

Lab Report

Title: Lab 4

Notice: Dr. Bryan Runck

Author: Zhaoxu Sui

Date: 12/1/2021

Project Repository: <https://github.com/KennethSui/GIS5571/tree/main/lab4>

Google Drive Link: https://drive.google.com/drive/folders/1ZZEpn-FVDsQ9uIDry_Uo6p3tjueFW7uK?usp=sharing

Time Spent: 5 hours

Abstract

In this project, I used a ETL to request the air temperature data for all monitor stations in NDAWN system. By importing the csv file, I acquired to the ArcGIS Pro, I performed dissolve tool to find the average temperature within the last 30 days for all these stations. Finally, I created three raster surfaces by using different kinds of interpolation methods, IDW, Kriging, and Spline.

Problem Statement

My objective is to get familiar with different interpolation methods with actual ArcPy practices. Within this exercise, I will utilize three different kinds of interpolation methods, including two methods of deterministic interpolation and one method of stochastic interpolation.

Also, in this lab exercise, I have an opportunity to review the ETL process and the ways to perform the tools in ArcPy.

Sheet 1: Required Datasets for this lab task

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Air Temperature Data	A database which could represent the average temperatures of monitor stations in Minnesota.	CSV File to Vector Feature Data Layer	Average Air Temperature Data per Hour per station	NDAWN	Derived by a built ETL request

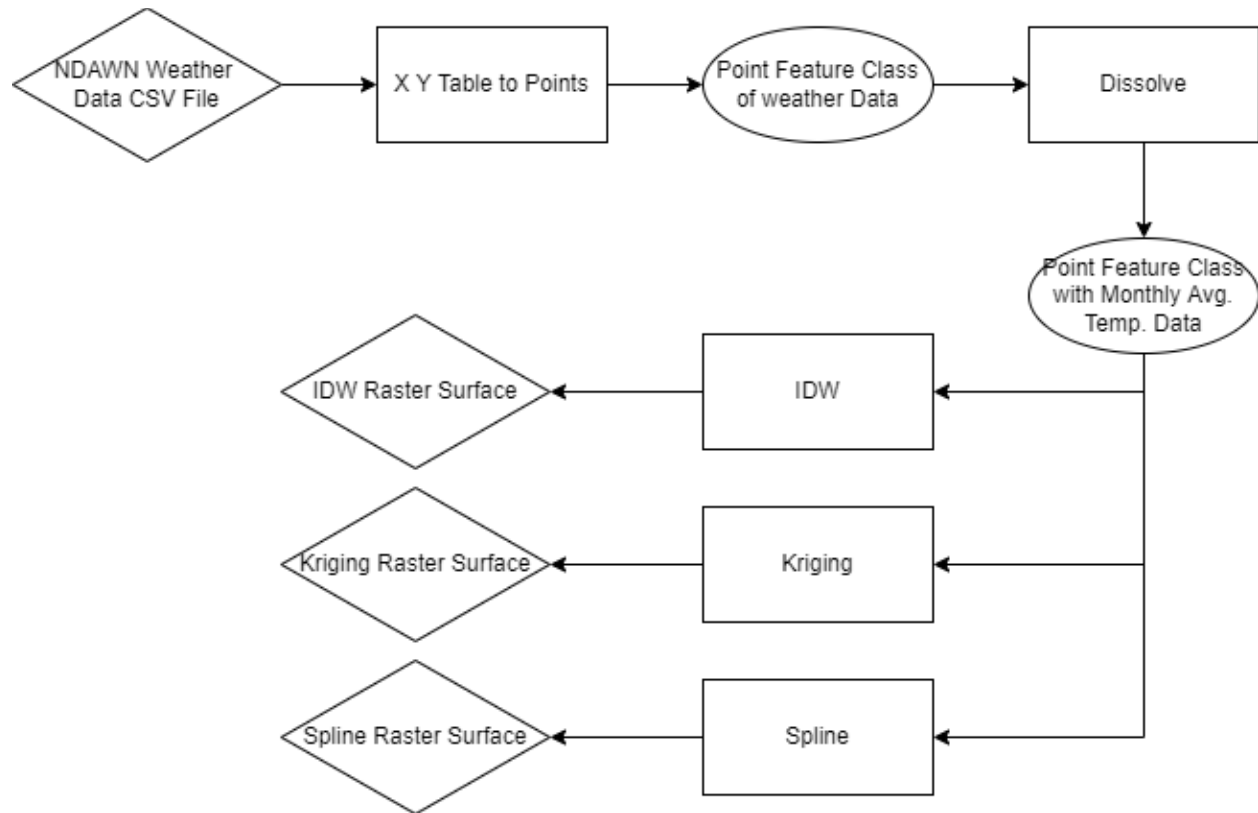
Input Data

Sheet 2: Input datasets

#	Title	Purpose in Analysis	Link to Source
1	Meteorological Data of weather monitor stations in Minnesota	To create raster surfaces to represent the continuous surfaces of temperatures in this area.	<u>NDAWN</u>

Methods

Figure 1 Flow Chart of the Whole Process



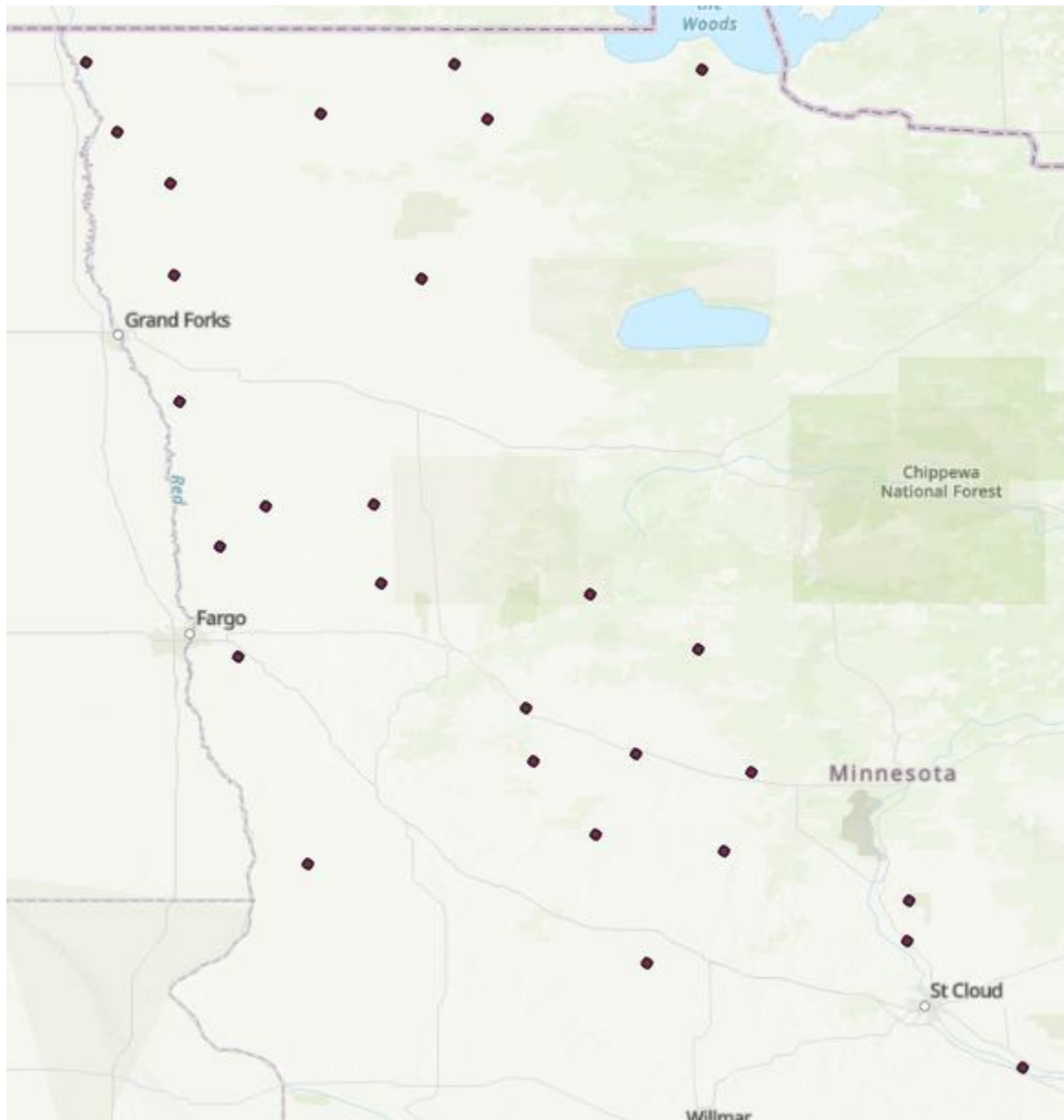
The whole process is programmed in the ArcGIS Pro notebook with the utilization of Arcpy functions.

Firstly, I used the given ETL codes to request the data of all monitor stations with their hourly air temperatures, and I export the data into a CSV file inside of the root directory of ArcPro Project.

Inside of the ArcPro Notebook, I used the X Y Table to Points function to import the data to my current map, but at this point, all the data are raw with points of same stations but different temperatures overlapping together.

Then, I utilized Dissolve functions to delete the overlapping points, with defined parameters of calculated average temperatures in each station. Now, we have one single point representing its own monthly average temperature.

Figure 1 Stations as points in the ArcGIS Pro



Finally, I generated three raster layers with different interpolation methods, IDW, Kriging, and Spline. The output three raster layers are the results of my lab exercise.

Results

After the computation, I found all three raster tables seems different from each other, but all these results show the general patterns of a decrease in the temperature as latitude increases. The differences including the smooth. For example, the IDW seems not smooth than Spline if we refer to the isotherms simulated by the interpolations. Also, we can see some cold spots and hot spots of splines and IDW interpolations more than Kriging method.

The Map of Stations with their temperatures are attached in the github directory.

Figure 2 IDW raster surface with a power of 1

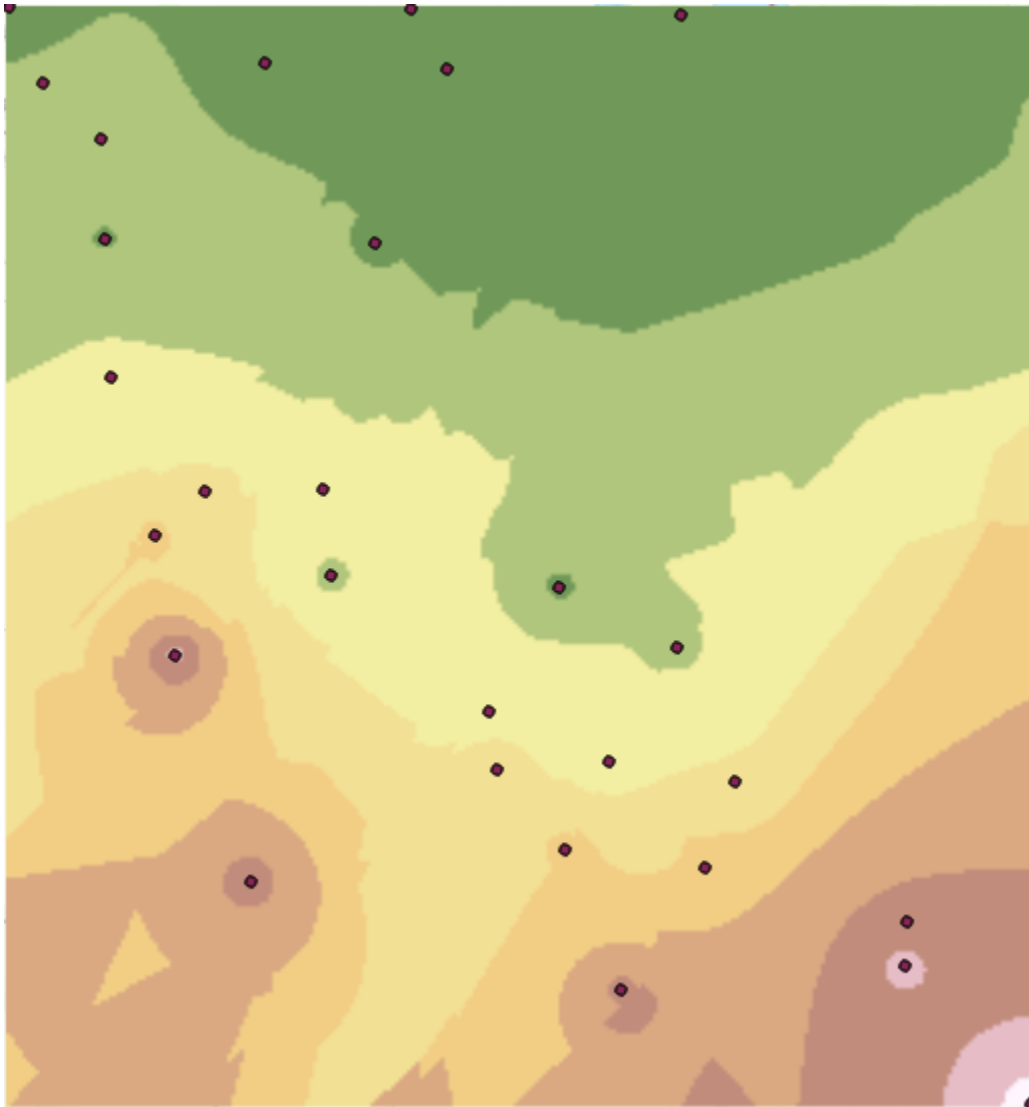
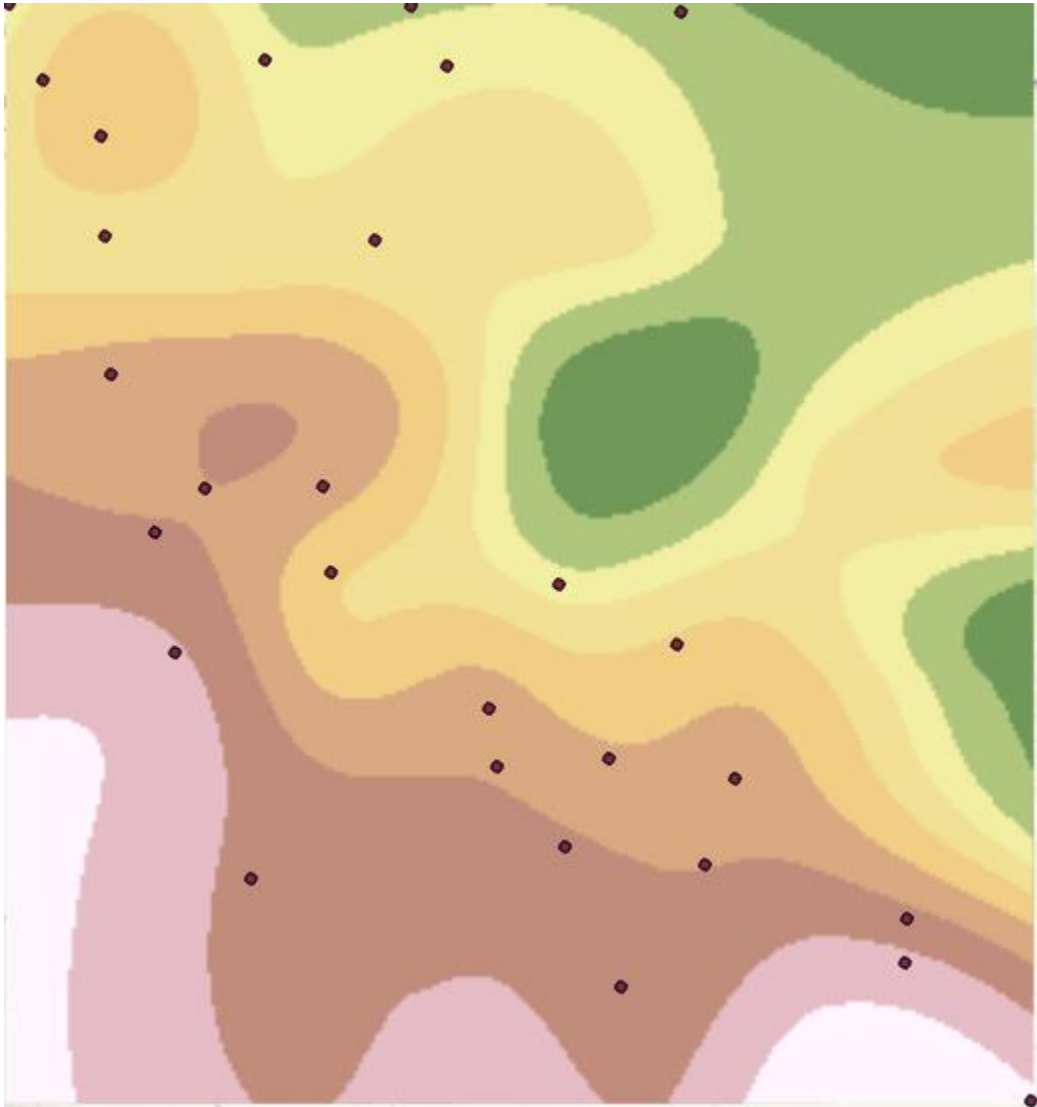


Figure 3 Kriging Interpolation Result



Figure 4 Spline Interpolation Results



Results Verification

This exercise is relatively simple, which means all the data are through carefully examined and verified. Despite the result codes in the Jupyter notebooks, the pictures above also shows that the approach of making these results make sense.

Discussion and Conclusion

In this exercise, utilized three different interpolation methods, IDW, Kriging, and Spline. IDW and Spline are deterministic interpolation, which means the functions are pre-assigned before computing the data. Kriging, however, is a stochastic or geostatistical interpolation, which will build a function according to the inherited spatial pattern of the given point datasets. Thus, Kriging designs different functions for different point datasets to be interpolated.

In this case, the advantage of IDW is that we can assign a small power into it. Due to the nature of temperature, it is hard to see whether there is a cluster of high values in a small area, so the distant points might be playing a huge part into our simulation. Whereas the power functions in IDW helps us to simulate the inherited nature of temperature better than other functions in this aspect.

The Kriging function, in turns, involves a strong prediction (GISGeography, 2021) because of its powerful semi-variograms generated from the given data. It is really complicated and compute intensive (Gentile et al, 2013. Table 8), which can be its drawbacks to this exercise. However, it gives a smoother surface with fewer abrupt depressions than the IDW methods. That is the reason I always prefer Kriging unless I have a confidence power to assign in IDW interpolation.

It is hard to understand what Spline does mathematically, but it is designed to avoid the abrupt changes in the raster surface, which is called Runge's phenomenon in math, that occurred in IDW and Kriging. On the base of preserving the credible simulated values, the Spline is, which I believe, most visual-pleasing method to generate the data interpolation. If there is not a high requirement on the accuracy of interpolation, I would like to do Spline more than Kriging function.

In conclusion, all three data interpolation methods I chose in this exercise are helpful in different scenarios. IDW works if you have confidence on how much distance weighted in the interpolation. Kriging works if you have no idea about the dataset but must do the interpolation, since it is versatile, and not pre-determined. The Spline function is visually pleasing and easy to do. After this exercise, I wish I can see more interesting interpolation methods in the future!

References

Esri. Classification trees of the interpolation methods offered in Geostatistical Analyst. Last Retrieved December 1, 2021.

<https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm>

Esri. How inverse distance weighted interpolation works (2021). Last Retrieved December 1, 2021. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm>

Gentile, Courbin, F., & Meylan, G. (2013). Interpolating point spread function anisotropy. *Astronomy and Astrophysics (Berlin)*, 549, A1. <https://doi.org/10.1051/0004-6361/201219739>

GISGeography. Kriging Interpolation – The Prediction Is Strong in this One (2021). Last Updated: October 28, 2021. Last Retrieved December 1, 2021. <https://gisgeography.com/kriging-interpolation-prediction/>

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	24
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	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	18
		100	93