Lab Report

Title: Lab 3 Part 2 Deliverable
Notice: Dr. Bryan Runck
Author: Taylor Andersen-Beaver
Date: 12/3/22

Project Repository: https://github.com/and03449/GIS5571.git
Google Drive Link: N/A
Time Spent: 30 hours

Abstract

The North Dakota Agricultural Weather Network (NDAWN) API has extensive weather data from 141 active weather stations. For this lab we are interested in the maximum temperatures, minimum temperatures, and average temperatures of the last 30 days. Weather data is typically estimated from active weather stations and then uses a series of interpolation techniques to assume weather patterns where there are no active weather stations. In this lab we will be using weather data from 141 active weather stations and comparing multiple interpolation methods to see which will work best for certain situations including Inverse Distance Weighting, Kriging, and Local Polynomial Interpolation.

Problem Statement

Using weather data from the last 30 days in 141 active stations, what is the estimated weather in areas that do not have an active weather station? Exploring multiple interpolation methods will result in many different visualizations of weather assumptions that can be compared to see the different interpretations.

Table 1. Steps Required to Analyze the NDAWN API

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	NDAWN API	API used to search for weather data in North Dakota and some surrounding states	Weather station locations	Temperatures, Wind, Precipitation	CSV with multiple possible variables	Understanding how to search the API for data
2	Pandas Python Package	A packaged used in Python to be able to manipulate data tables	N/A	N/A	NDAWN weather data	Import Pandas to ArcGIS Pro
3	Weather Station Locations	XY coordinates of the 141 active weather stations	XY coordinates	Station Name	NDAWN weather data	Export all stations from NDAWN API

4	Temperature information from the last 30 days	Daily min, max, and avg temperature from each weather station over the last 30 days	XY coordinates	Station Nane, min temps, max temps, avg temps	NDAWN weather data	Export min, max, and avg temps from each station
5	Data manipulation	Finding the 1 min, max, and avg temp of the 30 for each weather station	XY coordinates	Station Name, min temps, max temps, avg temps	NDAWN weather data	Using Pandas and the .groupby function to find the min, max, and avg temp of each weather station
6	Analyze possible interpolations	Understanding each possible interpolation for the weather data and choosing the 3 most appropriate	XY coordinates	Station name, min temps, max temps, avg temps	link	Manipulated weather dataset

Input Data

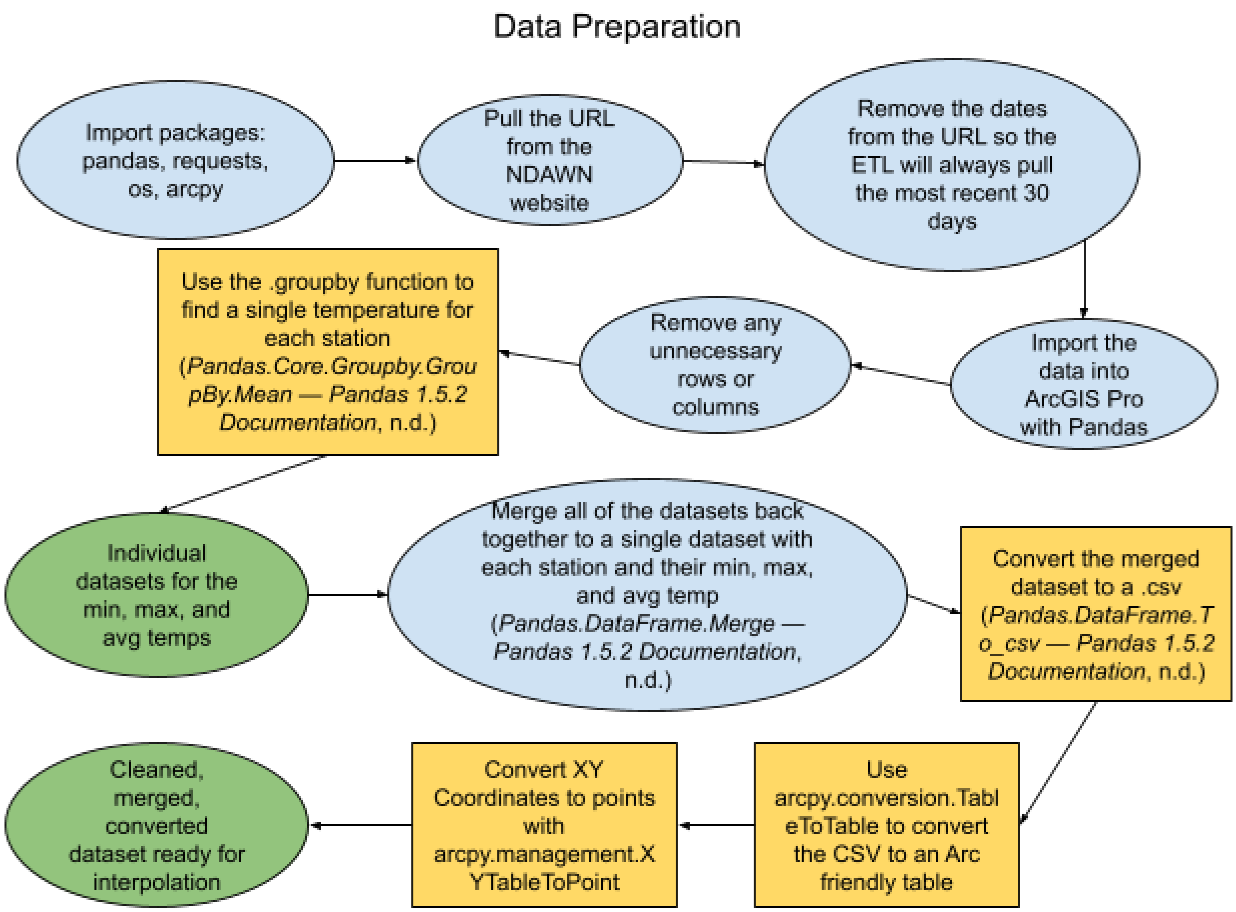
The data being used in this lab is the minimum, maximum, and average temperatures from all 141 active weather stations being collected by NDAWN over the last 30 days. These 30 days are defined by the day that the ETL is run and will include the 30 days prior, i.e. this does not have set days that are being collected each time the ETL is run. The input data will have many more rows than I will want for the weather interpolation - each weather station will have 30 minimums, 30 maximums, and 30 averages and I wanted a single min, max, and average for each station. This will require some manipulation of the data once it has been pulled by the ETL.

Table 2. Dataset from NDAWN API for Weather Data from the Last 30 Days

#	Title	Purpose in Analysis	Link to Source
3	NDAWN Weather data	To have a dataset with temp minimums, maximums, and averages at all NDAWN weather stations over the last 30 days	Link

Methods

The first step to be able to do any interpolation will be to prepare the data. Because the ETL pulled all of the maximums, minimums, and averages for each station for each of the 30 days I am going to use the Pandas package to find the highest maximum, lowest minimum and the average temperature for each station.



Now that the data has been manipulated to work for the interpolation methods in ArcGIS Pro I have to determine which methods will be the most appropriate for this project. To decide which method of interpolation to use I researched quite a few different methods listed in the table below.

Table 3: Determining the appropriate interpolation methods for the NDAWN weather data

Interpolation Method	Brief Summary	Decision to Use	Source
Global Polynomial Interpolation (GPI)	Fits a smooth surface (the polynomial) to the sample points. Best if the changes occur gradually and will capture coarse-scale patterns	No - I am afraid this will not be a great interpolation because the changes may not be as gradual as needed	(*How Global Polynomial Interpolation Works—ArcMap | Documentation*, n.d.)
Local Polynomial Interpolation (LPI)	Similar to GPI but best when used with short-range variation and it is sensitive to neighboring distance	Yes - this will be a good interpolation method because of the short-range variation and there are plentiful neighbors	(*How Local Polynomial Interpolation Works—ArcMap | Documentation*, n.d.)

Inverse Distance Weighted (IDW)	IDW is a very commonly used interpolation as it follows the simple concept that thing that are close are more alike however it can be sensitive to clustering and outliers	Yes - as mentioned before IDW is a commonly used interpolation because of its simple principle and I will be able to test multiple powers seeing the sensitivity of the weights	(*How Inverse Distance Weighted Interpolation Works—ArcMap | Documentation*, n.d.)
Radial Basis Functions (RBF)	Creates a smooth surface of a large number of points. This interpolation is not a good choice if there are large changes in short distances	No - although this interpolation would probably provide decent results, I am only working with a small number of points so I will focus on different interpolations	(*How Radial Basis Functions Work—ArcMap | Documentation*, n.d.)
Kernel Interpolation w/Barriers (KSB)	KSB is a variation of LPI but adds a small amount of bias to help prevent prediction errors.	No - KSB only considers the next closest point in its interpolation and I would prefer an average of the nearest points	(*How Kernel Interpolation With Barriers Works—ArcMap | Documentation*, n.d.)
Diffusion Interpolation w/ Barriers (DKB)	Diffusion interpolation is the process in which heat or particles diffuse with time and will flow over barriers smoothly	No - although this may be useful for future interpolation of weather data, I have found other interpolations I would prefer to use.	(*How Diffusion Interpolation With Barriers Works—ArcMap | Documentation*, n.d.)
Ordinary Kriging (OK)	OK interpolation has good flexibility and uses an assumed constant and estimates error	No - OK would be fine but is a little too simple and there is a better option	(*Understanding Ordinary Kriging—ArcMap | Documentation*, n.d.)
Simple Kriging (SK)	SK is similar to OK but there is a known trend and a known error	No - same as OK	(*Understanding Simple Kriging—ArcMap | Documentation*, n.d.)
Universal Kriging (UK or RK (Regression Kriging))	UK is different from OK and SK that there is not an assumed constant but rather a deterministic function	No - although UK is essentially a basic regression calculation there are better options	(*Understanding Universal Kriging—ArcMap | Documentation*, n.d.)
Indicator Kriging (IK)	Similar to OK but the values stay between 0 and 1. It also creates a binary variable	No - same as OK	(*Understanding Indicator Kriging—ArcMap | Documentation*, n.d.)
Probability Kriging (PK)	Same as IK but estimates two errors instead of one which involves a lot of estimating	No - too much estimating than it is worth. There are better options.	(*Understanding Probability Kriging—ArcMap | Documentation*, n.d.)

Disjunctive Kriging (DK)	IK is a special case of DK which predicts a value or an ‘indicator’	No - mathematically complicated and difficult to verify	(*Understanding Disjunctive Kriging—ArcMap | Documentation*, n.d.)
Empirical Bayesian Kriging (EBK)	EBK estimates parameters automatically whereas other forms of Kriging are adjusted manually.	Yes - more accurate than other forms of Kriging especially when using smaller datasets	(Krivoruchko & Gribov, 2019)

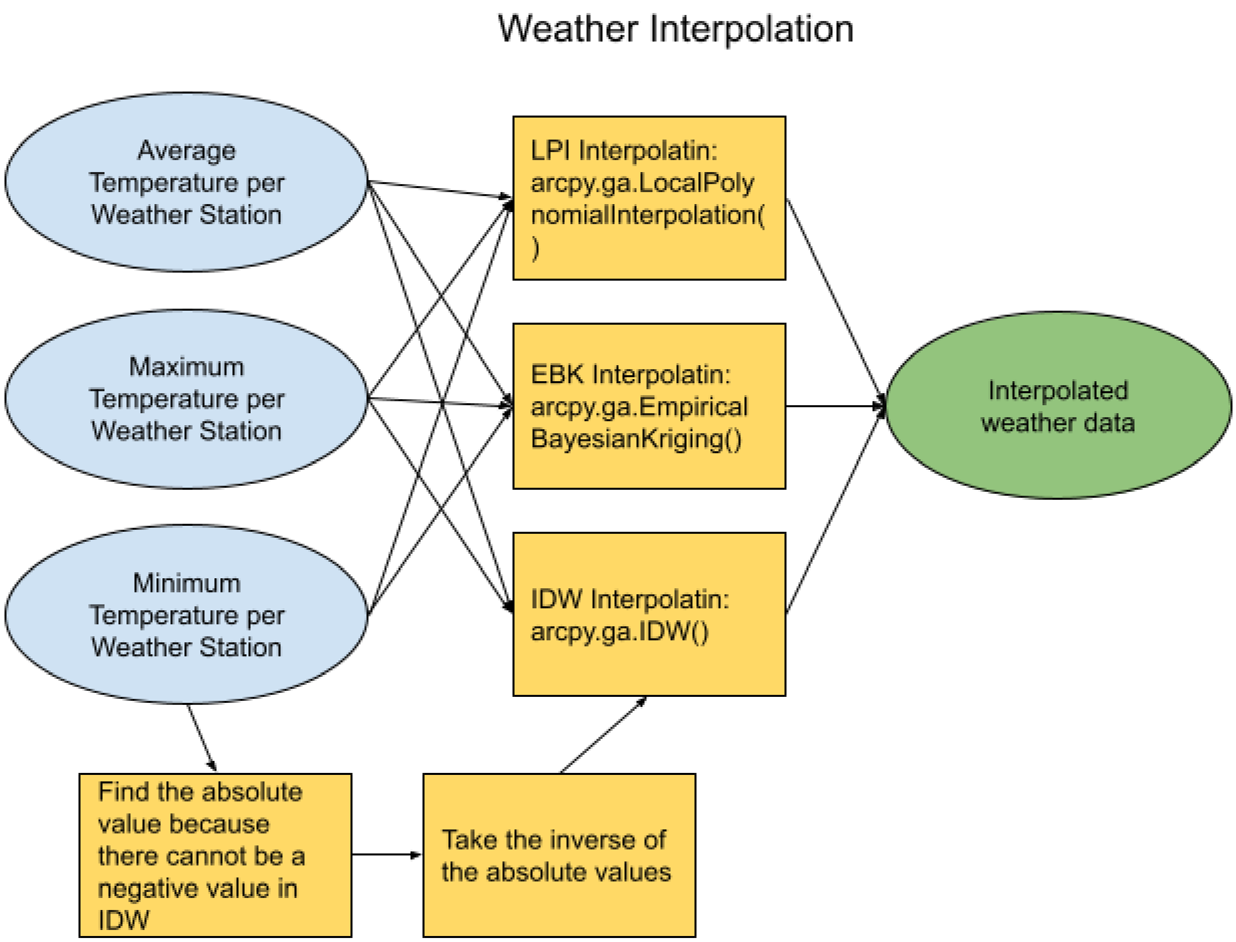
The IDW was an easy choice when it came to deciding which methods to use as it is a widely used interpolation method because of the simple concept that things that are near are more alike. So even though this was a simple choice to make I wanted to investigate a little further to fully understand what IDW does and what I can expect to see in the results. IDW is used to “predict a value for any unmeasured location” using the “measured values surrounding the prediction location” (Hodam et al., 2017). IDW assumes spatial autocorrelation which “assumes that each measured point has a location influence that diminishes with distance” (Hodam et al., 2017). While looking at the IDW function in ArcGIS Pro I determined the “Power” parameter to be the most useful in seeing changes in the IDW output. A higher power value results in less influence from distant points.

Kriging is a “geostatistical spatial interpolation” method that “measures the degree of spatial dependence among the known points” (Dalmau et al., 2017). As there are many Kriging options I did a good amount of research to help me decide which would be the best for this dataset. The main types of Kriging used are Simple Kriging, Ordinary Kriging, and Universal Kriging. Simple Kriging (SK) assumes a constant that covers the entire study area - because this study area varies so much in temperatures I decided this would not be a good fit (Dalmau et al., 2017). Ordinary Kriging (OK) is a little better in that instead of assumed a constant over the entire area it assumes a constant “only in the local neighborhood of each estimation point” - this is better but does not assume any error (Dalmau et al., 2017). Universal Kriging (UK) is similar to Ordinary Kriging except instead of assuming a constant it uses a deterministic function that varies for each location (Dalmau et al., 2017).

All of these would have been fine interpolation methods to use for the weather data but I decided to use Empirical Bayesian Kriging (EBK) instead. EBK “is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model” (Krivoruchko & Gribov, 2019). All of the other methods (SK, OK, and UK) require manual adjustments to the parameters where EBK does this automatically through subsetting and simulations (Krivoruchko & Gribov, 2019). EBK also accounts for error by “estimating the underlying semivariogram”, which other methods do not, which means that these other methods may be underestimating the errors (Krivoruchko & Gribov, 2019).

Lastly, I was deciding between Global and Local Polynomial Interpolation - both of which are uncommon to use in this sort of interpolation but piqued my interest (Hadi & Tombul, 2018). “Global interpolation is based on the assumption that the entire study area can be represented by a general formula” whereas Local interpolation uses a local formula determined by the measured points (Hadi & Tombul, 2018). Ultimately I decided to use Local Polynomial Interpolation because it is local though it can be inexact (Hadi & Tombul, 2018).

Once my interpolation methods were selected it was a simple ‘plug-and-chug’ with the minimum, maximum, and average temperatures for each of the stations. The only exception to this was the minimum temperatures when calculating the IDW - because there were some negative values in the minimum temperatures I had to take the absolute value of these temperatures which spiked them all up into positive values and then take the inverse of all of the temperatures to the scale that more closely aligned with the original temperatures.

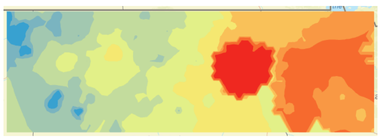
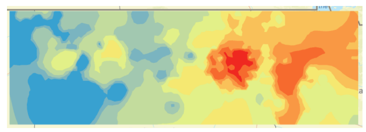
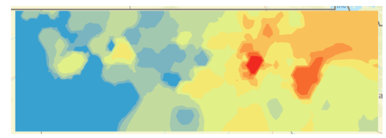
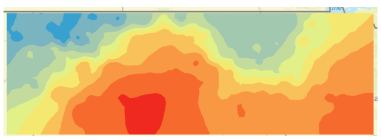
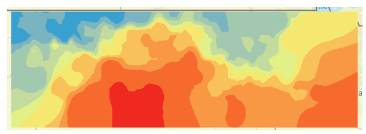
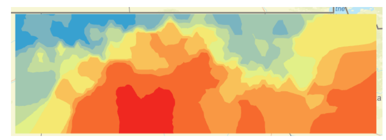
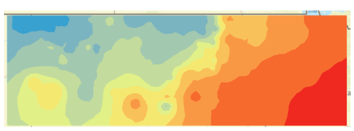
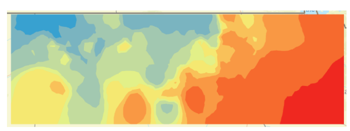
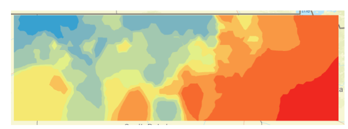


Results

The results, overall, were about as I would expect the temperatures to be in North Dakota around this time of year, with the exception of a few heat waves. Typically the eastern part of the state is warmer because there is more urban development there in both eastern North Dakota and western Minnesota. In addition, the western part of North Dakota has more gusting winds that fuel cooler temps. Of course, the northern parts of the state are typically cooler than the southern parts as well.

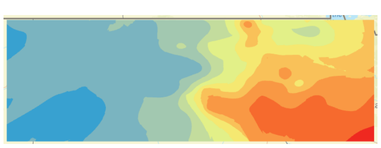
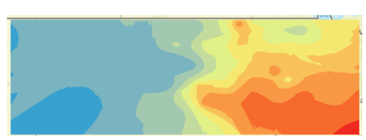
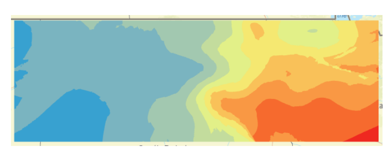
Looking at the IDW interpolation, the change in the power parameter gives a good example of how the distance to another data point can affect the interpolation - the higher the power value the less of an influence distance has on the interpolation which means the lower the power value the more smooth the interpolation appears. At a power of 2 the interpolation is relatively smooth across the surface with a few hot and cold spots respectively. Up to a power of 10 there are many more pockets of hot and cold spots that interrupt the smooth surface. Although overall the surface is never really smooth compared to some of the other interpolation methods below.

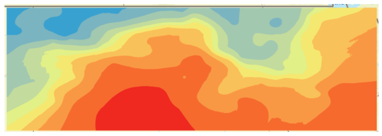
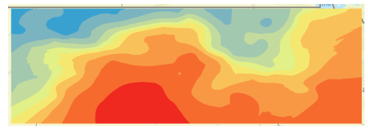
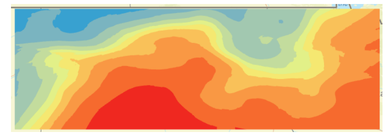
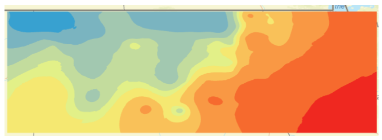
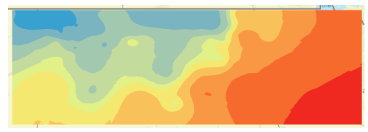
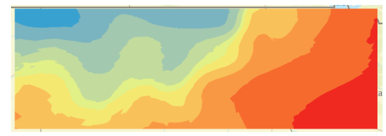
Table 4: Results from Inverse Distance Weighted Interpolation of NDAWN Weather Data

IDW			
	Power 2	Power 5	Power 10
Minimum Temps (-21 - 7 deg)			
Maximum Temps (61 - 79 deg)			
Average Temps (20-34 deg)			

Of the three different interpolation methods used, EBK did the best at eliminating outliers and has the most smooth surface. The hot and cold spots in each of the minimum, maximum, and average temps are clear in the images in Table 5 and smoothly glide across the surface without too many outlier interruptions. For scope and time's sake for this lab I will not get into each of the parameters of the EBK specifically but I did want to test each of them to visually be able to see which would have the best representation and Spline has overall the most smooth surface with little interruption.

Table 5: Results from Empirical Bayesian Kriging Interpolation of NDAWN Weather Data

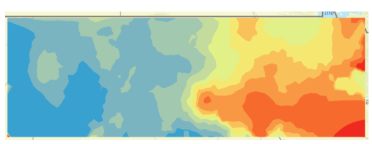
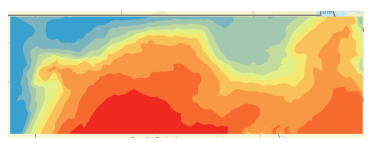
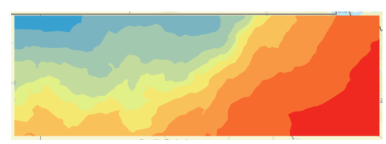
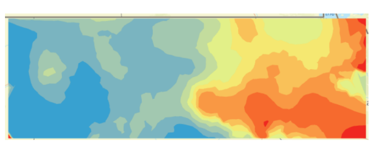
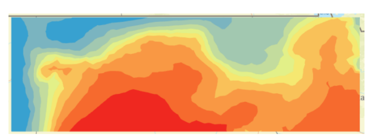
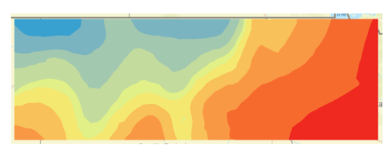
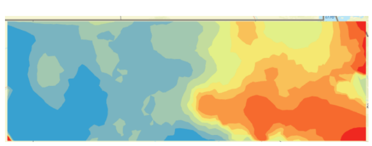
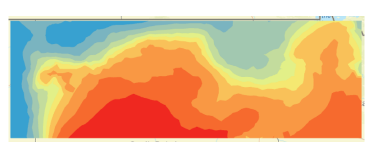
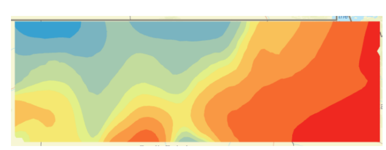
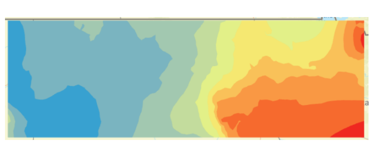
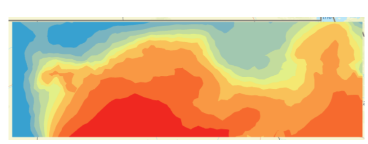
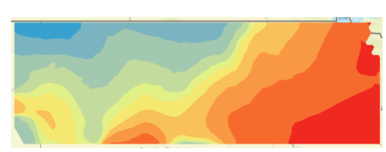
EBK			
	Linear	Power	Spline
Minimum Temps (-21 - 7 deg)			

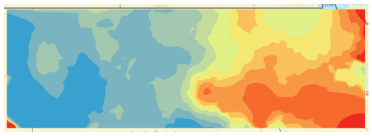
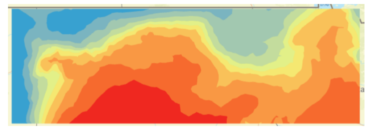
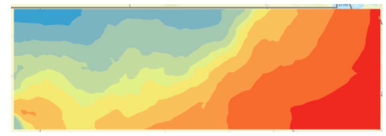
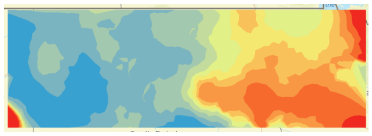
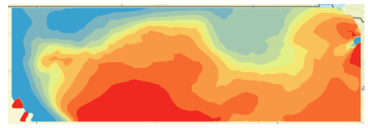
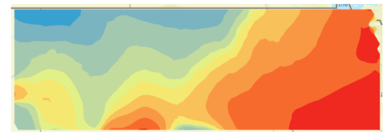
Maximum Temps (61 - 79 deg)			
Average Temps (20-34 deg)			

The final interpolation method I used was the Local Polynomial Interpolation which, as stated above, is not a commonly used method for interpolating weather data but I was interested to see why. I did not investigate too much into the different parameters available in ArcGIS Pro but did find that the Kernel Function is probably the one to show the biggest difference between the interpolations.

What is interesting about the LPI is that some of the interpolations are actually smoother than the examples from EBK but not all of them. The average temperatures, for example, have a very smooth surface but the minimum temperatures have many more interruptions. This leads me to believe that the LPI is not commonly used for this type of interpolation because it is inconsistent.

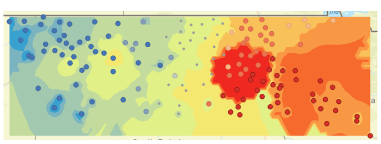
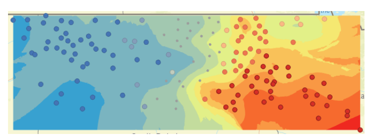
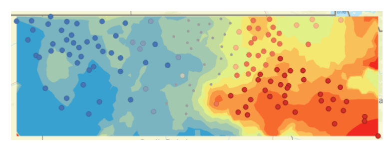
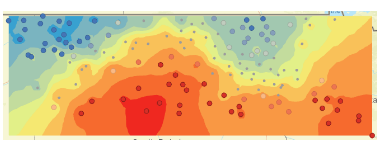
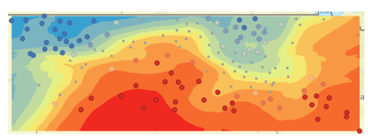
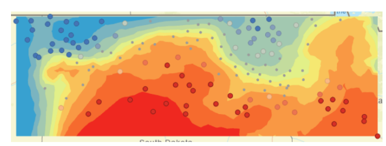
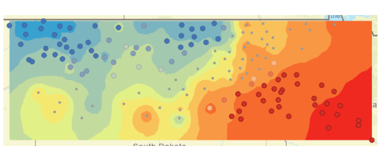
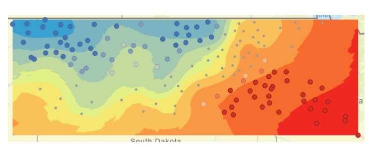
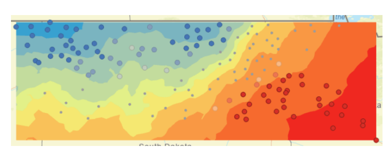
Table 6: Results from Local Polynomial Interpolation of NDAWN Weather Data

LPI			
Kernel Function	Minimum Temps (-21 - 7 deg)	Maximum Temps (61 - 79 deg)	Average Temps (20-34 deg)
Constant			
Exponential			
Gaussian			
Quartic			

Epanechnikov			
Fifth-order Polynomial			

Results Verification

To verify my results I decided to run a Hot and Cold Spot function on all of the weather stations and their minimum, maximum, and average temperatures. The Hot and Cold Spot function does verify the results I received that the temperatures are warmer over the more developed, urban areas and colder in the rural easter part of the state of North Dakota. Many of the Cold Spots (blue points) fall into the cold (blue) shaded interpolated areas and many of the Hot Spots (red points) fall into the warm (red) shaded interpolated areas. Running a Hot and Cold Spot function also verified some of the concerns I have about the interpolations themselves and where they fall short.

Hot/Cold Spots			
	IDW - Power 2	EBK - Spline	LPI - Constant
Minimum Temps (-21 - 7 deg)			
Maximum Temps (61 - 79 deg)			
Average Temps (20-34 deg)			

Discussion and Conclusion

If I were to do this lab again and had more time I would have done a few things differently. For one, I would reclassify the temperatures for each of the minimum, maximum, and average datasets so that they are all on the same scale and it doesn't appear as though the hot spots values are the same in the minimum as the maximum, for example. Another aspect I would change would be to further investigate all of the parameters available for each of the interpolation methods to be able to further understand the sensitivity of each parameter. I did not test most of the default settings in ArcGIS Pro and would like to revisit those to understand the differences better.

The purpose of this lab was to take a current weather dataset from NDAWN, investigate different interpolation methods, and perform the chosen functions in ArcGIS Pro. In order to be able to test the interpolation of the data, it had to be cleaned from the raw data pulled from the NDAWN website. Once the data was cleaned and manipulated to have the appropriate data needed for the interpolation I was able to perform multiple interpolation functions to see how these functions. I would say overall the EBK interpolation did the best in representing the temperature surface as it matches most closely with the Hot & Cold Spots.

References

Dalmau, R., Perez-Batlle, M., & Prats, X. (2017). Estimation and prediction of weather variables from surveillance data using spatio-temporal Kriging. *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 1–8. https://doi.org/10.1109/DASC.2017.8102132

Hadi, S. J., & Tombul, M. (2018). Comparison of Spatial Interpolation Methods of Precipitation and Temperature Using Multiple Integration Periods. *Journal of the Indian Society of Remote Sensing*, *46*(7), 1187–1199. https://doi.org/10.1007/s12524-018-0783-1

Hodam, S., Sarkar, S., Marak, A. G. R., Bandyopadhyay, A., & Bhadra, A. (2017). Spatial Interpolation of Reference Evapotranspiration in India: Comparison of IDW and Kriging Methods. *Journal of The Institution of Engineers (India): Series A*, *98*(4), 511–524. https://doi.org/10.1007/s40030-017-0241-z

How Diffusion Interpolation With Barriers works—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-diffusion-interpolation-with-barriers-works.htm

How global polynomial interpolation works—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-global-polynomial-interpolation-works.htm

How inverse distance weighted interpolation works—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm

How Kernel Interpolation With Barriers works—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-kernel-interpolation-with-barriers-works.htm

How local polynomial interpolation works—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-local-polynomial-interpolation-works.htm

How radial basis functions work—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-radial-basis-functions-work.htm

Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, *32*, 100368. https://doi.org/10.1016/j.spasta.2019.100368

pandas.core.groupby.GroupBy.mean—Pandas 1.5.2 documentation. (n.d.). Retrieved November 28, 2022, from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.core.groupby.GroupBy.mean.html

pandas.DataFrame.merge—Pandas 1.5.2 documentation. (n.d.). Retrieved November 28, 2022, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.html

pandas.DataFrame.to_csv—Pandas 1.5.2 documentation. (n.d.). Retrieved November 28, 2022, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html

Understanding disjunctive kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-disjunctive-kriging.htm

Understanding indicator kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-indicator-kriging.htm

Understanding ordinary kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm

Understanding probability kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-probability-kriging.htm

Understanding simple kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-simple-kriging.htm

Understanding universal kriging—ArcMap | Documentation. (n.d.). Retrieved November 29, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-universal-kriging.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	**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**