Lab 3 Part Two

Title: Applying Deterministic Interpolation Four Ways for NDAWN Temperature Data
Notice: Dr. Bryan Runck, Michael Felzan
Author: Alexander Danielson
Date: 11/30/2022

Project Repository: https://github.com/ardumn/GIS5571/tree/main/Lab3
Google Drive Link:
https://drive.google.com/drive/u/0/folders/11cDLxUWDv6QMdVnUu3WcG89pRrwmM9yB
Time Spent: 24 hours

Abstract

When it comes to interpolating data by a set of points in space, or by a grid of pixels. Interpolation methods such as inverse distance weighting (IDW), Kriging, natural neighbors (Voronoi polygons), Triangulated Irregular Networks (TIN), or Splines are primed for spatial analysis and data science. Each method conforms to their own parameters in interpolating points based on a reference point to other sample points i.e., IDW, or probabilistic interpolation based on i.e., Kriging. For this report, utilizing the North Dakota Agricultural Network (NDAWN) API to extract all station's average temperature data for the past thirty days (or month of October) and juxtapose and contrast four interpolation methods. Depending on the data used one methodology can be more advantageous over the other for temperature and spatio-temporal coverage. This is depicted by the map results for each of the NDAWN station points and gleam by an in-depth overview of each interpolation methodology and mostly governed by elevation, distribution of points, number of points, and other statistical implications. Then a literature review is highlighted for each of the following interpolation methods to justify the chosen methodologies and provide substantial evidence why these methods prevail over the other, as IDW can be better for temperature-given continuous measurements compared to perception which is discontinuous measuring in time.

Problem Statement

The goal predominantly is to juxtapose and contrast the plethora of interpolation methods that can be applied to spatial analysis/data science and to decompose their statistical/numerical underpinnings of them. For this study, the development of an ETL that extracts the past thirty days (or October) of temperature data from the North Dakota Agricultural Weather Network (NDAWN) and apply three interpolation methodologies are applied. To justify why we utilized these interpolation(s) methodologies, supportive literature and ESRI documentation will be highlighted about the various iterations and subtypes of interpolations classes. (such as ordinary and universal kriging) Including why one is advantageous over the other in applying them to temperature data.

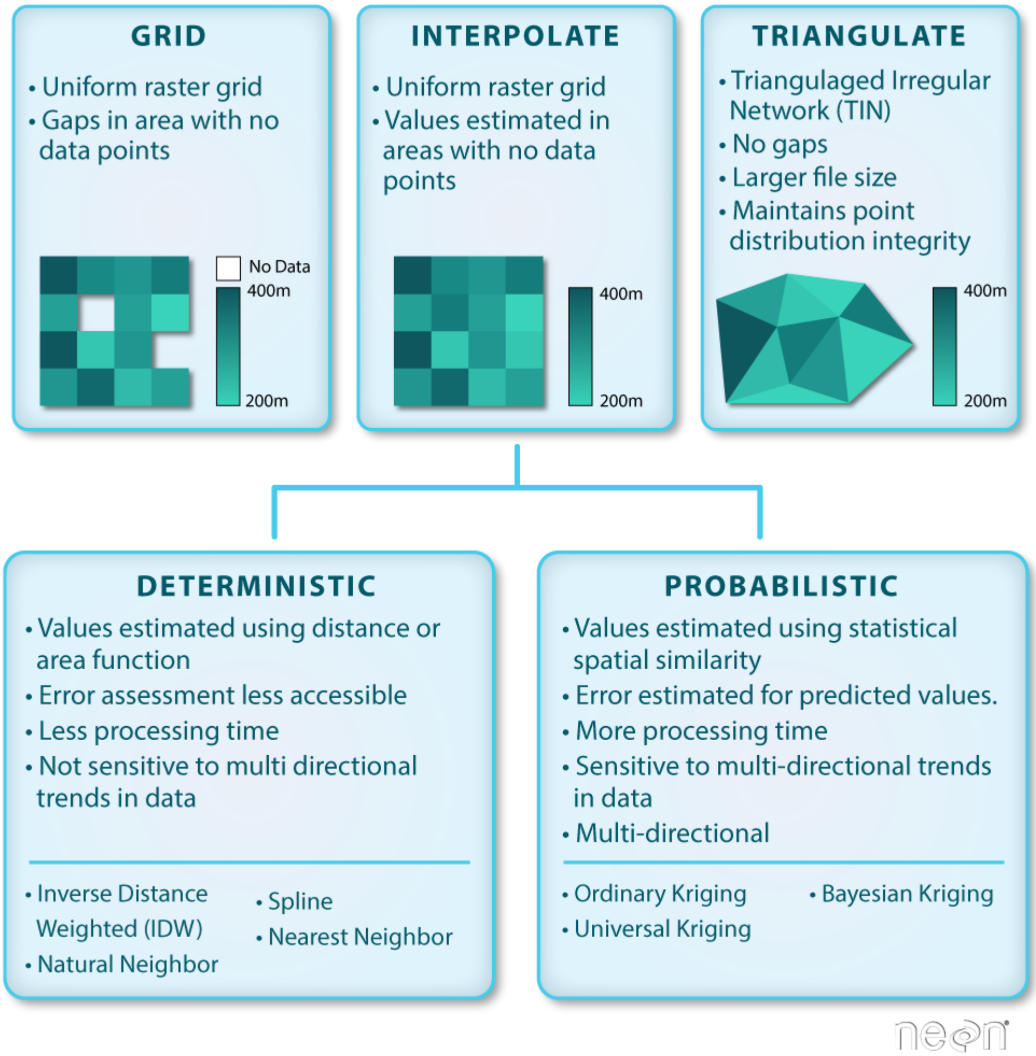


Figure 1. This context map breakdown serves as the precipice for the foundations of deterministic and probabilistic interpolation as the purview to be investigated in the report. (Neonscience.org)

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	NDAWN Average Air Temperature Data for All Stations from October 20th to November 18th	NDAWN Daily data for the month of October - November	Table	Tabular	NDAWN	Organized and rearrange columns in the table and created own table for XY Table to Point for extraction of tabular data.

2	Inverse Distance Weighted (IDW)	uses a method of interpolation that estimates cell values by averaging the values of sample data points in the neighborhood of each processing cell.	N/A, used on NDAWN Station points	Point	Point	None, Predefined tool in ArcGIS Pro
3	Kriging (Ordinary)	generates an estimated surface from a scattered set of points with z-values.	N/A, used on NDAWN Station points	Point	Point	None, Predefined tool in ArcGIS Pro
4	Kriging (Universal)	generates an estimated surface from a scattered set of points with z-values.	N/A, used on NDAWN Station points	Point	Point	None, Predefined tool in ArcGIS Pro
5	Spline	uses an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points.	N/A, used on NDAWN Station points	Point	Point	None, Predefined tool in ArcGIS Pro

Table 1. Inserted NDAWN Table utilized for study and predefined interpolation methods applied.

Input Data

The North Dakota Agricultural Weather Network averaged air temperature table for all stations is the preliminary data that is utilized for juxtaposing and contrasting interpolation methodologies. The table also contains ancillary data of minimum and maximum temperature statistics and standard deviations for all the stations for other analyses or inferences. The anemometers and thermometers utilized at each station to measure air and collect temperature data are collated by each day and then divided for the month for the table. The instrumentation/tools used in the field are paramount in relation to what is recorded in the table for veracity and geolocating the station within the API from NDAWN.

#	Title	Purpose in Analysis	Link to Source
1	NDAWN Average Air Temperature from October 20th to November 18th	CSV table for comparison and contrast on interpolation methods in ArcGIS Pro.	NDAWN

Table 2. The preliminary data was used to juxtapose and contrast statistical interpolation methods spatially and empirically.

Methods

The steps carried out for employing the North Dakota Agricultural Weather Network tabular data are as follows congruently in Figures 2 and 3 respectively. As depicted, (except for importing necessary modules for programming) the CSV file extracted from the NDAWN website is converted to XY points to be displayed in ArcGIS Pro, where then the points are utilized for interpolation. At least three interpolation methodologies need to be utilized, but for more robust comparison and optimization, this analysis is apt for four (additional kriging) for further comparison in literature analysis and inverse weight distance contrasts when using weather/temperature data.

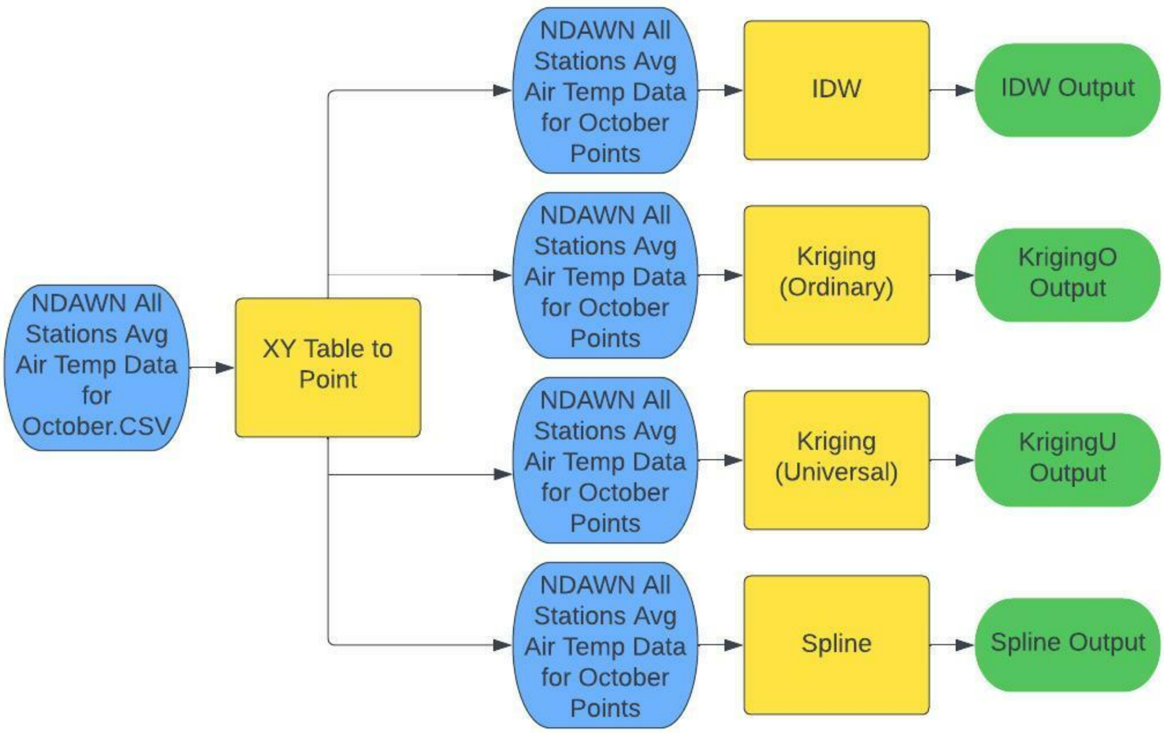


Figure 2. Data flow enumeration for implementation of NDAWN data table from XY Table to Point and interpolation methodologies employed.

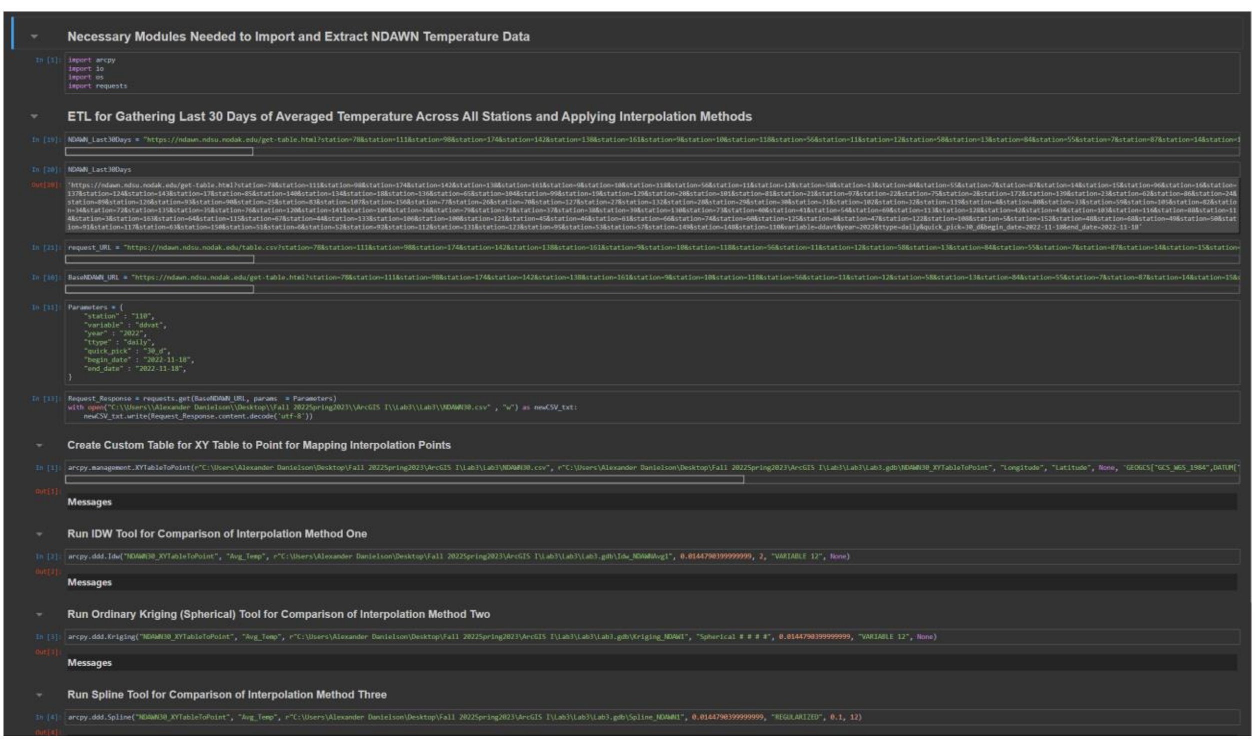
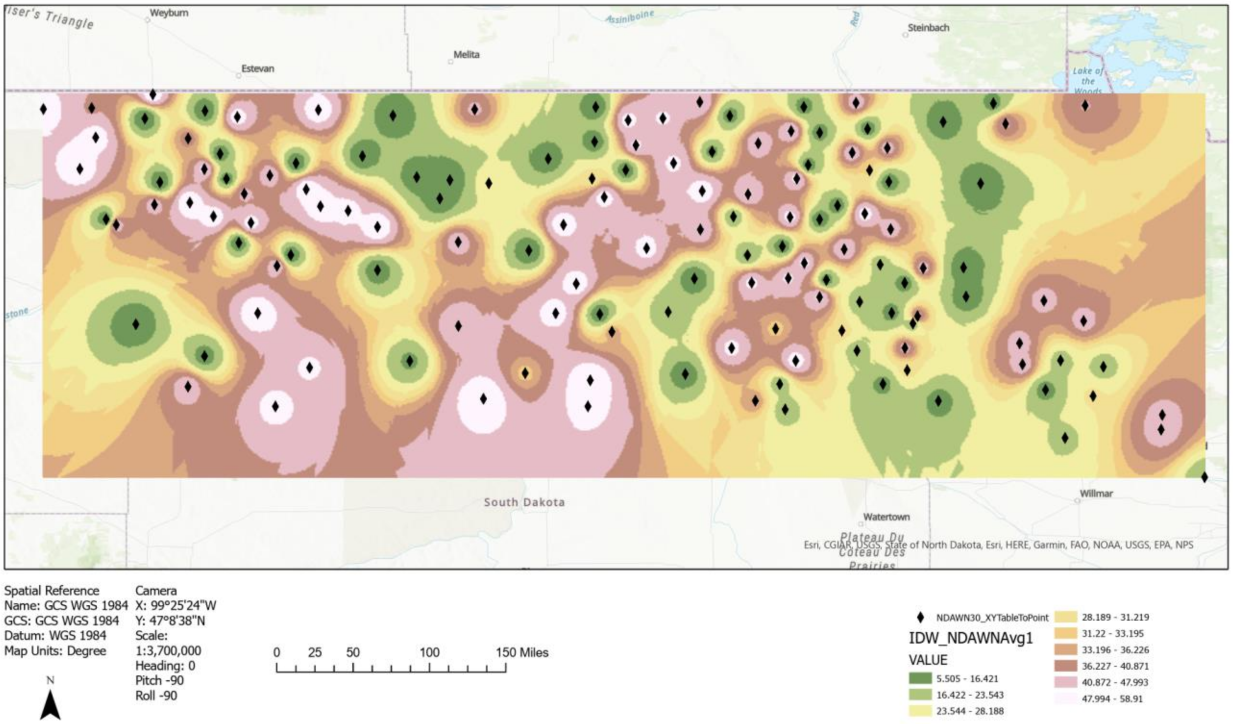
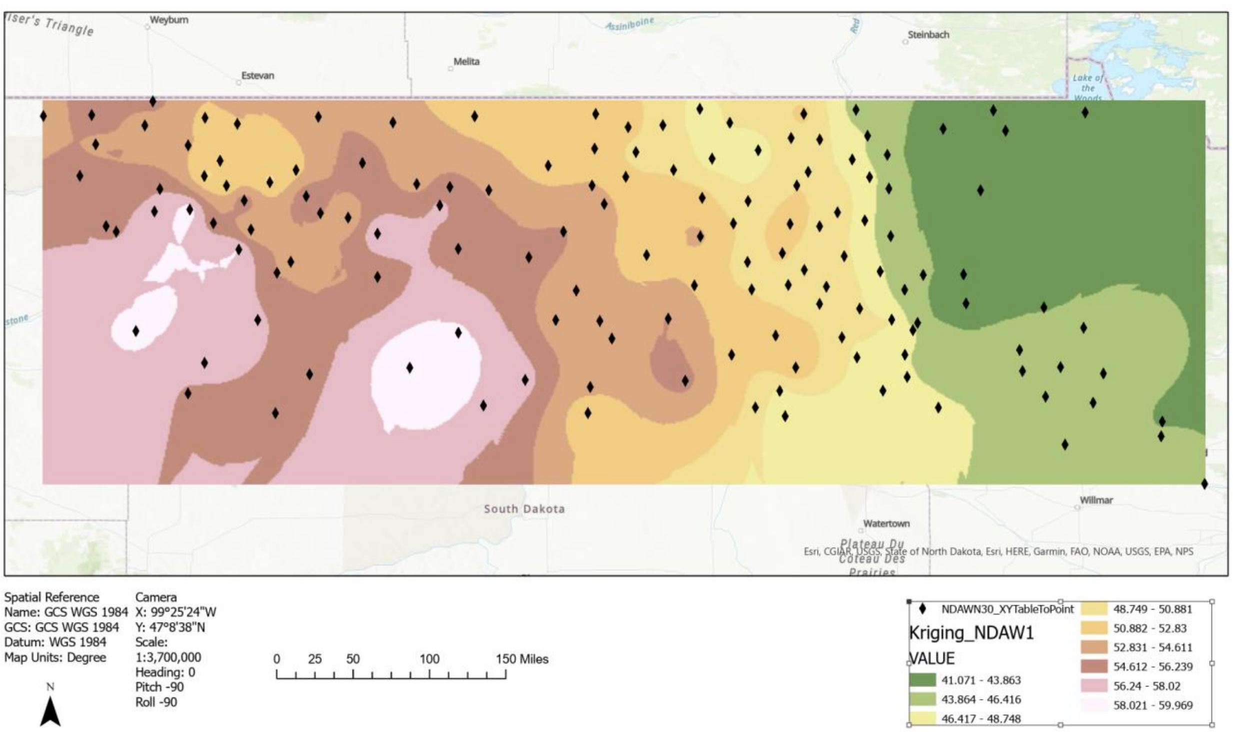


Figure 3. Programmatic enumeration of ETL and employing interpolation methodologies on weather stations from the NDAWN table data.

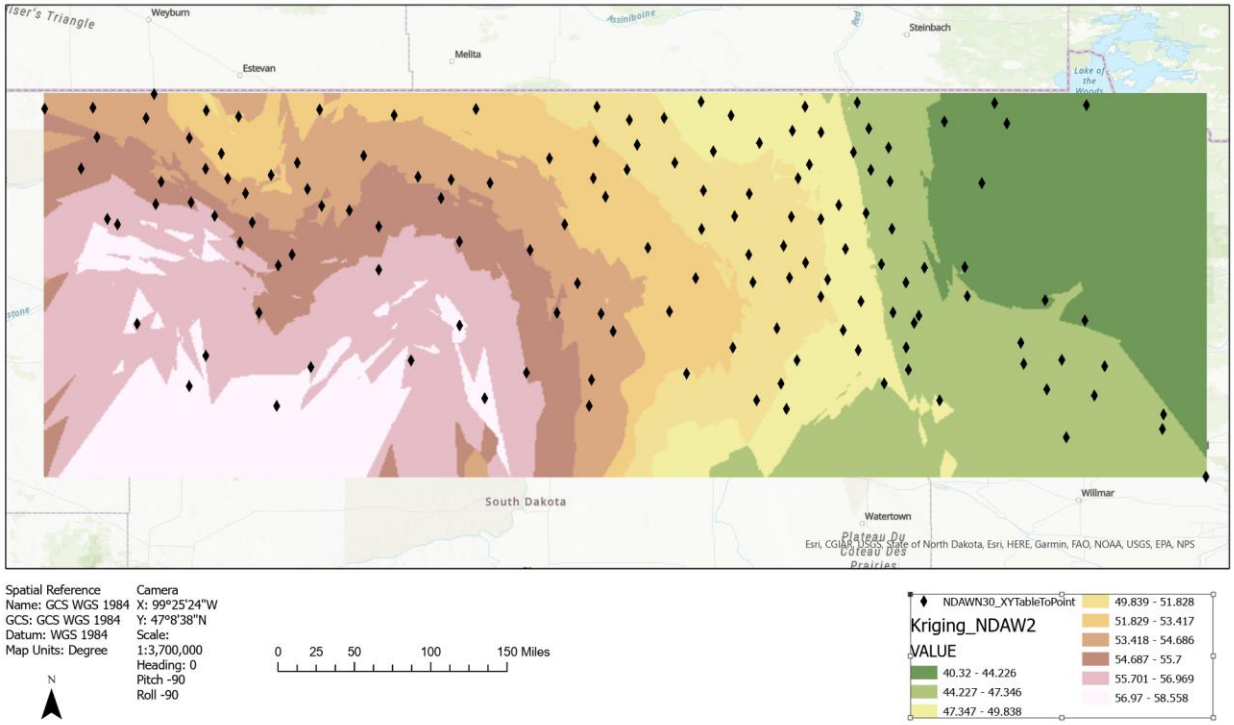
Results



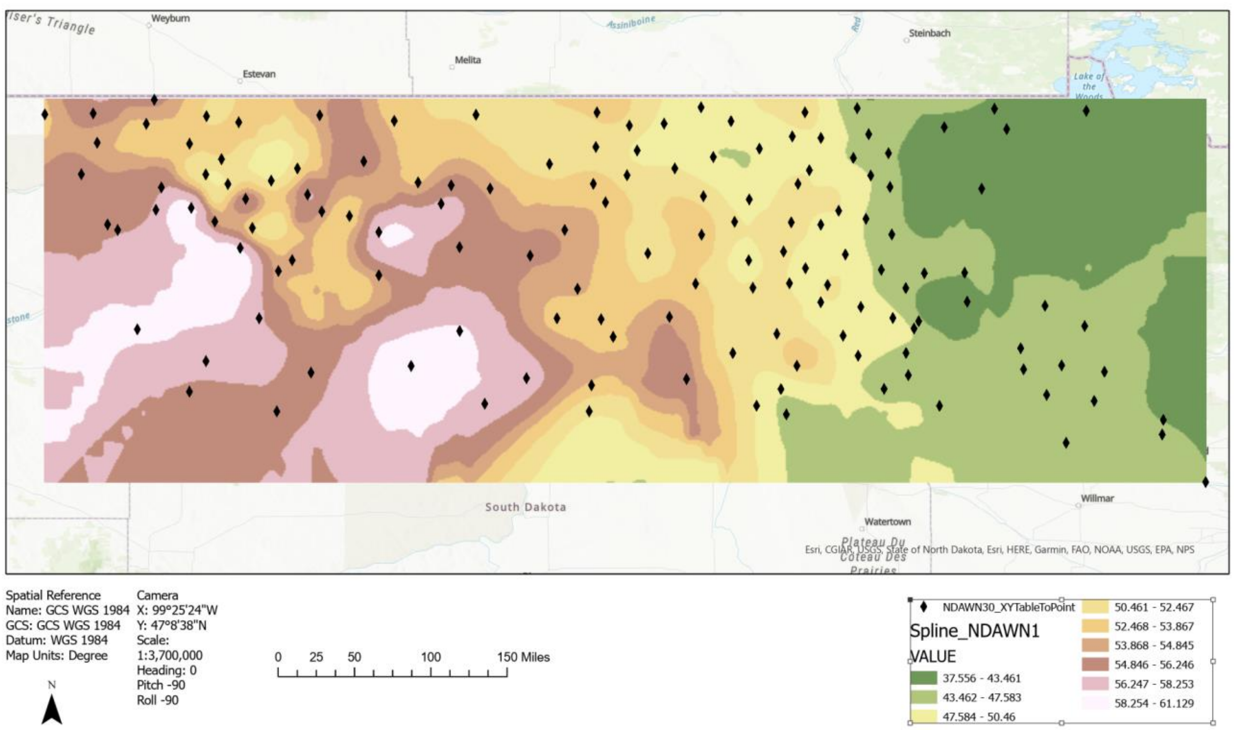
Map 1. Inverse Distance Weighted interpolation results for all NDAWN Stations.



Map 2. Kriging (Ordinary) interpolation results for all NDAWN Stations.



Map 3. Kriging (Universal) interpolation results for all NDAWN Stations.



Map 4. Spline interpolation results for all NDAWN Stations.

Results Verification

While the results are variable for each different interpolation methodology based on the station elevation and their respective location, each interpolation has different parameters and equations for estimating/calculating and assigning values. Based on Figure 1 for the overview of context mapping and partitioning the interpolation methods from deterministic to probabilistic interpolation; qualitatively IDW and Splines are not sensitive to error trends for accessibility, take less processing times (depending on data input) and use distance or area functions as parameters. Compared to Kriging statistics tre more prone to error based on predicted values, take more processing time, are sensitive to multi-directional trends, and values are estimated with statistical spatial similarity. (Neonscience.org)

In using each toolset for the NDAWN average temperature for October, there are stark contrasts in the results, these have to do with the continuity of the data being the temperature data being collected at different times and each statistical calculation being applied to the stations across spatio-temporal space. Each following paragraph gives the justification for the methodology used and verification for why it is applicable or unjustified for the analysis from ESRI documentation from the resulting Maps in the previous section.

The Inverse Distance Weighted map 1 result is an interpolation methodology that estimates cell values by averaging the values of sample data points (NDAWN) in the neighborhood of each processing cell. The closer a point is to the center of the cell being estimated, the more weight, it has on the averaging process. As depicted, we can denote by the legend and certain placements of points from raster cells' sphere of influence that some points based on elevation had more dominance over others compared to their distance and vice versa. Since there are no discrepancies in temperature data and elevation the IDW method is prudent for predicting these variables and reducing error and each input is the same.

Ordinary and Universal Kriging results in maps 2 and 3 are defined as geostatistical procedures that generate an estimated surface from a scattered set of points with z-values. A thorough investigation of the spatial behavior of the phenomenon represented by the z-values should be done before selecting the best estimation method for generating output surface. The mapping results are implicitly depicting jagged and almost zig-zagged waves for the stations (more prevalent for universal kriging), but this is the result of predefined values. As Kriging makes many assumptions based on z-values for normally distributed data, the NDAWN data isn't normally distributed so this explains the jaggy surface interpolation. Kriging is best for the probability of datasets than determined data, as the temperature is an interval data and kriging would be more applicable to ratio data. This explains variations for elevation as, the algorithm for ordinary kriging goes from the bottom left to the top right of data points with intersecting lines from most rigid to smoother (as seen). Similarly, universal kriging likewise, except using polynomial equations and going over a curve.

Finally, the Spline results in map 4 is defined as estimating values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes through exactly through the input points. The spline method can be best compared to that of IDW, it's properties aren't equivalent to that of IDW (in terms of mapping comparison) as it uses a polynomial function for each cell juxtaposed to the IDW for the cell comparison. Spline is best suited for the data types that have irregularly fitted data.

Discussion and Conclusion

In verifying the results of each interpolation methodology and how each varies with the NDAWN average weather temperature stations, further supplementing evidence is needed to show which method is the most "recommended" for temperature. The following provides a definitive amount of evidence and support and based on results conducted proved which methodology is advantageous over the other for temperature data.

Based on Hosfstra et. al study in "Comparison of six methods for the interpolation of daily, European climate data, there is notable "best" method, but in the case of coming down to higher elevation and more dense stations in the vicinities, ordinary kriging performed best, but not for higher elevation compared Inversed-Distance Squared or Weighted. (Hosfstra et. al, pg. 3) Likewise, methods of Natural neighbor interpolation (NNI) which chooses its neighbors based on geometry algorithmically and Angular Distance weighting (ADW) applied towards monthly climatic data, the first iteration (two types) that contributes to a grid-point estimate using constant search radius of 250 km for precipitation and 500 km for temperature, with the distance components of the weights decaying to zero at the search radius. (Hosfstra et. pg. 3)

Also, a thin splines (TPS) interpolation was utilized, and the author gives justification that this is like Kriging, but covariances are in reference to measuring of the directional relationship between two random variables and not the spread of data. This then governs the inference to the ideation that splines are best suited for data that are heterogenous in scope and not uniform. Depending on which Spline methodology, regular spline tool in ArcGIS Pro seems adequate for temperature data for interpolating.

In another paper, Kusuma et. al stated that IDW is the best interpolation methodology in terms of sea surface temperature and weather currents and gives empirical evidence based on RSME charts and tables:

First Author (Last name, Initial first name et al.) / Journal of Fisheries and Marine Science XX(2018) XX-XX

Table 2. Assessment Test (December 2015 – May 2016)

Interpolation Methods	Assessment Method						
	SP	Min	Max	Mean	RMSE	Pearson	STDev
December 2015							
IDW	5	28.99	31.16	30.04	0.62	0.60	-0.05
Kriging	5	29.00	31.16	30.01	0.62	0.58	-0.07
NNI	5	29.00	31.15	30.00	0.63	0.57	-0.08
Spline	5	28.99	31.16	30.00	0.63	0.57	-0.08
Aqua MODIS		28.54	31.82	30.01			
January 2016							
IDW	5	29.60	30.55	30.22	0.64	0.13	-0.25
Kriging	5	29.59	30.54	30.20	0.66	0.11	-0.24
NNI	4	29.60	30.54	30.18	0.68	0.09	-0.23
Spline	5	29.58	30.55	30.17	0.70	0.06	-0.20
Aqua MODIS		29.60	31.52	30.70			
February 2016							
IDW	5	29.43	30.78	30.32	0.80	0.34	-0.11
Kriging	3	29.42	30.78	30.29	0.83	0.31	-0.11
NNI	5	29.42	30.78	30.27	0.85	0.28	-0.11
Spline	3	29.42	30.78	30.27	0.85	0.28	-0.10
Aqua MODIS		30.09	31.86	30.99			
March 2016							
IDW	5	30.11	31.68	30.95	0.87	0.26	-0.17
Kriging	4	30.11	31.67	30.92	0.91	0.21	-0.17
NNI	4	30.10	31.67	30.89	0.94	0.15	-0.16
Spline	4	30.10	31.67	30.89	0.95	0.14	-0.15
Aqua MODIS		30.24	33.07	31.66			
April 2016							
IDW	5	29.76	31.54	30.70	0.44	0.58	-0.02
Kriging	4	29.75	31.54	30.67	0.44	0.55	-0.03
NNI	4	29.75	31.54	30.65	0.45	0.52	-0.03
Spline	4	29.75	31.55	30.65	0.45	0.52	-0.03
Aqua MODIS		29.57	31.49	30.54			
May 2016							
IDW	5	29.04	30.19	29.70	0.40	0.08	-0.03
Kriging	5	29.12	30.10	29.72	0.38	0.10	-0.07
NNI	5	29.04	30.19	29.68	0.40	0.06	-0.03
Spline	3	29.03	30.20	29.67	0.41	0.04	-0.02
Aqua MODIS		28.96	30.35	29.63			

Table 3. Assessment Test (June - November 2016)

Interpolation Methods	Assessment Method						
	SP	**Min**	**Max**	**Mean**	**RMSE**	**Pearson**	**STDev**
June 2016							
IDW	4	27.95	30.08	28.81	0.85	-0.30	-0.01
Kriging	4	27.94	30.07	28.78	0.88	-0.30	0.01
NNI	4	27.95	30.07	28.81	0.86	-0.28	-0.01
Spline	4	27.94	30.08	28.78	0.89	-0.29	0.02
Aqua MODIS		28.58	30.42	29.35			
July 2016							
IDW	2	27.04	29.64	28.02	0.91	-0.04	0.01
Kriging	2	27.03	29.62	27.99	0.94	-0.04	0.03
NNI	2	27.03	29.62	28.02	0.91	-0.05	0.01
Spline	3	27.03	29.64	28.00	0.94	-0.04	0.03
Aqua MODIS		27.60	29.94	28.55			
August 2016							
IDW	2	26.51	29.03	27.39	0.77	0.16	0.00
Kriging	2	26.50	29.01	27.37	0.80	0.13	0.01
NNI	2	26.50	29.01	27.40	0.77	0.16	0.00
Spline	2	26.49	29.03	27.36	0.82	0.12	0.03
Aqua MODIS		27.44	29.37	27.96			
September 2016							
IDW	2	26.46	28.45	27.55	1.20	-0.29	-0.34
Kriging	2	26.45	28.44	27.52	1.24	-0.37	-0.33
NNI	3	26.45	28.43	27.54	1.20	-0.26	-0.34
Spline	2	26.43	28.45	27.49	1.28	-0.43	-0.30
Aqua MODIS		27.38	30.01	28.39			
October 2016							
IDW	2	27.40	29.56	28.58	1.20	0.37	-0.24
Kriging	3	27.39	29.55	28.55	1.24	0.30	-0.25
NNI	2	27.40	29.54	28.57	1.22	0.33	-0.24
Spline	2	27.38	29.55	28.52	1.29	0.20	-0.24
Aqua MODIS		28.39	31.84	29.59			
November 2016							
IDW	2	28.37	30.31	29.62	1.32	0.64	-0.43
Kriging	4	28.36	30.31	29.59	1.36	0.61	-0.44
NNI	3	28.38	30.31	29.60	1.35	0.61	-0.44
Spline	3	28.36	30.31	29.57	1.39	0.57	-0.43
Aqua MODIS		28.60	32.69	30.71			

Figure 4. Kusuma et al. Giving a ranking for best performance (5 being the best) and RSME for how interpolation methods adapt from 2015 to 2016 in the Indian Ocean.

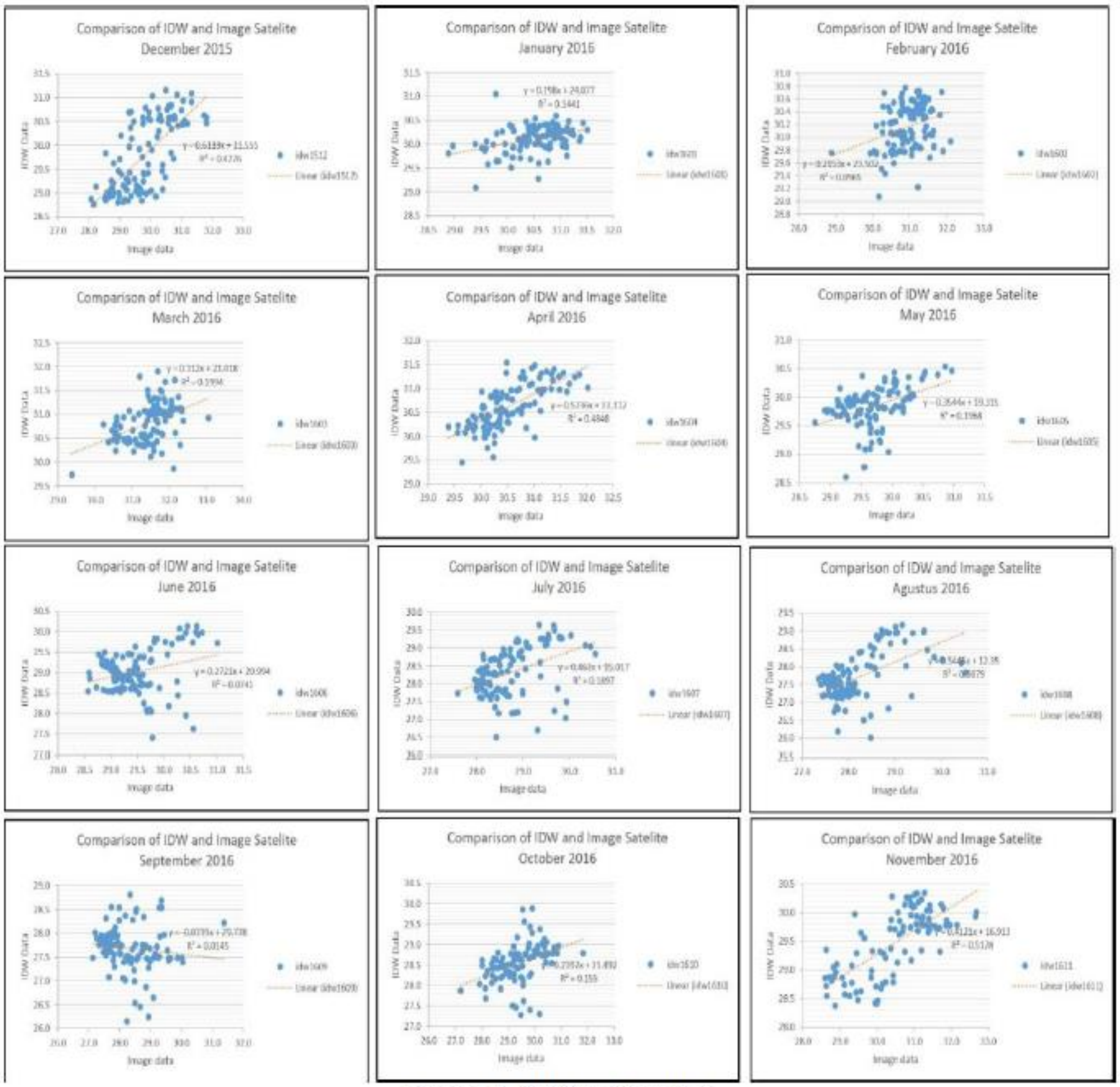


Figure 2. IDW scatter plot

Figure 5. Kusuma et al. Scatterplots statistics depicting IDW methodology for each month of sea surface temperature.

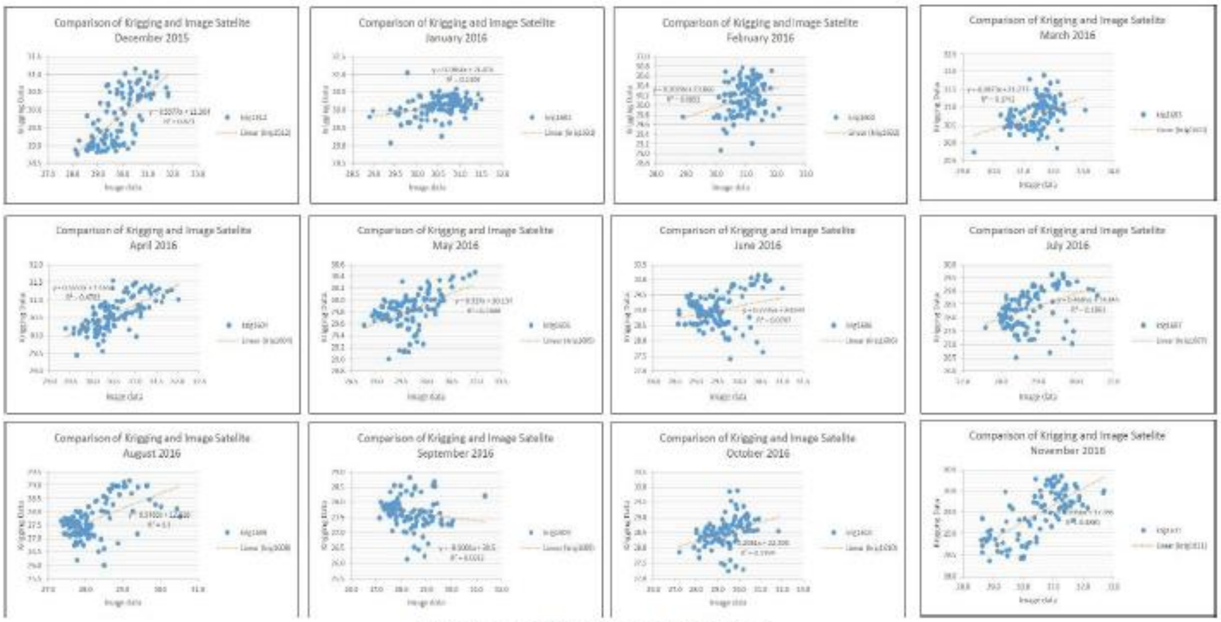


Figure 4. Kriging scatter plot

Figure 6. Kusuma et al. Kusuma et al. Scatterplots statistics depicting Kriging methodology for each month of sea surface temperature.

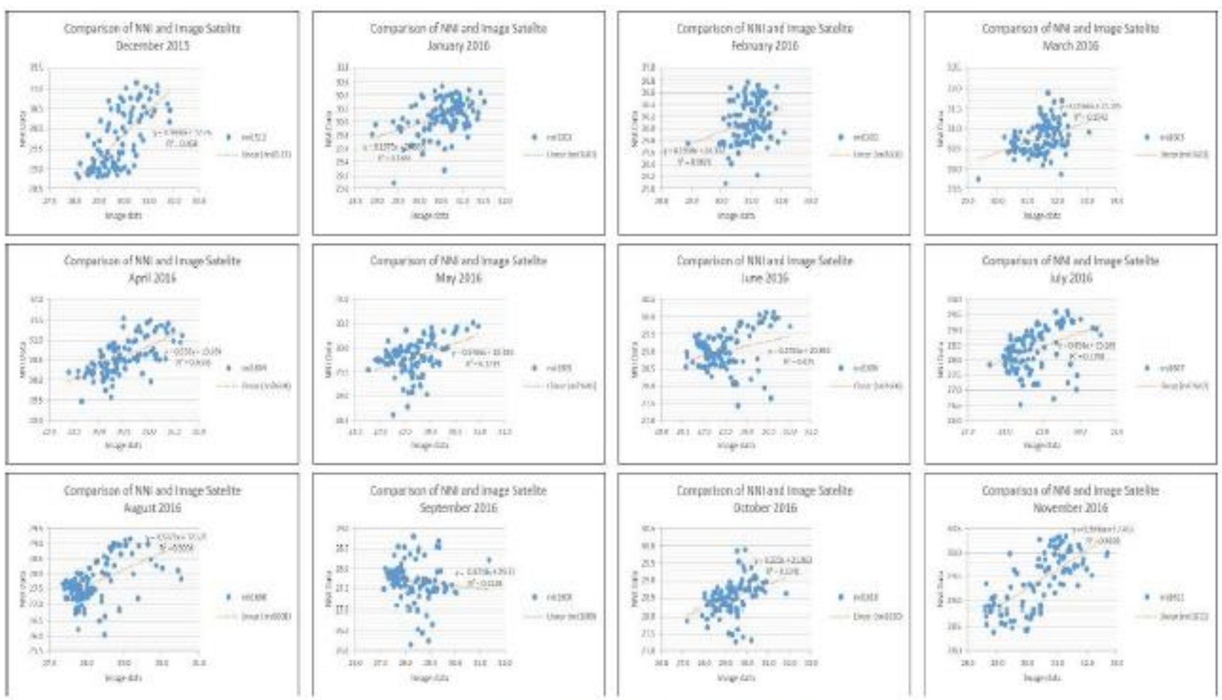


Figure 6. NNI Scatter plot

Figure 7. Kusuma et al. Kusuma et al. Scatterplots statistics depicting NNI methodology for each month of sea surface temperature.

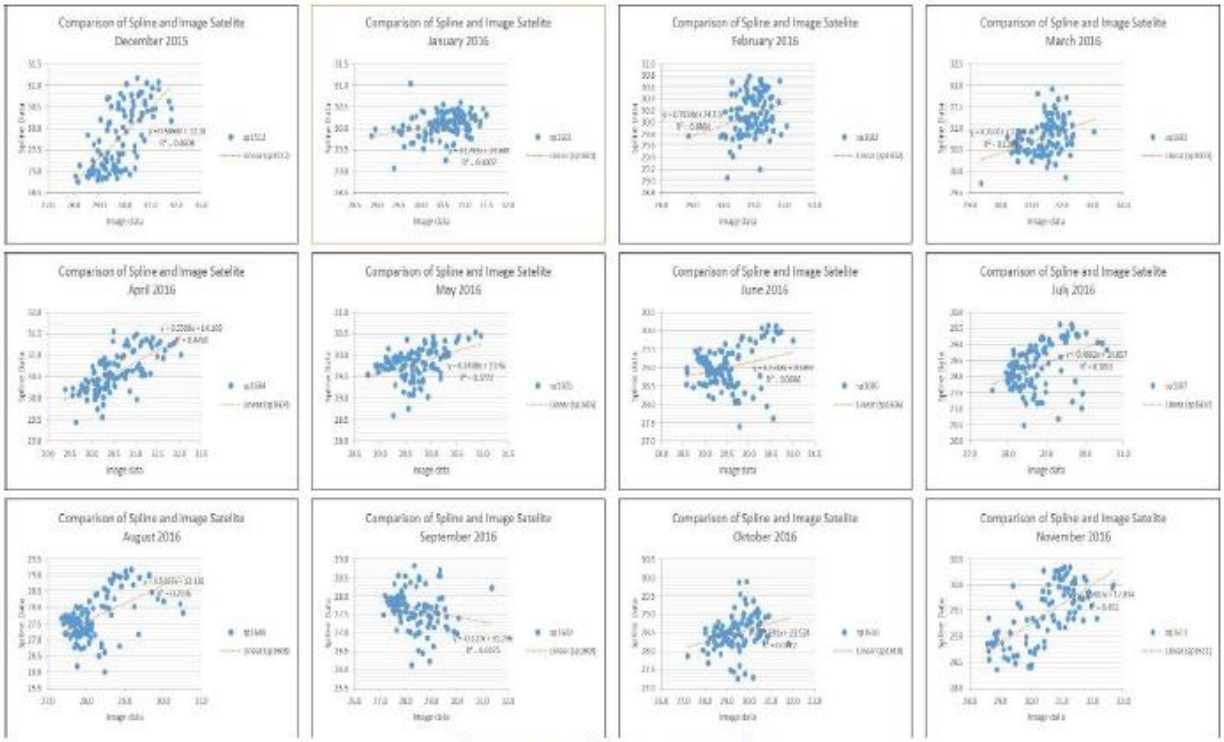


Figure 8. Spline scatter plot

Figure 8. Kusuma et al Kusuma et al. Scatterplots statistics depicting Spline methodology for each month of sea surface temperature.

References

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

Kusuma, D., Murdimanto, A., Sukresno, B., & Jatisworo, D. (2018). Comparison of interpolation methods for sea surface temperature data D. JFMR-Journal of Fisheries and Marine Research, 2(2), 103–115. https://doi.org/10.21776/ub.jfmr.2018.002.02.7

Classification trees of the interpolation methods offered in Geostatistical Analyst—ArcMap. (n.d.). Arcgis.com. Retrieved November 14, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

Comparing interpolation methods. (n.d.). Arcgis.com. Retrieved November 13, 2022, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/comparing-interpolation-methods.htm

(N.d.). Neonscience.org. Retrieved November 13, 2022, from https://www.neonscience.org/resources/learning-hub/tutorials/spatial-interpolation-basics

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	**23**
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 the verification is clearly stated **(5 points)**.	20	**20**
		100	**99**