**Lab Report**

Title: Exploring Interpolation Methods on 30-day NDAWN Temperature Datasets in ArcGIS Pro

Notice: Dr. Bryan Runck

Author: Maisong Francis

Date: 4/22/2021

**Project Repository:** [*https://github.com/msongfrancis/GIS5572.git*](https://github.com/msongfrancis/GIS5572.git)

**Abstract**

Temperature data is dynamic and can be published frequently. Tools that can retrieve temperature data can be utilized on the fly to update and add-on to datasets. This type data is also collected at stations like NDAWN stations and due to the nature of that, is missing values where stations are not located. Interpolation can be used to solve this problem by creating and inferring the values in between based on the existing sample values collected. There are different methods of interpolation such as IDW, various Kriging methods, and RBF. There are differences such as exact and non-exact consideration of sample values between these different methods. They are all able to produce an interpolated surface. Overall, there is no best method, but in some studies, various types of Kriging methods have produced the “best” interpolated surfaces with considerations to the scenario variables.

**Problem Statement**

Temperature data is dynamic and changes over time. NDAWN provides temperature data on the web for different temporal resolutions. To examine data on a monthly frequency, 30 days of temperature data must be extracted on the fly. Temperature data like maximum and minimum daily temperatures can also be extracted on the fly to examine the high and low temperatures. The high temperature is defined in this lab as the largest value of the maximum daily temperatures recorded for the 30-day period. The low temperature is defined as the lowest value of the minimum daily temperature recorded for the 30-day period. Furthermore, NDAWN stations are dispersed unevenly. Temperatures in areas between station collection sites must be inferred through interpolation, which can be done with different methods.

Table 1. *Components needed to get NDAWN temperature data and interpolation analysis.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **Spatial Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | NDAWN station information | Where NDAWN stations are located and their name. | Point geometry | Station lat and long |  | Extract lat long from html and create a table for each station |
| 2 | Temperature data | Average, minimum, and maximum daily temperature for the 30-day period for each NDAWN station |  | Average temp, minimum temp, maximum temp |  | Extract from NDAWN site |

**Input Data**

The information about the NDAWN stations and the daily temperature data can be extracted form the NDAWN website. The information was extracted from the CSV and HTML for each station. The time which all data is outputted and calculated is Feb 01, 2021 to March 03, 2021.

Table 2. *Data needed to perform analysis.*

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | NDAWN stations coordinates | Used to create point features to show where stations are located. | [NDAWN](https://ndawn.ndsu.nodak.edu/) |
| 2 | Highest temperature | Maximum daily temperature collected by each station. | [NDAWN](https://ndawn.ndsu.nodak.edu/) |
| 3 | Lowest temperature | Minimum daily temperature collected by each station. | [NDAWN](https://ndawn.ndsu.nodak.edu/) |
| 4 | Average daily temperature | Used to calculate the average 30-day temperature | [NDAWN](https://ndawn.ndsu.nodak.edu/) |

**Methods**

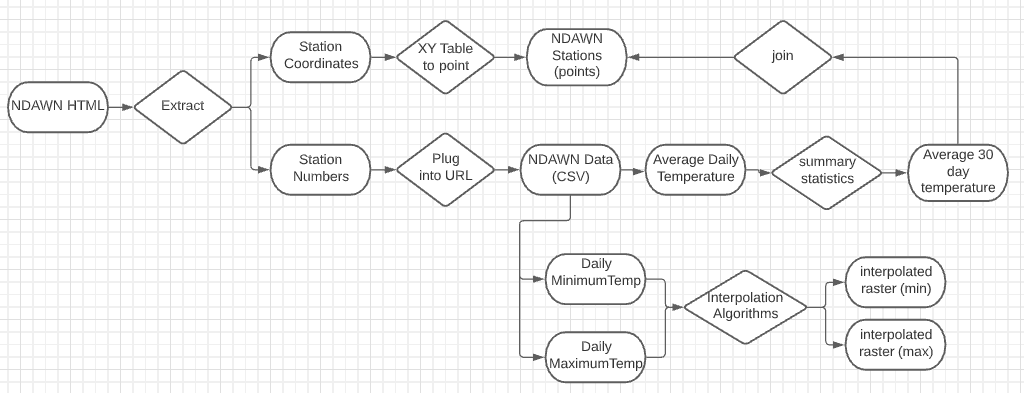
*Get NDAWN Data*

The station names and numbers were obtained by extracting the information from the HTML of the NDAWN website. The URL for each station was constructed after obtaining the station numbers and the coordinates for each station were obtained once again from the HTML. The temperature data was obtained from the CSV for each station containing the daily average temperature, daily maximum temperature, and daily minimum temperature (variable codes: ddavt,ddmxt,ddmnt respectively).

*Interpolating Maximum and Daily Temperatures*

A point feature class of the stations were created using the coordinates obtained. The high and low from each station for the 30-day period was calculated using the Summary statistics tool. Then the high low temperature data was joined to the stations feature class and different interpolation tools, specifically IDW, Kriging, Radial Basis Functions, and Empirical Bayesian Kriging, were used to interpolate the low and high temperatures for the data extent.

Figure 1. Data flow diagram to get 30-day NDAWN data and perform interpolation analysis in ArcGIS Pro.



**Results**

There was a total of 131 NDAWN stations. The daily temperatures collected by the NDAWN stations varied. The minimum daily temperatures for all stations had a range from -40.864 to 34.704 and a maximum temperature range of -21.388 to 62.24. The average temperature for the 30-day period for all stations ranged from -0.37 to 11.72 F (Fig. 2).

The interpolated results for the 30-day period high and low temperatures from Feb 01, 2021 were different for each algorithm used. Each output raster was symbolized using natural break classification with 10 classes. The IDW low interpolated results ranged from -40.86 to -24.41 and the high interpolated results ranged from 38.23 to 62.23. The kriging­­­ low interpolated results ranged from -36.23 to -25.97, and the high interpolated results ranged from 43.43 to 58.91. The EBK low interpolated results ranged from -42.18 to -24.88, and the high interpolated results ranged from 38.46 to 63.53. The RBF low interpolated results ranged from -40.89 to -24.41, and the high interpolated results ranged from 38.06 to 63.27 (Fig 3).

The highest maximum daily temperatures appears to be recorded in the south west portion of North Dakota and the lowest minimum daily temperatures appears to be in areas outside of the red river valley of the north river. In terms of smoothness, EBK produced the most “smooth” interpolated surface and Kriging had the least smotth with edges/straight line sin the interpolated results (Fig.3).

A picture containing chart

Description automatically generated

Figure 2. NDAWN station locations and the average temperatures interpolated using empirical Bayesian kriging from Feb 01.2021 to March 02, 2021.

**Results Verification**

I do not know that the result is correct because it is using different algorithms to interpolate maximum and minimum temperatures. One way to check the accurateness of a produced interpolated is to see how similar the values around each other are. Furthermore, you could test the interpolated results by removing a point and seeing if the interpolated results were like the temperatures collected. Another option to verify the interpolated surface is to look at other resources like satellite sensor datasets.

**Discussion and Conclusion**

Using the interpolation decision tree provided by ArcGIS, I chose to do interpolation with IDW, RBF, and ordinary Kriging, and empirical Bayesian Kriging. First, these methods are suitable because there is one predication per location (the high and low value for each station). Secondly the output type was prediction from the input. Also, IDW and RBF incorporated exact values at the input value locations, while Kriging is not.

The IDW algorithm assumes that areas closer to the point are more alike. The interpolated surface shows that around the station locations the interpolated values are similar, and we get a spotted look where values around the stations are similar but become less similar as they cells are further away. In this IDW interpolation scenario, a distance power of 2 was used, but if a larger value were used, the weights of the distant points would decrease ((ESRI, n.d. -a)

The other exact interpolation method I used was Radial Basis functions, specifically completely regularized spline. This method fits the interpolated values through the sample values which is why it is an exact interpolation. The output surface has similarities with the IDW and kriging. The values closest to the sample points appear to be more similar and as distance increases, less similar. The edges where there are no sample values appear to have the same values for larger areas. According to the ArcGIS Pro documentation, RBF is best for producing smooth surfaces datasets with many datapoints but are inappropriate for large changes in surface values within short distances (ESRI, n.d. -b).

The first Kriging algorithm I used was ordinary kriging. Ordinary kriging assumes an unknown constant mean and can be used for data that seems to have a trend (ESRI, n.d. -c). The overall surface output is not exact at the station points and shows large areas appear to have the same interpolated values (Fig.3). The second kriging algorithm I used was empirical Bayesian kriging. EBK is different from other kriging methods because it accounts for errors in the semi variogram and more accurate with small datasets (ESRI, n.d. -d). The surface produced from the EBK was very smooth apart from areas without many stations/input temperatures. (Fig 3).

The best method for interpolation will vary based on the scenario. Cronqvist (2018) found that using regression kriging always had a smaller mean square error than IDW when interpolating the temperature data for Sweeden where distance was 2.5 km between each sample point. Cronqist (2018) also states that this method could lead to violations if the required assumptions were not known, but in this case, they were known. In another scenario, Hofstra et al. (2008) was testing different interpolation methods for temperature in Europe and found that the best method was global kriging because it was best for their climate variables and produced the smallest interpolation error. In both scenarios, Kriging was used as opposed to other interpolation methods because it was the “best” for the scenario and because it was the best for the estimation of continuous data like temperature.

**References**

CRONQVIST, F. (2018). Interpolation of temperature data for improved weather forecasts. *DEGREE PROJECT IN MATHEMATICS, SECOND CYCLE*, 1–98. https://kth.diva-portal.org/smash/get/diva2:1214240/FULLTEXT01.pdf

ESRI. (n.d.-a). *How inverse distance weighted interpolation works—ArcGIS Pro | Documentation*. ArcGIS Pro. Retrieved April 21, 2021, from https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm

ESRI. (n.d.-b). *How radial basis functions work—ArcGIS Pro | Documentation*. ArcGIS Pro. Retrieved April 21, 2021, from https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-radial-basis-functions-work.htm#:%7E:text=Radial%20basis%20functions%20(RBFs)%20are,Spline%20with%20tension

ESRI. (n.d.-c). *Understanding ordinary kriging—ArcMap | Documentation*. ArcGIS Desktop. Retrieved April 21, 2021, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm

ESRI. (n.d.-d). *What is empirical Bayesian kriging?—ArcGIS Pro | Documentation*. ArcGIS Pro. Retrieved April 21, 2021, from https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*, *113*(D21), 1–19. https://doi.org/10.1029/2008jd010100

Figure 3. Interpolation results for high and low temperatures collected from NDAWN stations from using different interpolation methods: Inverted distance weighted (IDW), kriging, empirical Bayesian kriging (EBK), and radial basis function (RBF).

|  |  |
| --- | --- |
| a) IDW Low Temperature  Graphical user interface  Description automatically generated | b) IDW High Temperature  A picture containing surface chart  Description automatically generated |
| c) Kriging Low Temperature  Graphical user interface  Description automatically generated | d) Kriging High Temperature  A picture containing chart  Description automatically generated |
| e) EBK Low Temperature  A picture containing graphical user interface  Description automatically generated | f) EBK High Temperature  A picture containing surface chart  Description automatically generated |
| g) RBF Low Temperature  Graphical user interface  Description automatically generated with medium confidence | h) RBF High Temperature  Surface chart  Description automatically generated with low confidence |

**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 | **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 | **18** |
|  |  | 100 | **98** |