**Lab Report**

Title: Lab 4

Notice: Dr. Bryan Runck

Author: Holly Leaf

Date: 12/1/2021

**Project Repository:**[*https://github.com/hleaf/GIS5571/tree/main/Lab4*](https://github.com/hleaf/GIS5571/tree/main/Lab4)

**Time Spent:** *3.5 hours*

**Abstract**

The purpose of this lab is to compare and contrast three interpolation methods: inverse distance weighted (IDW), Kriging, and local polynomial interpolation (LPI). These three interpolation methods were performed using 30-day air temperature data from NDAWN using the data from their Minnesota stations.

An ETL was created in Python to download the previous 30-days of air temperature data, calculate averages of all temperatures readings for each station, creature a feature class (points) for these stations, and perform the three methods of interpolation using this feature class. This ETL allows us to create “real-time” interpolation of the last 30-days of temperatures.

**Problem Statement**

How do different interpolation methods compare when estimating air temperature values across a region using data from several weather stations over a 30-day period?

*Table 1. The data required for this analysis*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **(Spatial) Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Air Temperature station data | Air temperature station data from NDAWN for last 30 days | Station coordinates | Air temperature in degrees F | [NDAWN](https://ndawn.ndsu.nodak.edu/) | Calculate air temperature means |

**Input Data**

The data used is from NDAWN (North Dakota Agricultural Weather Network). The weather stations used are all in northwestern Minnesota. There are 28 weather stations in the dataset. Data from the past 30 days was used (11/1/21 to 12/1/21). Each weather station records the air temperature on the hour. In all, there were approximately 20,000 weather readings that were downloaded from NDAWN over the 30 day period.

*Table 2. <insert caption>*

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Air Temperature station data | Interpolating air temperatures across the spatial extent (NW Minnesota) | [NDAWN](https://ndawn.ndsu.nodak.edu/) |

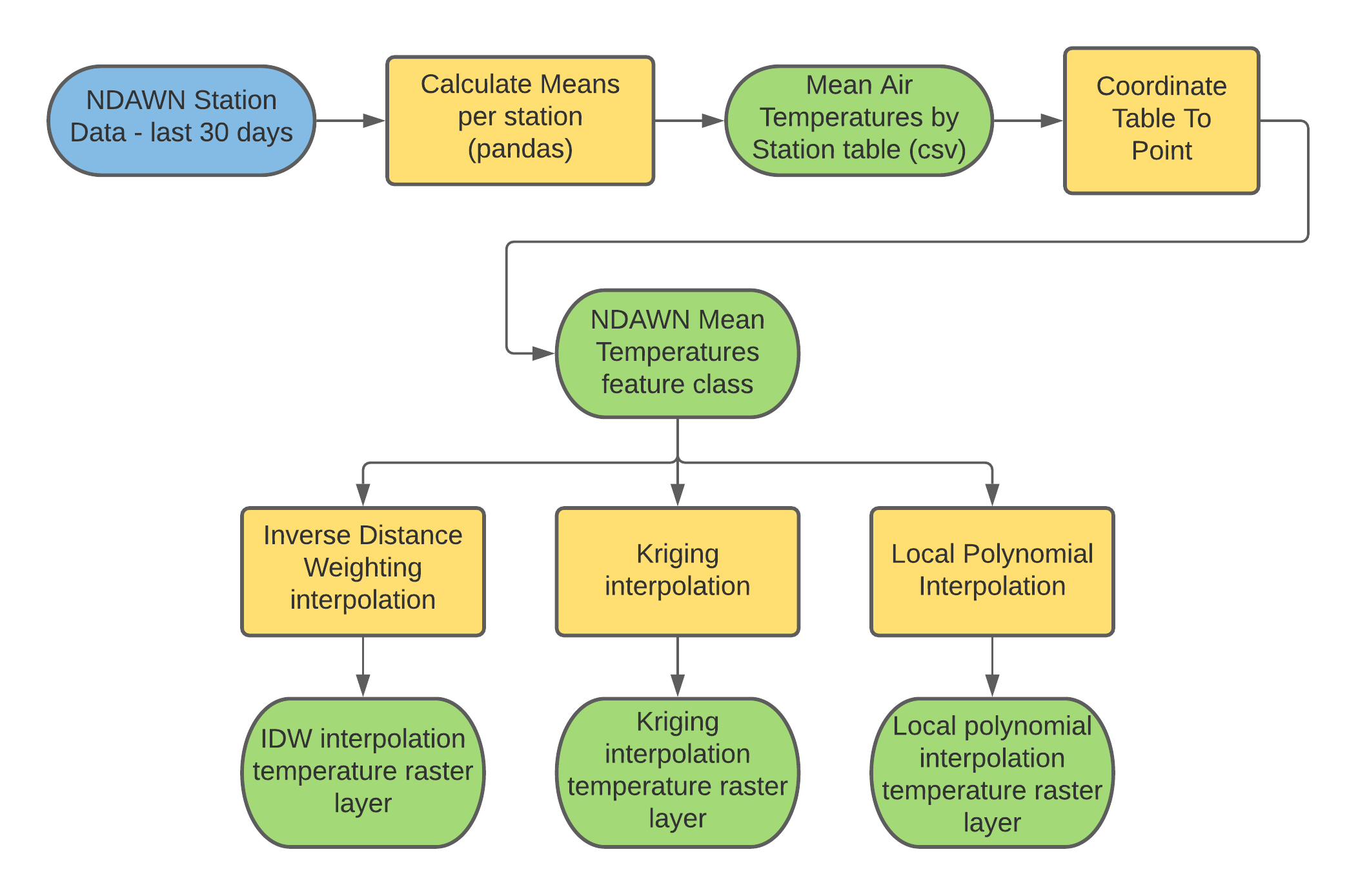
**Methods**

An ETL pipeline was created to download the weather station data from NDAWN’s Minnesota weather stations for the last 30 days from when the pipeline runs, providing “real-time” weather station data. The air temperatures for each weather station were averaged using pandas data frames, to provide a single air temperature average at each weather station location.

These averages along with the latitude and longitude of each weather station were used to create a feature class (points) in ArcGIS Pro. Using this feature class, three methods of interpolation were run in ArcGIS: inverse distance weighting (IDW), Kriging, and local polynomial interpolation (LPI). Each of these interpolation methods generated a raster layer within the spatial extent of a rectangular box that contains all weather station points.

These raster layers were used to generate three maps to compare the results of each interpolation method.

*Figure 1. Data flow diagram*

**

**Results**

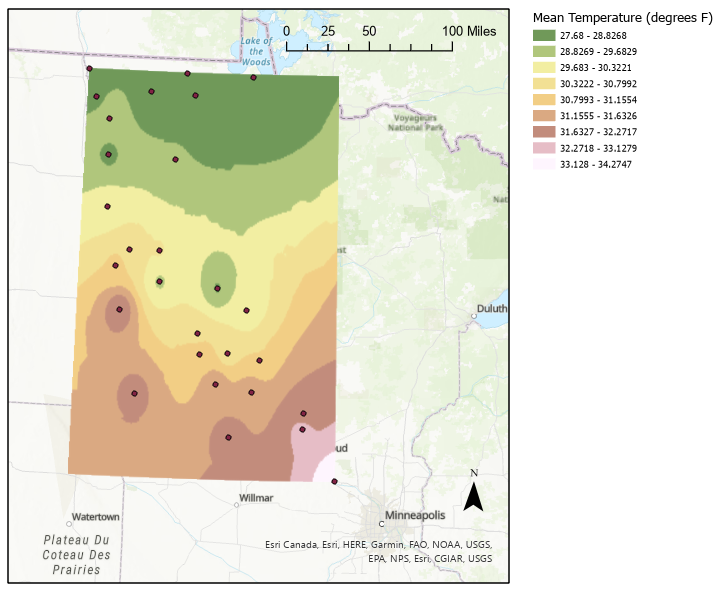
The results for each interpolation method can be seen in the maps below (Figures 2, 3, and 4). Each of these interpolation methods only cover the spatial extent of a rectangle bound by the northwestern- and southeastern-most weather stations.

The IDW interpolation method seems to provide a map which is the most similar to weather models that I have seen as a layperson, with relatively smooth contours and each weather station being within the temperature range in the interpolation raster.

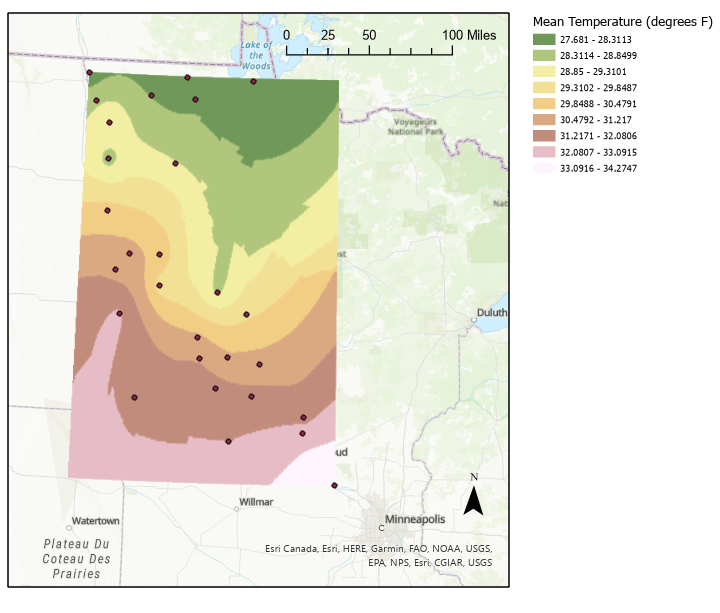
The Kriging interpolation method appears to have fewer “islands” around weather stations. Instead, there are two points that seem to “pull” the results strongly in one spatial direction. There are also some areas with sharp lines compared to IDW.

The LPI method seems to have several artifacts with sharp lines and strange shapes, particularly in the northeastern region where there are few or no weather stations. This method does not appear to function well where there are gaps in data. This method also seems to have less local effect from a single weather station compared to the IDW and Kriging methods.

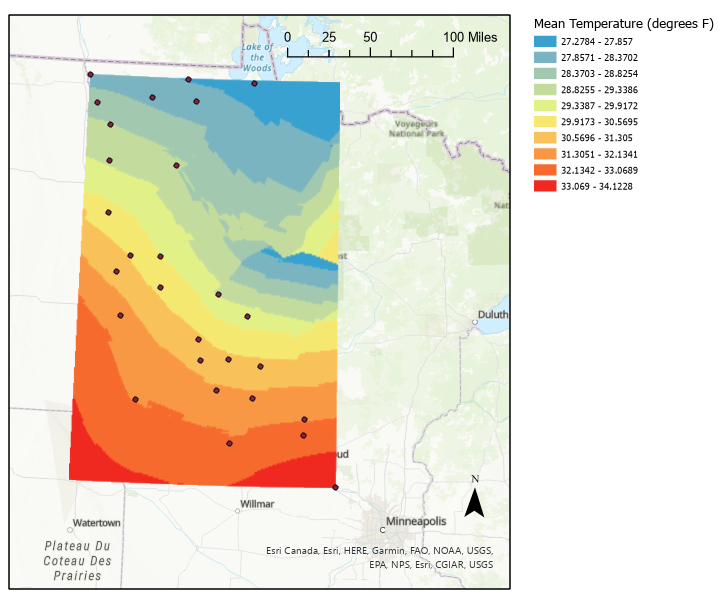
*Figure 2. IDW interpolation map*

**

*Figure 2. Kriging interpolation map*

**

*Figure 4. LPI interpolation map*

**

**Results Verification**

To check that the weather stations were placed correctly, I examined their locations in ArcGIS compared to the map on the NDAWN website. The mean temperature calculations were also compared to the (daily) air temperature averages on the NDAWN website to make sure that they were similar.

Each interpolation method differed in exact results (as expected), but largely agreed on temperature ranges across the region overall. For example, higher temperatures in the south and cooler temperatures in the north as well as similar temperature ranges across the region.

**Discussion and Conclusion**

Of the interpolation methods, IDW seems to have the most local sway by each weather station. On one hand, this creates “islands” around some weather stations (which may in fact be accurate). IDW also creates smoother lines which looks most similar to weather models that I have seen in the past. Also with IDW, stations appear to be towards the middle of temperature bands, versus the other methods where stations often appear right on the border between two bands. It seems to me that Kriging and LPI take the entire range of datapoints into account when creating the field, whereas IDW puts more weight behind distance from the individual datapoints.

In addition to the interpolation methods, I learned more about using pandas data frames to manipulate my data. I have had trouble in the past in using Python to download, save, and manipulate data. But on this lab I was able to create more of a start-to-finish ETL in Jupyter Notebooks and did not rely on the GUI very much.

**References**

ArcGIS Documentation. “How local polynomial interpolation works” <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-local-polynomial-interpolation-works.htm#GUID-0547DD7E-B720-4FBB-8B0E-1944C94B49D0>

ArcGIS Documentation. “Classification Trees of the Interpolation Methods Offered in Geostatistical Analyst” <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm>

ArcGIS Documentation. “How Kriging Works” <https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm>

ArcGIS Documentation. “IDW (Spatial Analyst)” <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/idw.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 | **22** |
| **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 | **96** |