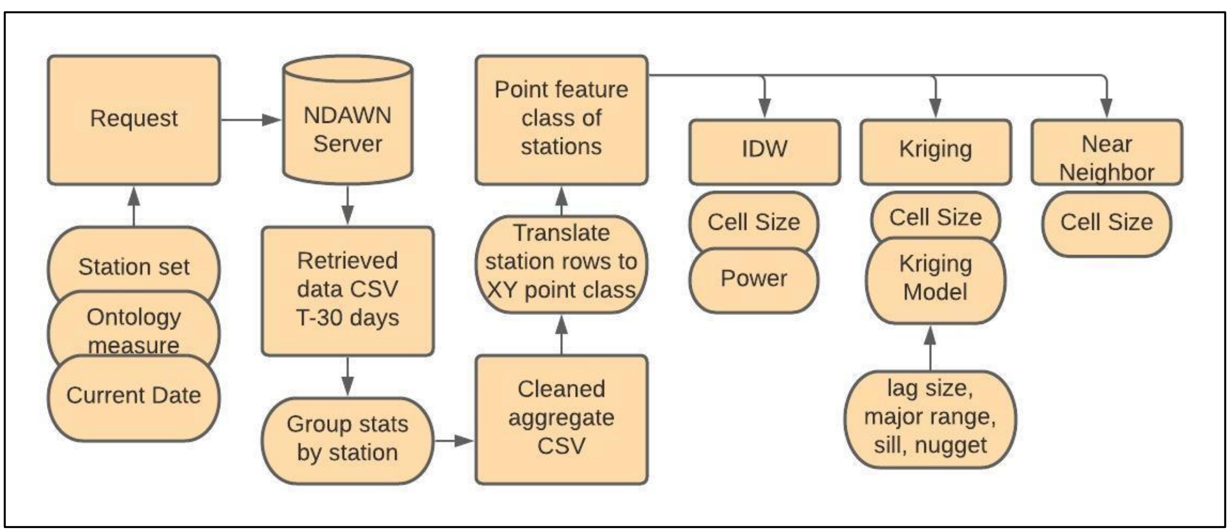
#	Title	Purpose in Analysis	Link to Source
1	NDAWNdata	Source data for aggregation and interpolation comparison	NDAWN Monthly Data

Methods

The initial set up for this Lab was to find the correct tag markers for each of the NDAWN station locatons. This information can be found under the Inspect window. These values were assigned to the correct location (key) in a dictionary. The next step was to create the **ndawn_request** class. The initialization function **init** accepts a start date, end date, desired weather variable(s), and a list of temperatures, as well as a save option. Essentially, this function organizes the request syntax. The next **get_data** function inside the class receives the syntax text from the initial function and strings them together. It sends it out the entire correct URL string to the NDAWN server as a request. NDAWN returns a dataset that is translated to a dataframe and saved to file. In this case, the dataframe required aggregation before saving, with all the rows for each station being averaged. There are then as many rows as stations, saved as **GroupedNDAWN.csv.** In order to accomplish retrieving 30 days of data before the current date, the 'start date' was set as datetime function **date.today().** The end date was set as whatever date was 30 days before the start using **timedelta(30).** Once the desired data was saved into a csv, it had to be translated to actual station points. Fortunately, the data table had latitude and longitude columns, so an **arcpy.management.XYTableToPoint** was the only needed step. The feature class was then interpolated as described immediately below.

For IDW Interpolation, the arcpy.sa.Idw function was used. This function requires a cell size, a power to raise the inverse distance to, the point class, and the magnitude field in said class. A cell size of .01 seemed to work nicely, the default 200 was far to large. The power was set to 3. **For Kriging Interpolation**, the arcpy.sa.Kriging function was used. This function requires a cell size, point class, and magnitude variable, but also a 'Kriging Model.' This model consists of search method (circular) in addition to lag, range, sill, and nugget parameters. The cell size was kept at .01, while the model variables were randomly set as from 0-2. **For Natural Neighbor Interpolation**, the arcpy.sa.NaturalNeighbor_3D function was used. This function requires a cell size, the point class, and the magnitude field in said class. A cell size of .01 seemed to work nicely. All of the method outputs were saved as TIFF files to be placed in an ArcPro data layout. The outputs were saved as a TIF, but if the script is run in ArcGIS a map layer will also appear. These can be seen immediately below in Results.

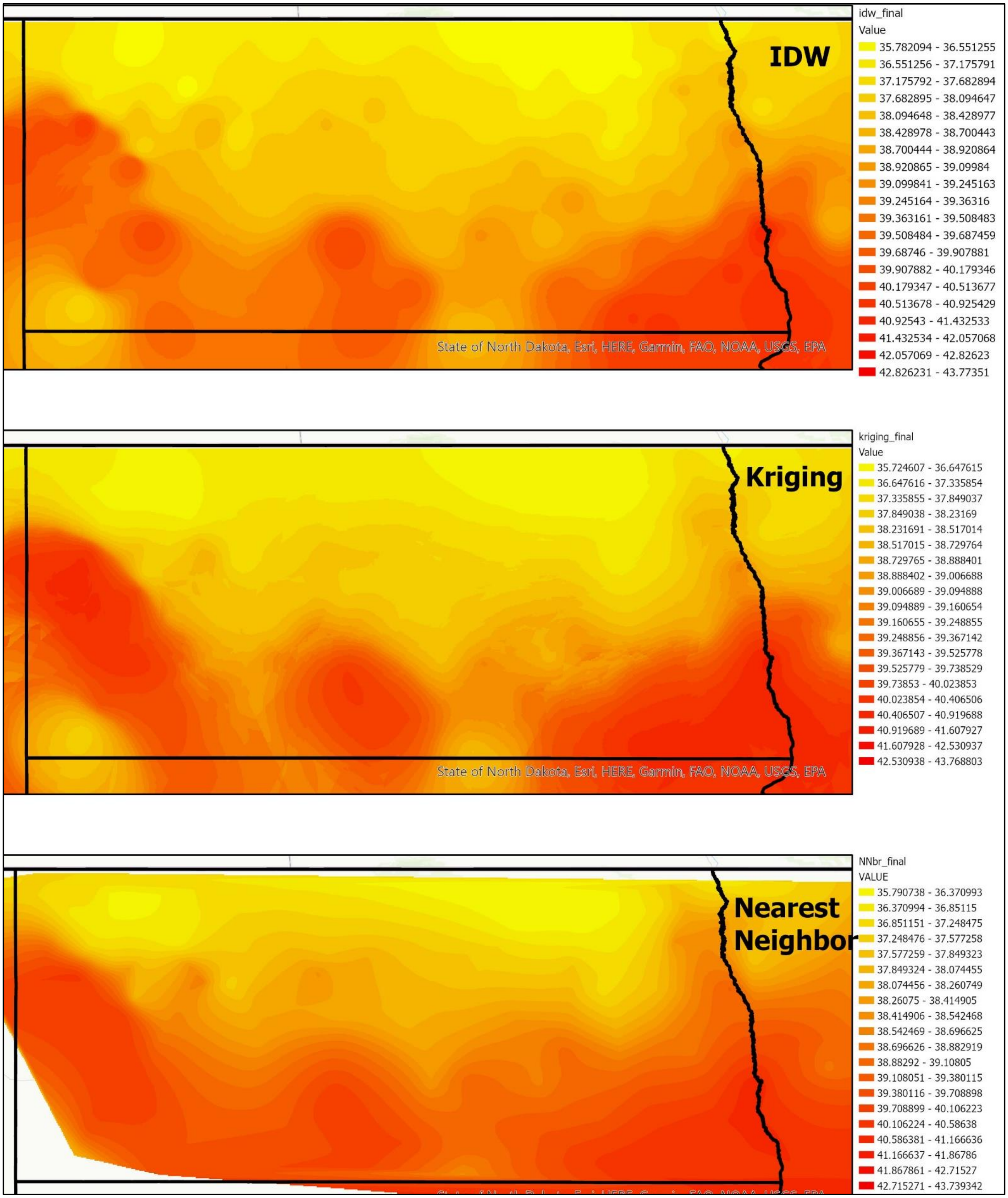
Figure 1: Spatial Processing Workflow



Results

These are the output rasters from each of the interpolation methods. The legends on the right are all very similar, but the values in each class do vary slightly.

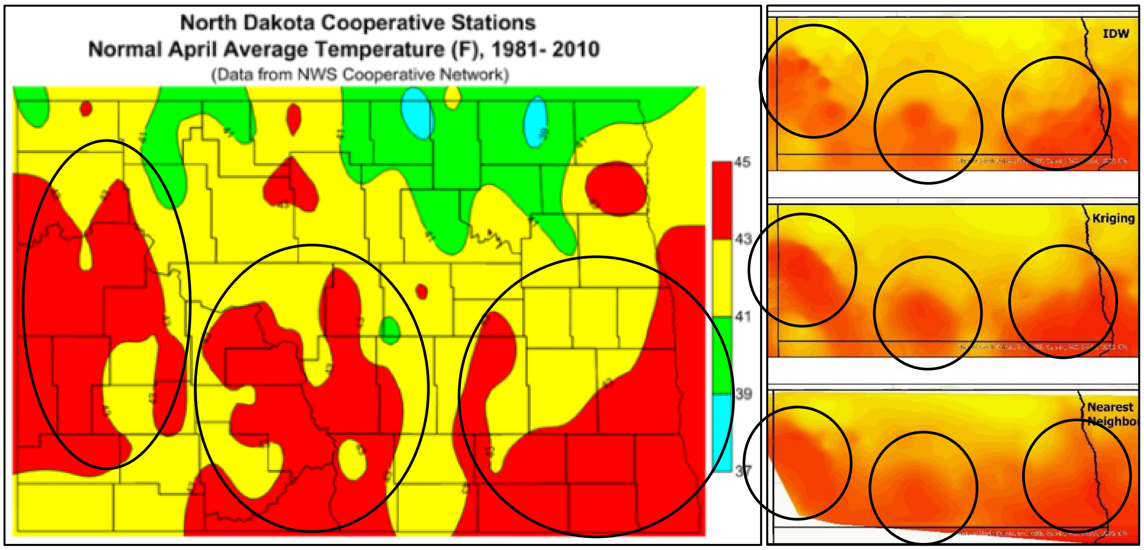
Figure 2: Interpolation results



Results Verification

To verify the results, I compared a April temperature average map from NDSU with the interpolation maps. This map uses a different data source, but should be similar. Transposing the two maps after georeferenced, the temperature areas can be compared. The general pattern is correct.

Figure 3: Comparison of Interpolation with NDSU map.



Actually, none of the interpolation maps are terrible. All of the methods contain similar high temperature 'blobs' in the southeast corner, in the bottom center, and off to the center left; these are indicated with the circles on the maps above. Natural Neighbor does have a warmer bridge between the hotspots, but these hotspots are still well defined. All methods get cooler to the north as well, with a sort of 'Λ' shaped light-yellow area. This lines up decently with the green spots on the NDSU map.

The more accurate methods seem to be IDW and Kriging because of the bridges between hotspots in Natural Neighbor. Then again, these bridges may be lost in the classification scheme in the more basic NDSU map. The nearest neighbor method also doesn't cover the desired study area, so is eliminated.

Discussion and Conclusion

One goal for this lab was to choose an interpolation method based on ESRI documentation. The main comparisons page offers the following related to IDW and Kriging:

- A pre-evaluation of spatial behavior is required for accurate kriging
- Kriging is processor intensive and slow compared to other methods
- IDW is a deterministic method (which assigns values based on a pre-specified formula and surrounding data points), while Kriging is geostatistical (based on specific statistical models)
- IDW requires an even and dense distribution of data point for an accurate result
- IDW cannot preserve ridges or pits

The remaining options were Trend, Spline, and Natural Neighbor. According to the documentation, all of these are deterministic. Natural Neighbor uses the "closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value." It is also called Sibson interpolation. Trend best captures gradual change and coarse patterns. Spline tries to minimize curvature when estimating values.

Another goal for this lab was to explore what method(s) of interpolation are best for weather data like temperature. For this lab, the 2018 dissertation *"Interpolation of temperature data for improved weather forecasts"* by Frida Cronqvist at the KTH Royal Institute of Technology School of Engineering Sciences was the primary guide. The paper primarily compares IDW interpolation to regression kriging under the assumption that known temperature points were held fixed. The conclusion is that kriging generally offers less mean square error than IDW, under the circumstances that the additional required assumptions are true. In order to accurately make such assumptions, more data than location, temperature, and elevation are required. Kriging is also computationally more intensive than IDW, so there are tradeoffs between both.

References

Jarvis, C. H., & Stuart, N. (2000). A Comparison among Strategies for Interpolating Maximum and Minimum Daily Air Temperatures. Part I: The Selection of "Guiding" Topographic and Land Cover Variables. *Journal of Applied Meteorology*, *40*(6), 1060–1074. https://doi.org/10.1175/1520-0450(2001)040<1060:acasfi>2.0.co;2

Bishop, J. (2021, March 11). *NDAWN Weather Scraper*. GitHub. https://github.umn.edu/BISH0227/NDAWN_Weather_Scraper.

ESRI. (2020). *Comparing interpolation Methods*. Comparing interpolation methods-ArcGIS Pro | Documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/comparing-interpolation-methods.htm.

Cronqvist, F. (2018). *Interpolation of temperature data for improved weather forecasts* (dissertation). KTH Royal Institute of Technology School of Engineering Sciences.

https://www.ndsu.edu/ndsco/data/30yearaverage/averagetemperature/#c343106

Self-score

Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.

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	**27**
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**	24	**23**

	points).		
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	**20**
		100	**97**