

## **Lab 3 Report: NDAWN Temperature Interpolation**

Title: NDAWN Temperature Interpolation

Notice: Dr. Bryan Runck

Author: Diego Osorio

Date: 11/30/2022

**Project Repository:** <https://github.com/osori050/GIS5571/tree/main/Lab3>

**Google Drive Link:**

**Time Spent:** 10 hours

### **Abstract**

Real-time data visualization models were generated by interpolating daily maximum and minimum air temperatures from all NDAWN stations for the last 30 days from now. Inverse distance weighting (IDW), global polynomial interpolation (GPI), Kriging ordinary (KO), and Kriging universal (KU) were utilized and their parameters were chosen after a sensitivity analysis was performed. KO Spherical, Circular, Exponential, and Linear yielded the best models which correspond to what has been reported by the literature.

### **Problem Statement**

This exercise aims to build a fully functional real-time data visualization of the maximum and minimum temperature recorded in the last 30 days by all the North Dakota Agricultural Weather Network (NDAWN) stations. That is, regardless of when the script is run, it will always retrieve the data from the last 30 days, which can be defined as a moving time window. Additionally, that data will be interpolated using inverse distance weighting (IDW), global polynomial interpolation (GPI), Kriging ordinary (KO), and Kriging universal (KU) to see the temperature variation in the region. Figure 1 shows the map of the locations of NDAWN stations in Montana, North Dakota, and Minnesota.

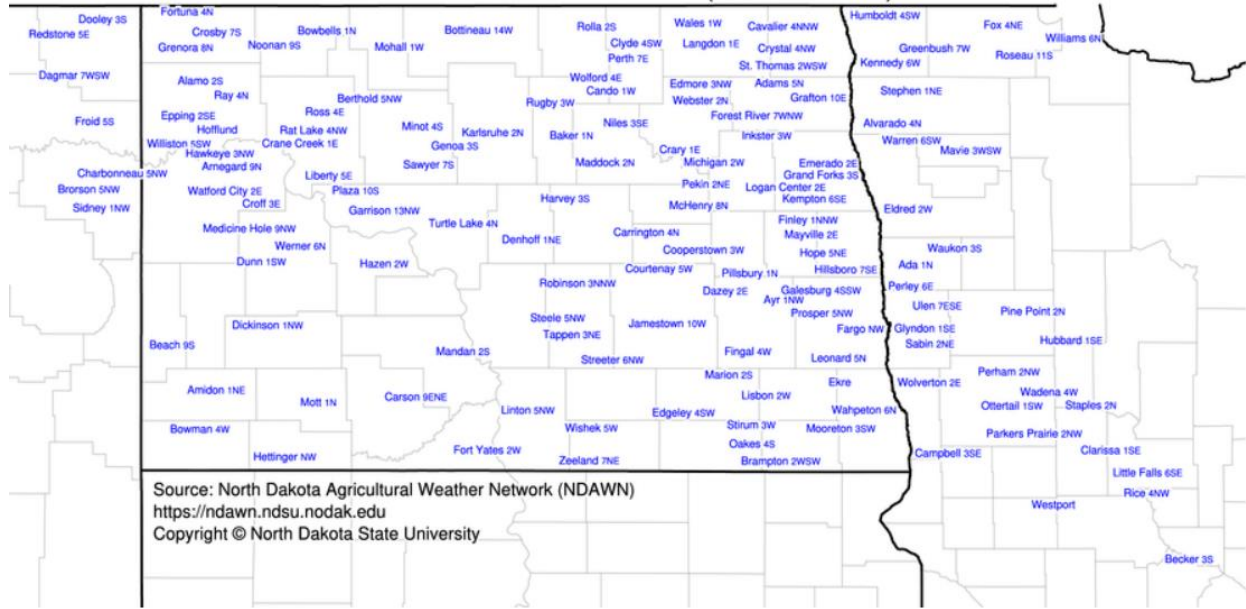


Figure 1. Location of NDAWN stations in Montana, North Dakota, and Minnesota (North Dakota State University, n.d.)

Table 1. Data required for the analysis

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Meteorological data	Daily measurement of the max and min temperature of the last 30 days from now	Coordinates	Temperature (F)	NDAWN Center	Convert csv files to vector layers and then, define spatial reference to WGS 1984

## Input Data

The dataset downloaded from NDAWN corresponds to a csv file listing all the meteorological stations with their coordinates. The variables of interest are the daily maximum and minimum air temperatures in the last month. Therefore, the table contains 30 readings for each station.

Table 2. Input data

#	Title	Purpose in Analysis	Link to Source
1	NDAWN daily max and min air temperature for the last 30 days	Raw input dataset in csv file to analyze the daily maximum and minimum air temperature recorded in all the stations in the last 30 days	<a href="https://ndawn.ndsu.nodak.edu">NDAWN</a>

## Methods

To begin with, the current date was calculated using the package `datetime` and a string variable was created to save it following the convention `mm-dd-yyyy`. Then the date a month ago from today was also generated and saved into another variable. These variables were included as date parameters when retrieving the data from NDAWN's API. The datasets downloaded as csv files were converted into DataFrames where some columns and rows were dropped. Later, the data was grouped by Station Name and the 30 readings for each station were averaged. Also, geometry columns were created and populated using the coordinates to create GeoDataFrames and then, shapefiles. Finally, the coordinate system was defined by utilizing the Define Projection tool of `arcpy`.

Regarding the interpolators, they were selected to have a representation of each kind, when possible, in the different categories. That is, as shown in Table 3, each interpolator falls into one or more kinds for each category. For the time constraints of this lab, all the interpolators are fast in the processing speed category. Additionally, for the Kriging methods, ordinary and universal were selected to compare the variability within this group as well.

Table 3. Classification of the selected interpolators (ESRI, 2021)

Category	Kind	IDW	GPI	Kriging
Type of information	One prediction per location	X	X	X
	Quantile value			X
	Many predictions per location*			
	Predicted values and errors			X
Measurement of spatial autocorrelation	Yes			X
	No		X	
	Implicit	X		
Output type	Prediction	X	X	
	Prediction error			X
	Probability			X
	Full distribution of possible values*			
Level of assumptions	Few	X	X	
	Intermediate			
	Many			X
Type of interpolation	Exact	X		

	Inexact		X	X
Smoothness of the output	Smooth		X	
	Intermediate			X
	Not smooth	X		
Uncertainty of predicted values	Yes			X
	No	X	X	
Processing speed	Slow			
	Intermediate			
	Fast	X	X	X

\* None due to complexity of the simulations

Figure 2 illustrates the conceptual model of this methodology.

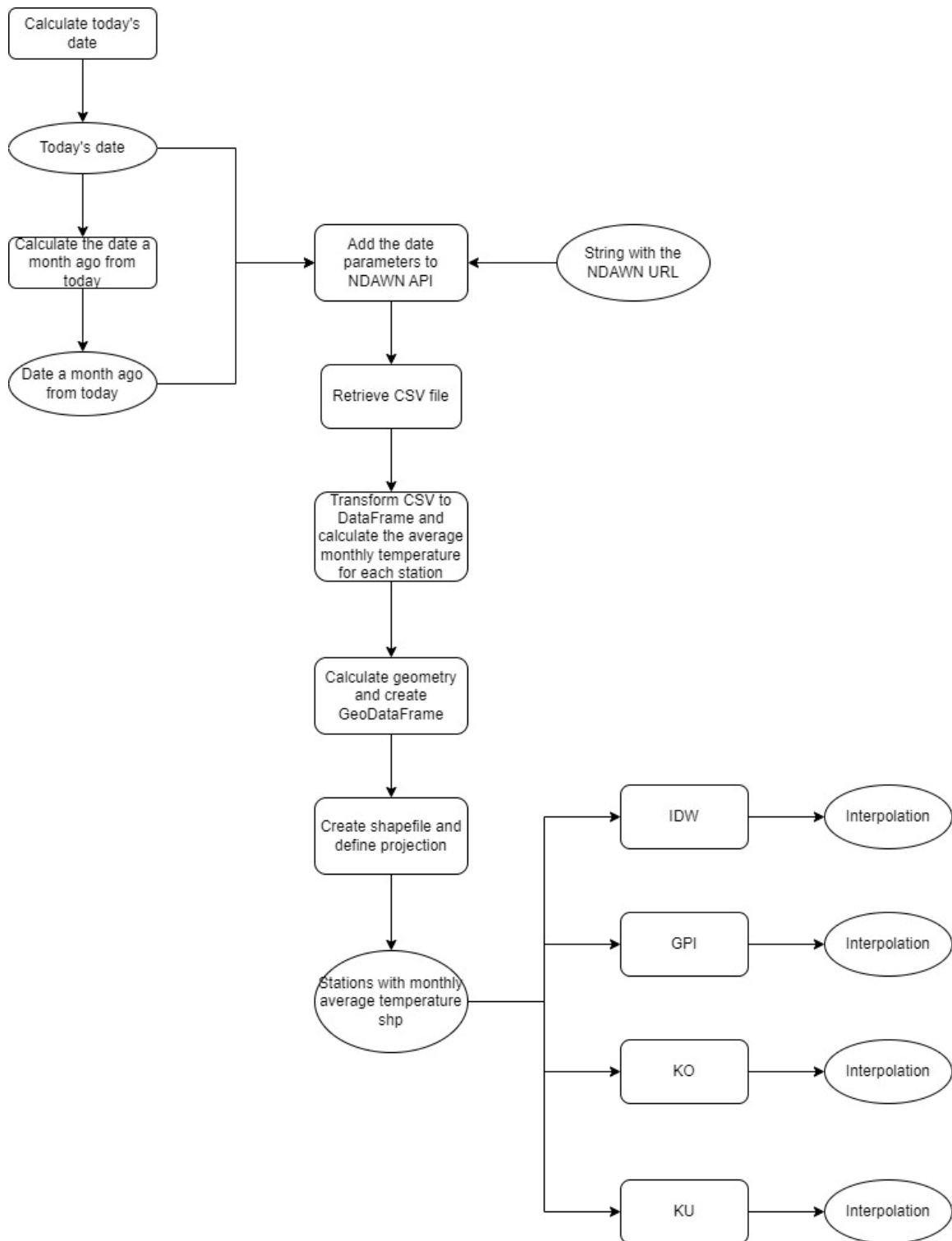


Figure 2. Flow diagram of data interpolation

## Results

It was found that the higher the alpha, the rougher the IDW interpolation is as shown in Figure 3. Since temperature changes tend to be rather smooth and subtle, alpha values of 2 or 1 seem to be optimal when modeling temperature. Likewise, fixed radius interpolations were discarded since they produce rough results.

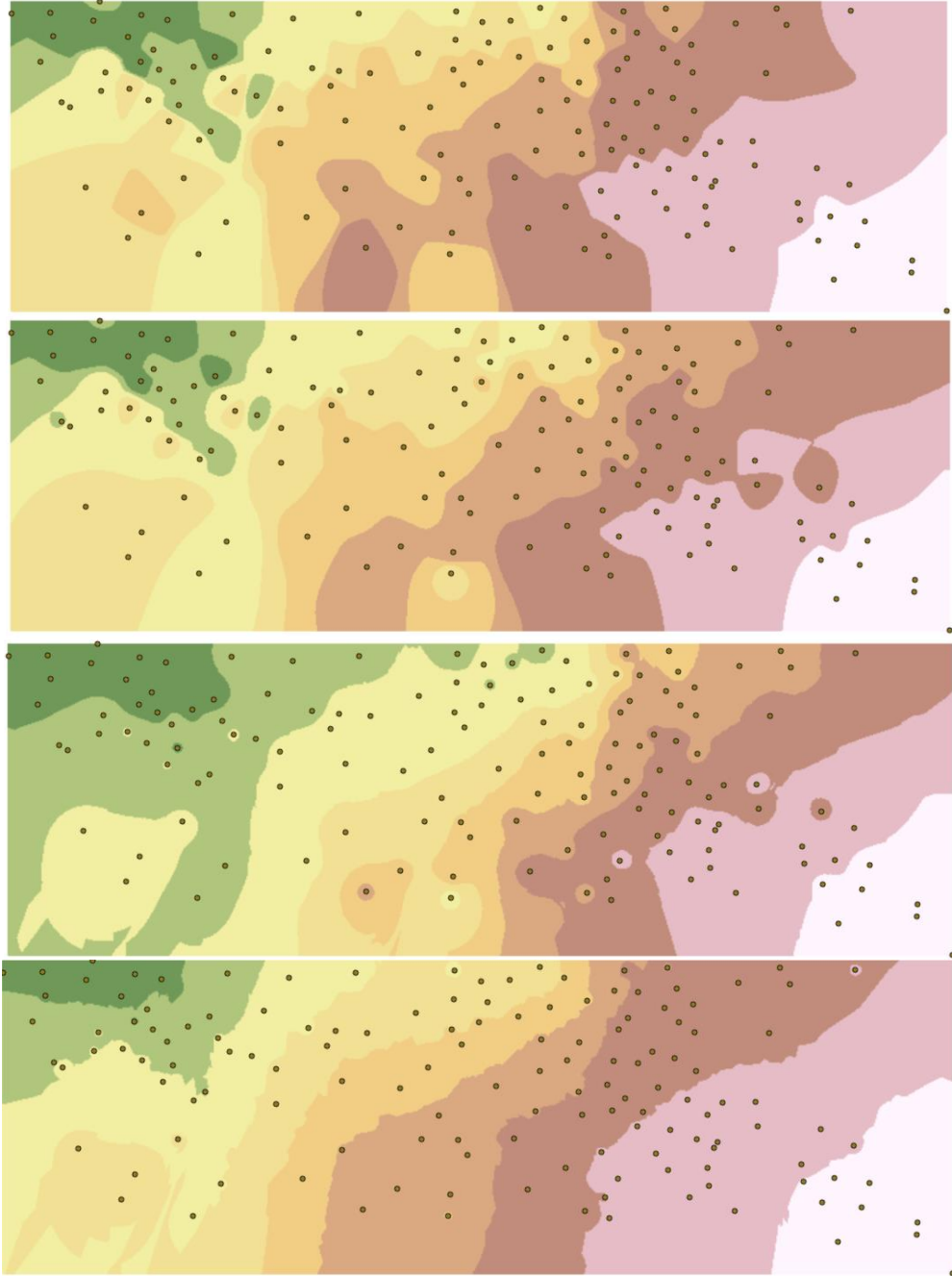


Figure 3. Different parameters for IDW. From top to bottom,  $\alpha = 10, 5, 2$ , and  $1$  respectively.

The polynomial order was changed for GPI (Figure 4) and the results show 2 is the best fit since 1 does not yield a realistic model and 3 distorts too much the values. Moreover, “it should be noted that the more complex the polynomial, the more difficult it is to ascribe physical meaning to it” (ESRI, n.d.).

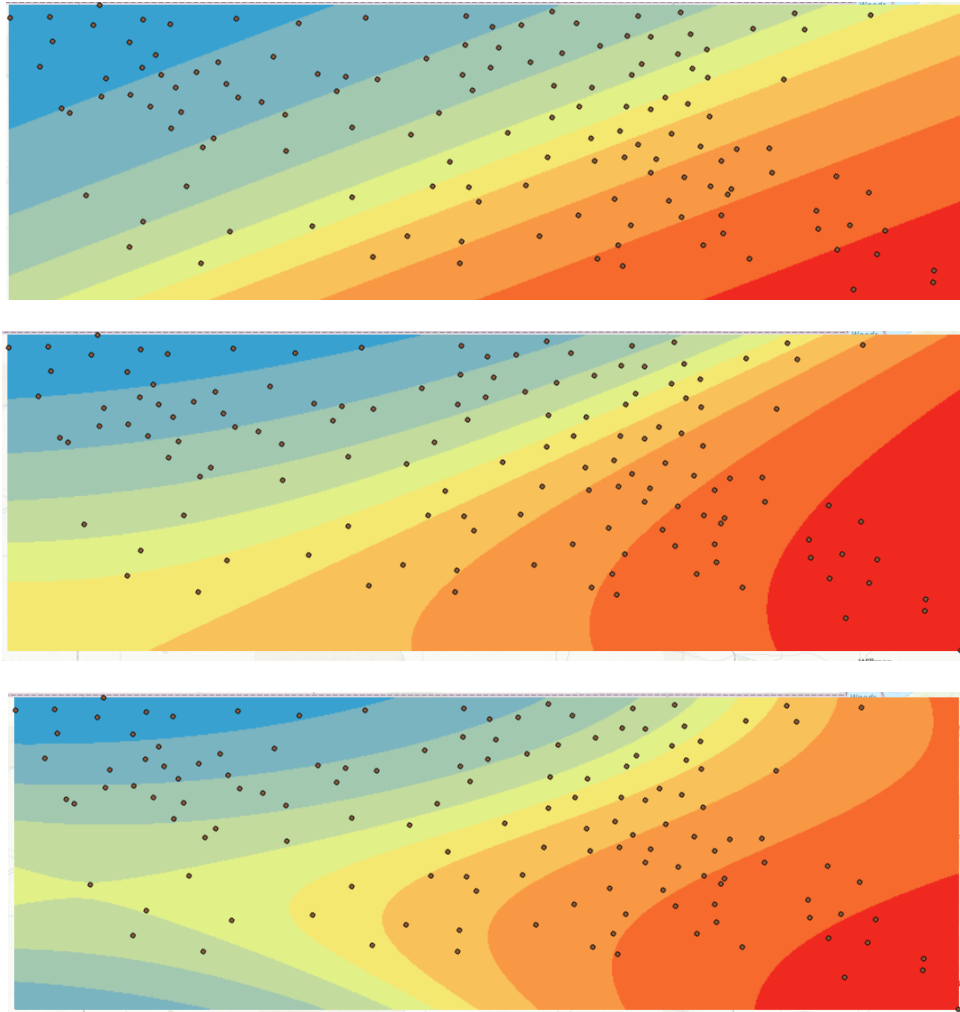


Figure 4. Different polynomial orders for GPI. From top to bottom, order = 1, 2, and 3 respectively

Regarding Kriging, the semi-variogram models for KO yielded the same result except for Gaussian or normal distribution which, albeit smoother than IDW, it is still rough to model temperature. For KU, both semi-variogram models, linear and quadratic, generate results similar to KO Gaussian which, consequently, have the same problem when applied to model this temperature data. Figure 5 and Figure 6 illustrate the outputs of the Kriging group.

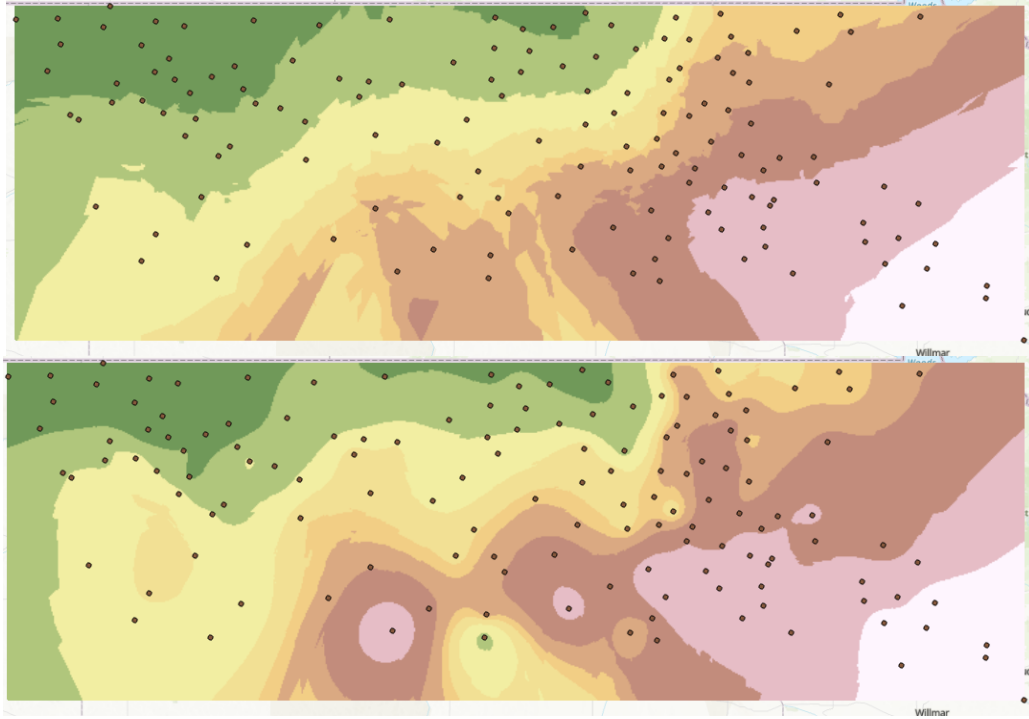


Figure 5. Kriging ordinary. Top: semi-variogram model = Gaussian; Bottom: semi-variogram model = Spherical, Circular, Exponential, and Linear

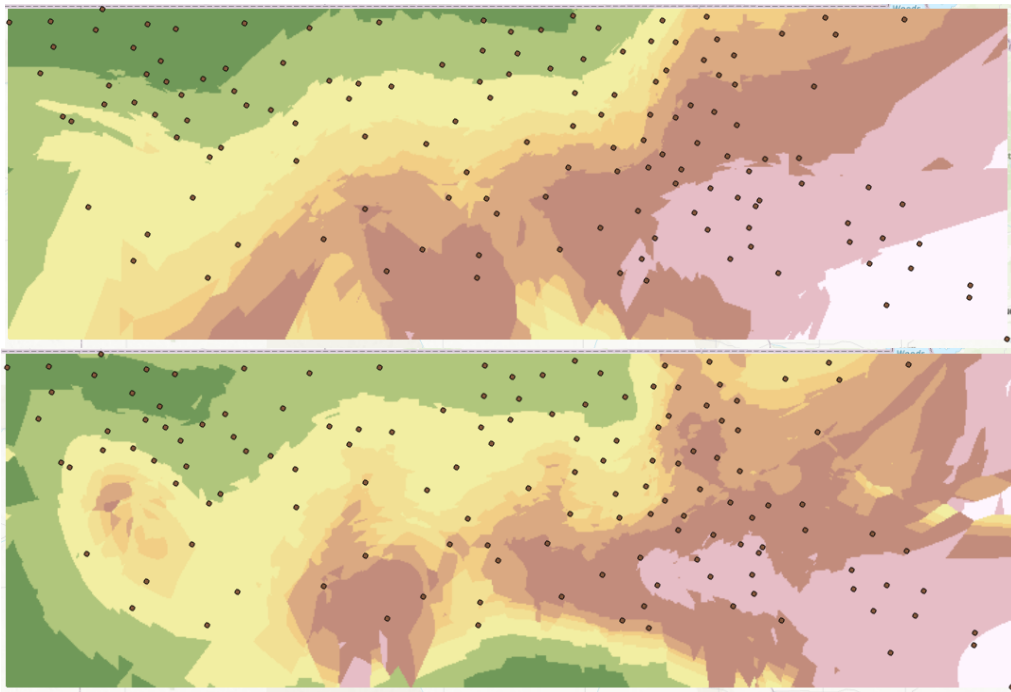


Figure 6. Kriging universal. Top: semi-variogram model = linear drift; Bottom: semi-variogram model = Quadratic drift

Overall, all the interpolators and their variants show low temperatures in the northwest and warmer ones in the southeast for both maximum and minimum temperature readings. Additionally, the



best fits for the data were the non-Gaussian KO's, i.e., Spherical, Circular, Exponential, and Linear.

## Results Verification

The results are consistent with what Cao et al. (2009) found when interpolating temperature values from 327 weather stations in China. Their findings showed that KO Spherical and Exponential have the highest accuracy, IDW is less accurate, and Gaussian KO is the least accurate.

## Discussion and Conclusion

The Kriging ordinary method with the semi-variograms Spherical, Circular, Exponential, and Linear yield the best models to interpolate temperature data. The rest of the interpolators and variants used here failed due to rough, unrealistic, or distorted outputs.

All in all, this exercise taught me to select an interpolator that fits best my needs by following decision trees. Likewise, I carried out a sensitivity analysis for each interpolator and found the best parameters for each to model temperature. This same methodology can be applied to my final project when selecting and comparing interpolators for water quality variables. Additionally, I learned to create code to request data in real time.

## References

- Cao, W., Hu, J., & Yu, X. (2009). A study on temperature interpolation methods based on GIS. *2009 17th International Conference on Geoinformatics* (pp. 1-5). IEEE.
- ESRI. (2021). *Classification trees of the interpolation methods offered in Geostatistical Analyst*. Retrieved from ArcGIS Desktop: <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm>
- ESRI. (n.d.). *How global polynomial interpolation works*. Retrieved from ArcGIS Pro: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-global-polynomial-interpolation-works.htm>
- North Dakota State University. (n.d.). *NDAWN Station locations*. Retrieved from NDAWN Center: <https://ndawn.ndsu.nodak.edu/>

## Self-score

Category	Description	Points Possible	Score
<b>Structural Elements</b>	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

<b>Clarity of Content</b>	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 ( <b>12 points</b> ). There is a clear connection from data to results to discussion and conclusion ( <b>12 points</b> ).	24	24
<b>Reproducibility</b>	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
<b>Verification</b>	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated ( <b>10 points</b> ), the method of comparison is clearly stated ( <b>5 points</b> ), and the result of verification is clearly stated ( <b>5 points</b> ).	20	20
		100	100