Assignment 4: Interpolation

Geog 418/518

March 24, 2022

Introduction

Taking measurements across large areas can be economically and temporally challenging. The resources needed to get accurate measurement data over province wide scale is no easy task. However, using spatial interpolation methods within GIS systems can aid in filling gaps between pre-existing measurement sites can provide an efficient and cost-effective mode for estimating a measured variable (Mitas, 1999).

This report will be introducing and comparing a various spatial interpolation methods and determining the strengths and weaknesses of each technique. Other similar method analysis' have been performed in the past comparing different interpolation methods such as (Naoum, 2004), who estimated rainfall variability over space via Thiessen Polygons, IDW, Kriging and a few other techniques. For their application however their results concluded that estimation of rainfall variation was ineffective using those interpolation styles. Other studies have been performed interpolating min and max temperature using kriging and distance weighting methods complimented with elevation data to determine the relative accuracy of interpolation methods, one within BC the other across the Northeastern United States respectively, (DeGaetano & Belcher, 2007; Stahl et al., 2006). Though all three of these studies analyze the effectiveness of different methods in a specific way, none of them combine to cover the scope of methods, specific variable and over regions which will be monitored in this analysis.

For this method analysis spatial interpolation estimation for the maximum temperature across the province of British Columbia in May of 2017 will be performed using a few of the previously mentioned interpolation techniques. Firstly, Thiessen Polygons, which is a discrete method of estimating areas of influence. Inverse Distance weighting, which bases its values upon a linear weighted interaction with existing observations (Simpson, 2014). This method is the first to use parameterization within the interpolation to achieve a better fit, which in this case determines the intensity of a measurements weight over distance. Finally, kriging, which estimates variance over some distance within a data set, featuring the largest amount of parameterization in fitting data to a generated variogram which then applies weights by a process of matric algebra relating to the known and unknown semivariances of the data / interpolated surface.

The objective for this analysis is to determine the strengths and weaknesses of each style of spatial interpolation as well as generating a cohesive interpolation of maximum temperature of British Columbia in May of 2017.

Methods

Study Area & Data

The study area for this analysis is British Columbia Canada. The westernmost coastal based province in Canada, which is known for its presence of islands along its coastline as well as the versatile topography inland including the coastal and Rocky Mountains.

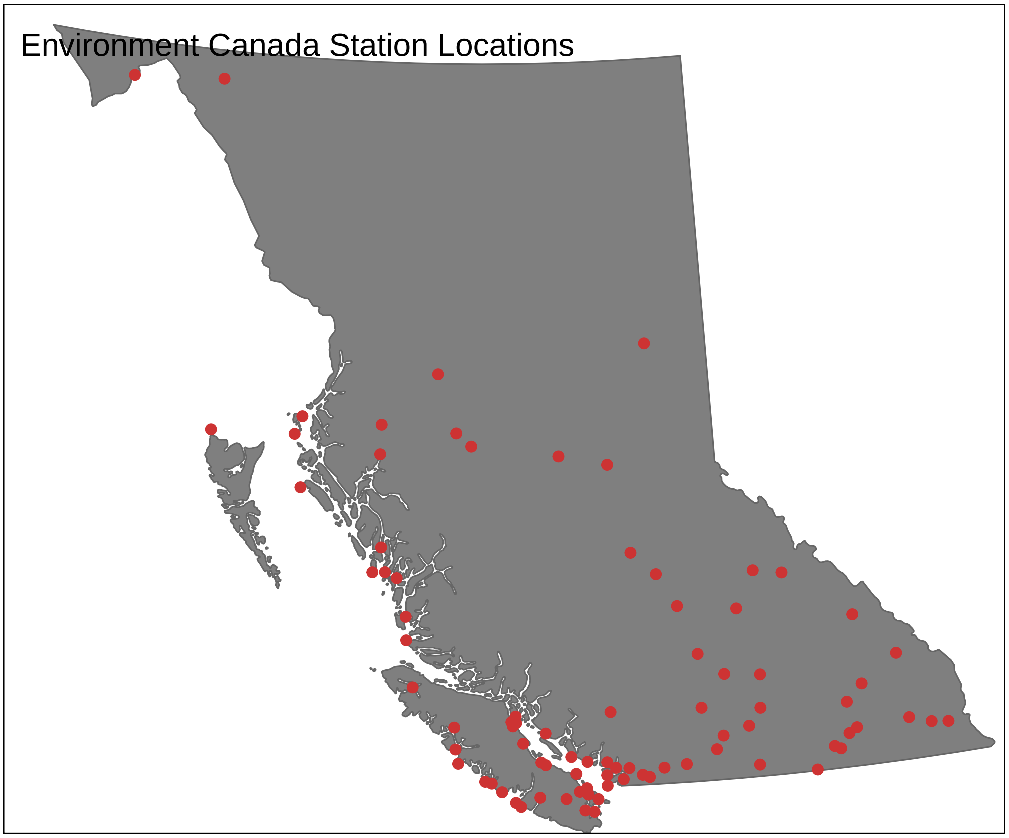

Figure 0. Study Area and original data

The data retrieved for this examination was taken from the Pacific Climate Impacts Consortium's open data portal, from which max temperature data from Environment Canada weather stations was collected. Though it should also be stated that the placement of EC weather stations is NOT uniform throughout BC making some regions far less likely to get an accurate representation when compared to the greater Vancouver area as well as the Southern Island which featured a substantially larger amount of stations.

Spatial Interpolation Styles / Methods

Thiessen Polygons

The first method of spatial interpolation: Thiessen polygons, arguably the most basic and straight forward of all the analysis' that will be performed. This method takes the supplied spatial data the respective attribute and generates a tessellated surface where each measurement is given its own polygon to represent data/ observation.

The application for this style of spatial interpolation would is best in applications where regions of influence are predetermined. More specifically the tessellation generated to fit the interpolating surface is arbitrary and is designed to fit each data point with its own polygon with minimal contradictions. The resulting surfaces are often irregularly shaped and not uniform in their size. Possibly the greatest flaw in this style of interpolation is that it does not feature a gradient. The generated tessellation/ polygons are assigned their specified measured value and there is no further estimation beyond the creation of the polygons. In instances such as this sample design or the original location of observations has a critical influence on the outcome of the interpolation. It would be ill advisable to use this method if the sample is inconsistent and the data set features high variability.

With that considered, the generated tessellation does remove some of the potential issues with human error in decision making for accurate results which will be seen in future methods. As well edge effects / edge cases are handled appropriately by this model as well due to the functionality of the tessellation. Finally it should also be said that this method is the least computationally intensive technique of all the methods today, making it more time and resource efficient than the other methods, with sacrifices being made to accuracies.

IDW Inverse Distance Weighting

The next method of interpolation is IDW or Inverse Distance Weighting, which takes into consideration the distance between locations and directly weights how much impact a measurement point has on the interpolated surface. Generally, the farther the distance an estimated surface is from the nearest measurement the less impact the measurement has on the estimated surface. This relation is of course Toblers 1st law of geography, objects which are closer are more similar. As mentioned prior this method has parameterization which is adjusted by the user that determines how quickly the intensity of a measurements influence decreases over distance. The IDP value is then adjusted to 1. Match the interpolated surface to the original data sets measurements and 2. Ensuring the root mean square error is an acceptable value.

$$\hat{Z}_i = \frac{\sum_{i=1}^{n} Z_j/d_{ij}^p}{\sum_{i=1}^{n} 1/d_{ij}^p}$$

[1] *where Z is the measured variable, and d is distance*

To validate the selected IDP value, a leave one out method is performed on the interpolated surface wherein each point / station is removed to determine the effectiveness of the interpolation. This jackknife resampling method was also performed in order to process all the data in a another leave one out method, retrying the entire interpolation n times, n being the total sample size. All iterations of the Jackknifed IDW are comparing the accuracy of the interpolated surface with each of the missing station to determine the accuracy of the interpolation and validity of the IDP value.

In general for this style of interpolation is more complex and effective when compared to Thiessen polygons. it would be wise to have a greater amount of measurements over space the greater the variability within the data set. i.e. For this application weather and temperature specifically remains relatively consistent over distances in prairie locations in contrast with mountainous regions with substantial climactic variations over smaller distances.

Kriging

The last interpolation technique used is ordinary kriging, which is somewhat similar to IDW in that it is assigning a weight to a total of values. However, it is being optimized to set weights such that the surrounding interpolated surfaces are fitting the generated semivariogram model. However, there are assumptions made about the data to maximize this methods' effectiveness. These assumptions are as follows: 1. There is a consistent average with no underlain trend in the data; 2. The variation in the data is the same in all directions away from measurements, the surface is isotropic; 3. The data is normally distributed and is not too variable that simple characteristics of the data can be defined; 4. These assumptions together allow the creation of the appropriate semivariogram which applies to the entire study area, due to consistency and trendlessensss in the data (O'Sullivan & Unwin, 2003).

The semivariogram created is a measure of the variance of an attribute over distance, however the important part to this is not the difference over distance, but the differences in variation over distance. More elaborately we are to aggregating the measured spatial variance at a specific distance X grouping that data such that spatial variances are then passed into function [2] where yh is the calculated spatial variance at distance X.

$$Y(h) = \frac{1}{2|N(h)} \sum_{N(h)} (z_i - z_j)^2$$

[2]

Once the semivariogram is created a semivariogram model can be selected to best fit the calculated semi variance profiles. When fitting a semivariogram model there are a variety of types to select from namely spherical, exponential, or gaussian though there are many others to choose from. See figure xx for comparison of variogram models.

$$\gamma(h) = c_0 + c(\frac{h}{a})$$

where $h > 0$ [3]

With the model and variogram prepped there is more parameter adjustments to be made to best fit the data, specifically nugget, range and sill. These variables represent the variance at zero distance(nugget), the distance at which the variance plateaus (range), and

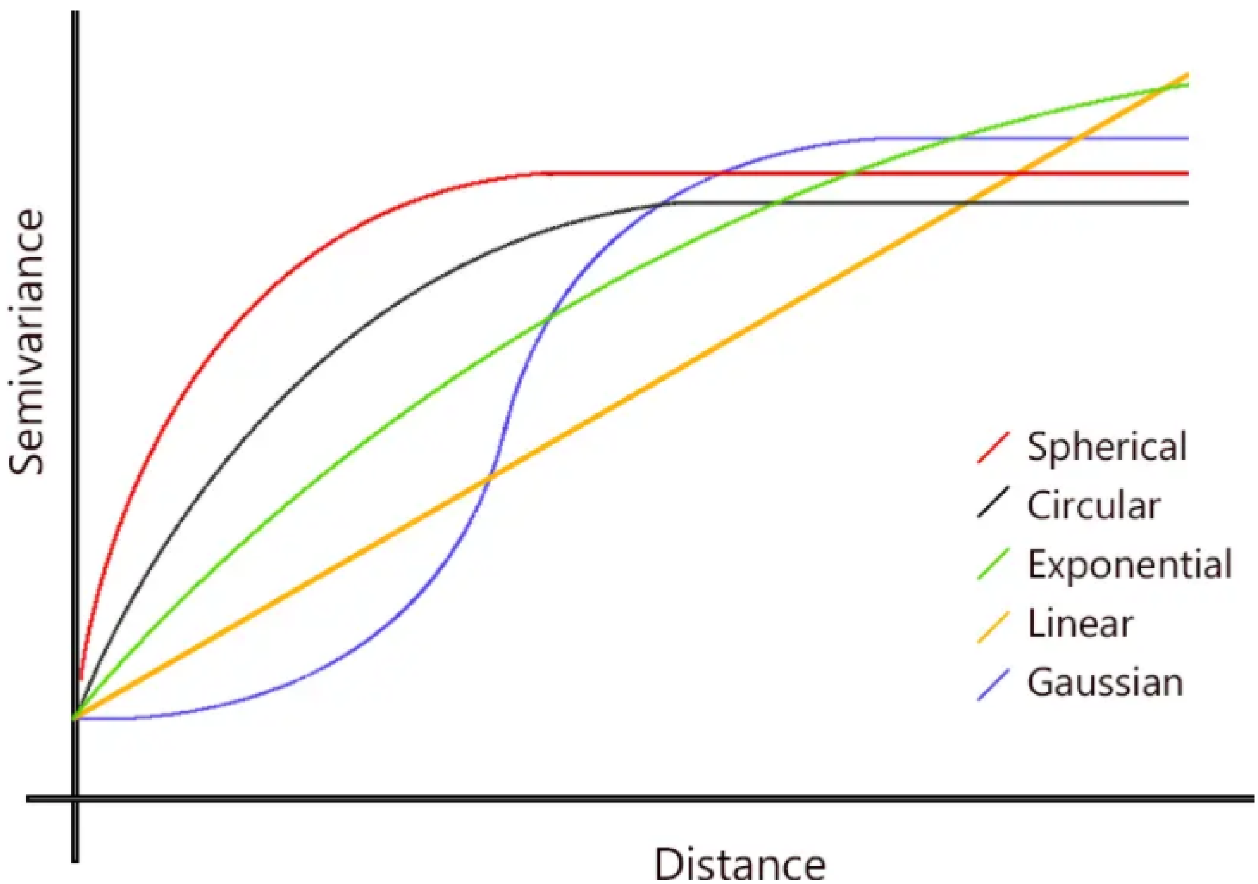

Figure 1: Model Types

the magnitude at which the variance plateaus (sill) (O'Sullivan & Unwin, 2003).

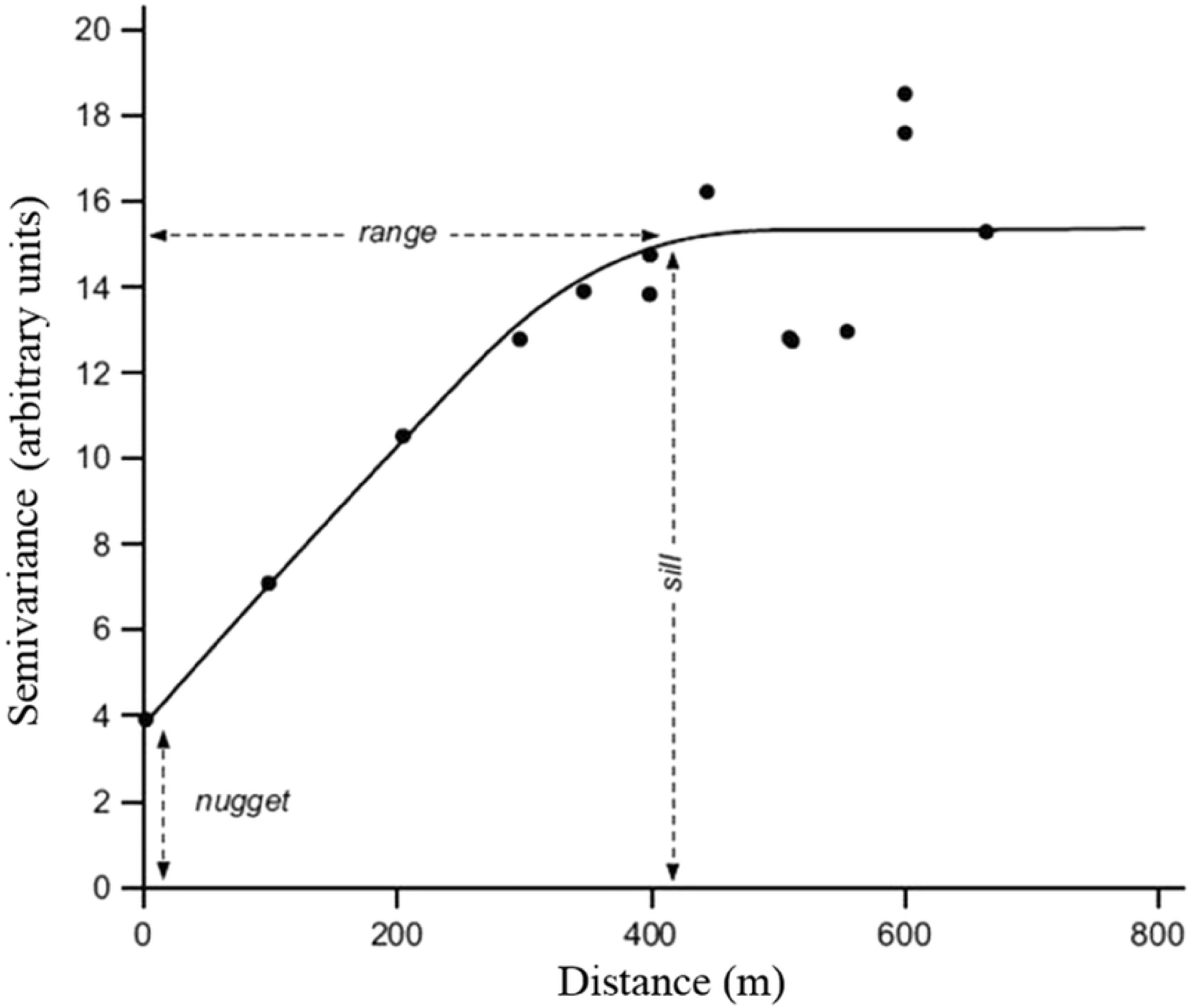

Figure 2: Variogram Parameters

Now weights can be applied to the unknown locations using formula [4] where w equates to weights applied to the known locations and z represents the values at the sampled locations.

$$\hat{Z}_i = w_1 + z_1 + w_2 + z_2 + ...w_n + z_n = \sum_{j=1}^{n} w_j Z_j$$

[4]

Then the computationally intensive matric algebraic calculations which a lots the weights applied to unknown locations using the known semivariance and weights from the previous formulation.

Results

Thiessen Polygons Results

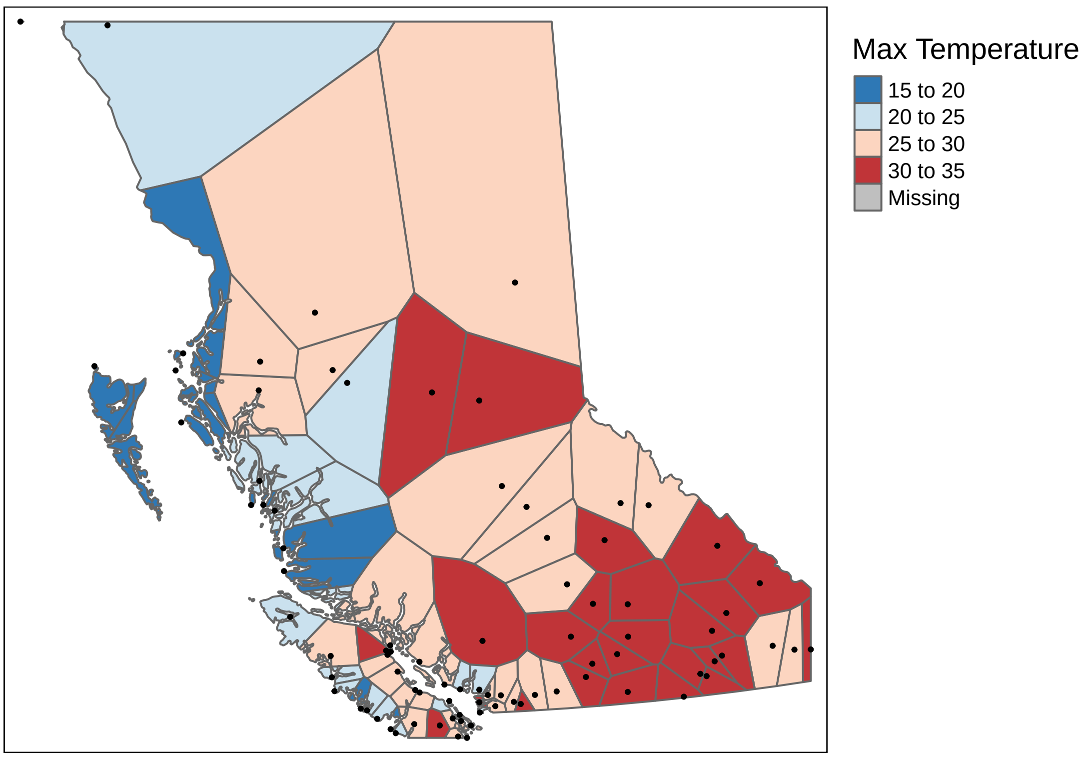

Figure 3. Thiessen Polygons Map for Maximum Temperature May 2017

In figure 3 regarding temperature it can be seen that many polygons which are not sharing vertices with the Pacific Ocean are warmer, as well the lower interior of BC is noticeably warmer than the rest of the province. Regarding the effectiveness of Thiessen polygons in interpolation, we can see the discrete data and lack of sample locations leads to very large and clunky outputs, with little to no gradient for steady interpretation.

IDW Results

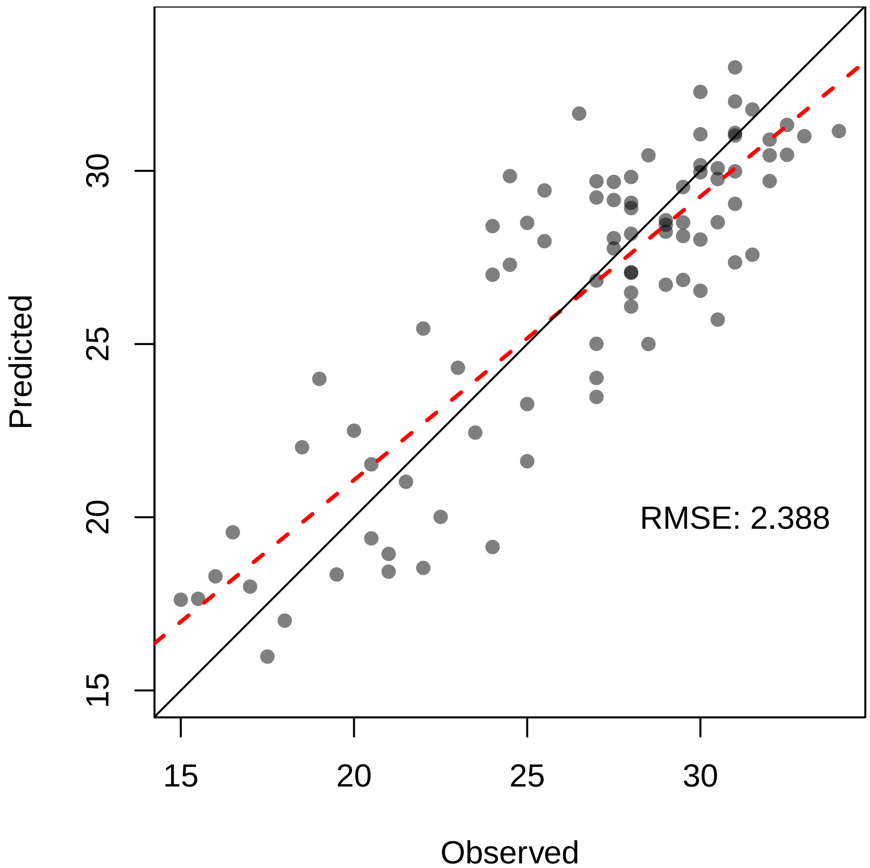

Figure 4. IDW Validation Plot with an IDP value of 5

As seen in figure 4s validation plot, the IDP value of 5 suits this data set rather well and will be used for this analysis.

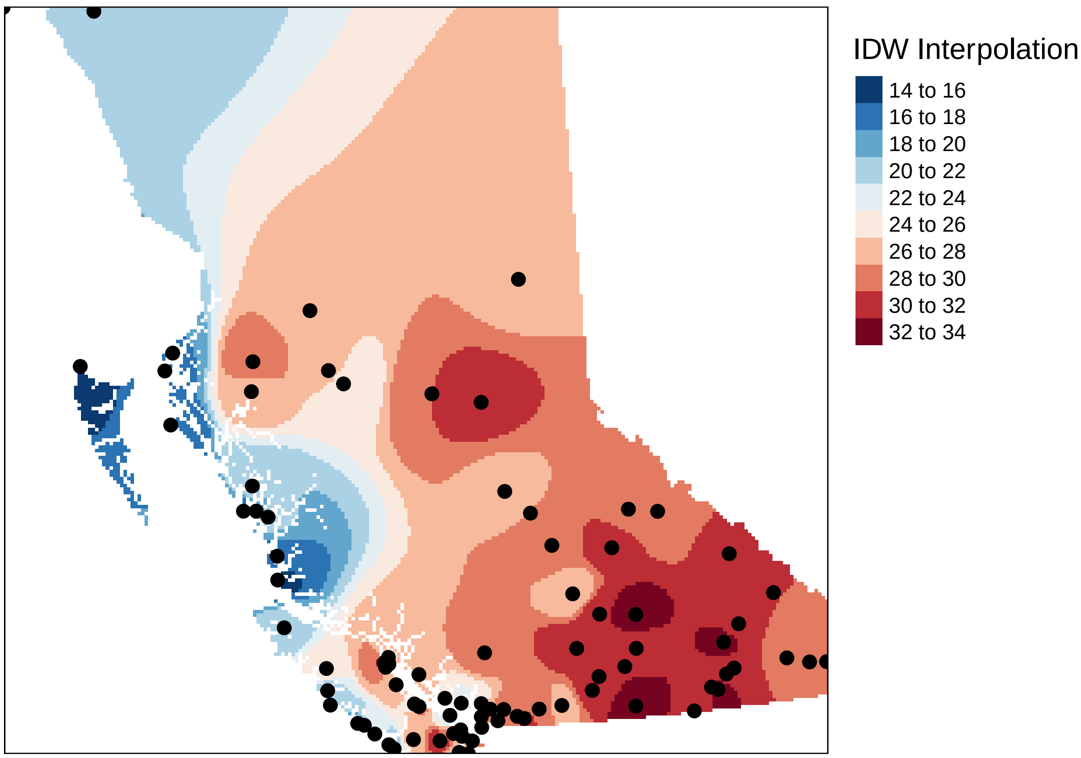

Figure. 5: Inverse Distance Weighted max temperature BC, May 2017

Comparing the results between figure 5 and figure 3 Thiessen polygon interpolation we can see that IDW have generated a much smoother surface, with a gradient which can be better interpreted.

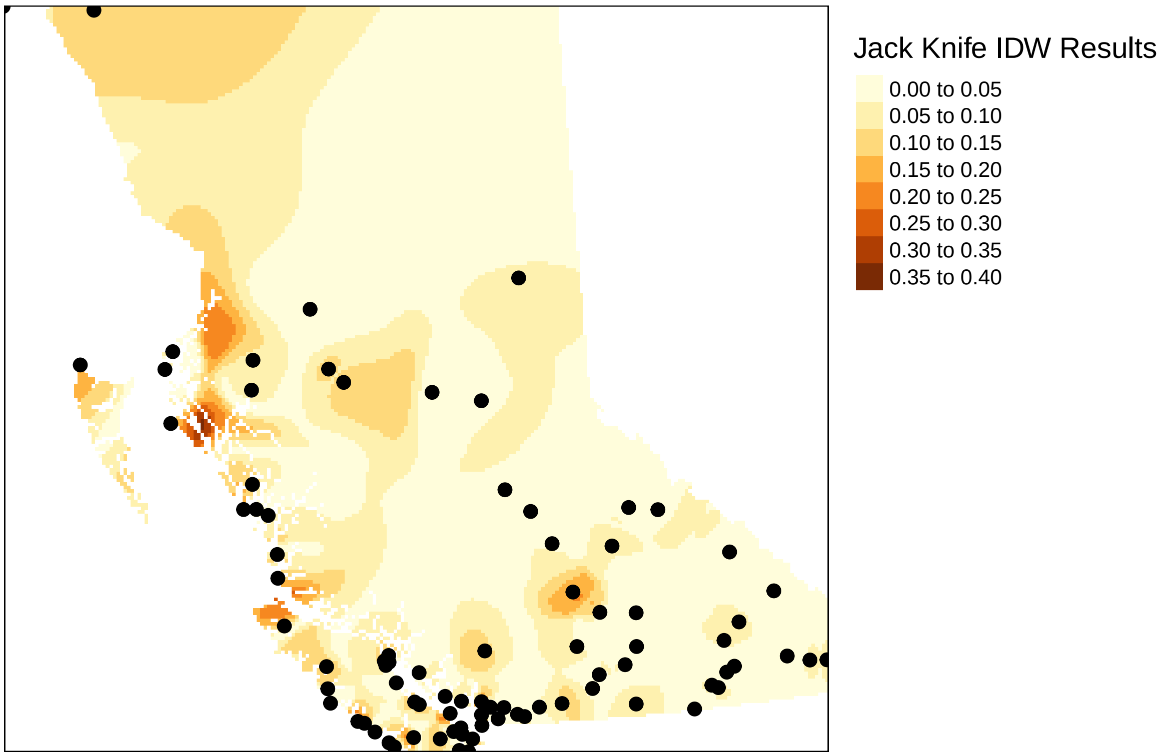

Figure 6. IDW JackKnife Raster

Looking to figure 6s jackknife results we can see there are a few regions which provide trouble for the interpolation.

The coastline along the Hecate Strait appears to provide inconsistency in the results, likely due to the lower than normal maximum temperatures seen there in previous figures clashing with more nearby stations. As well a station in the central interior is also producing some inconsistency, but with an opposite temperature difference from what was noted in the data.

|

|

Kriging Results

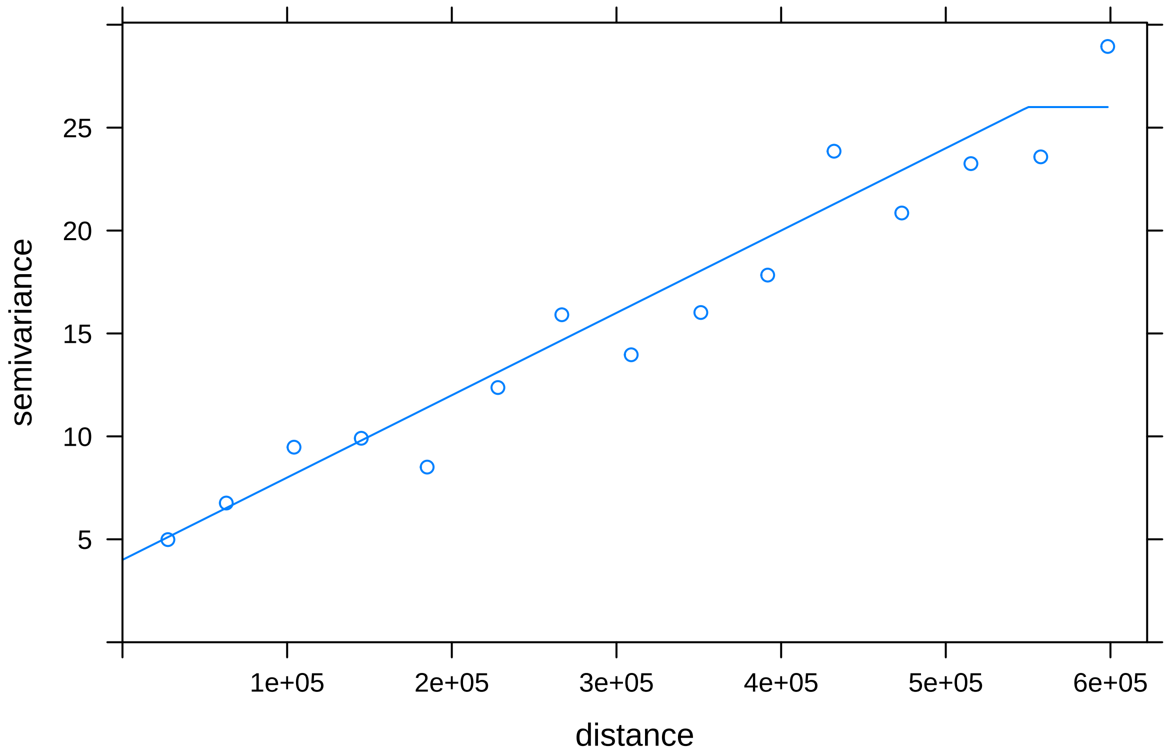

Figure 7. Kriging Variogram Linear Plot

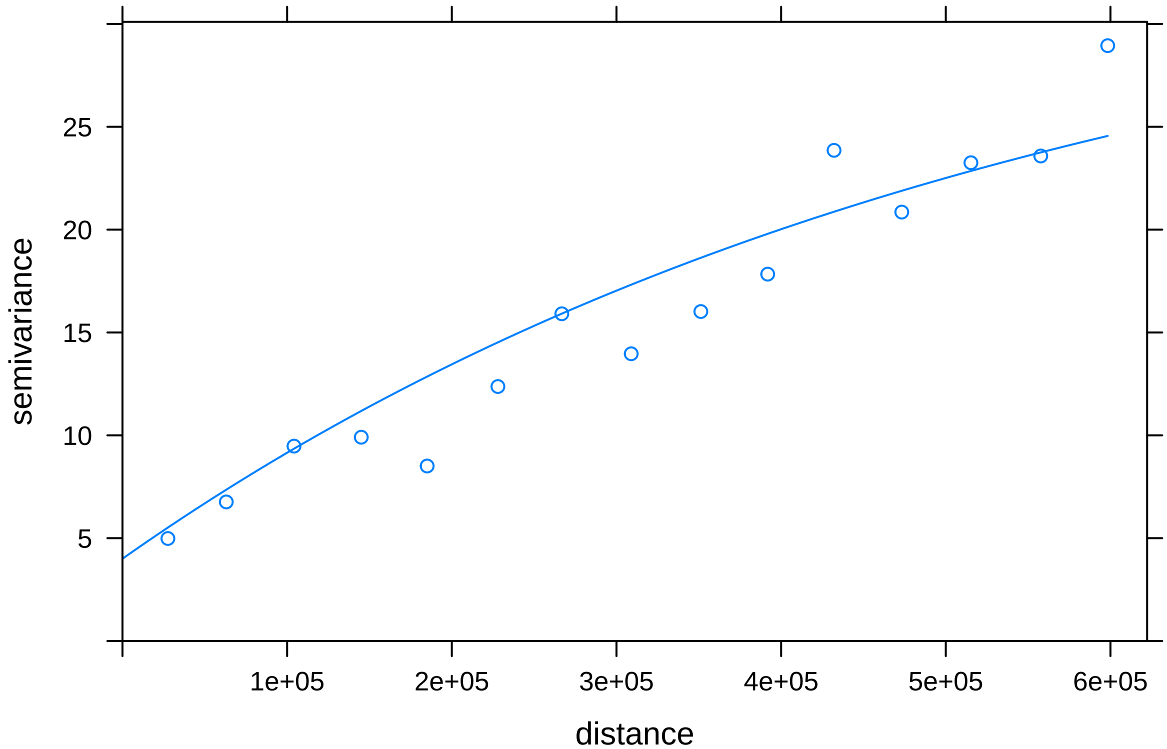

Figure 8. Kriging Variogram Exponential Plot

As seen in figure 7, though not typically used, the linear model for the semivariance within this data set better than the alternative Exponential. So the linear model was used further along with this analysis.

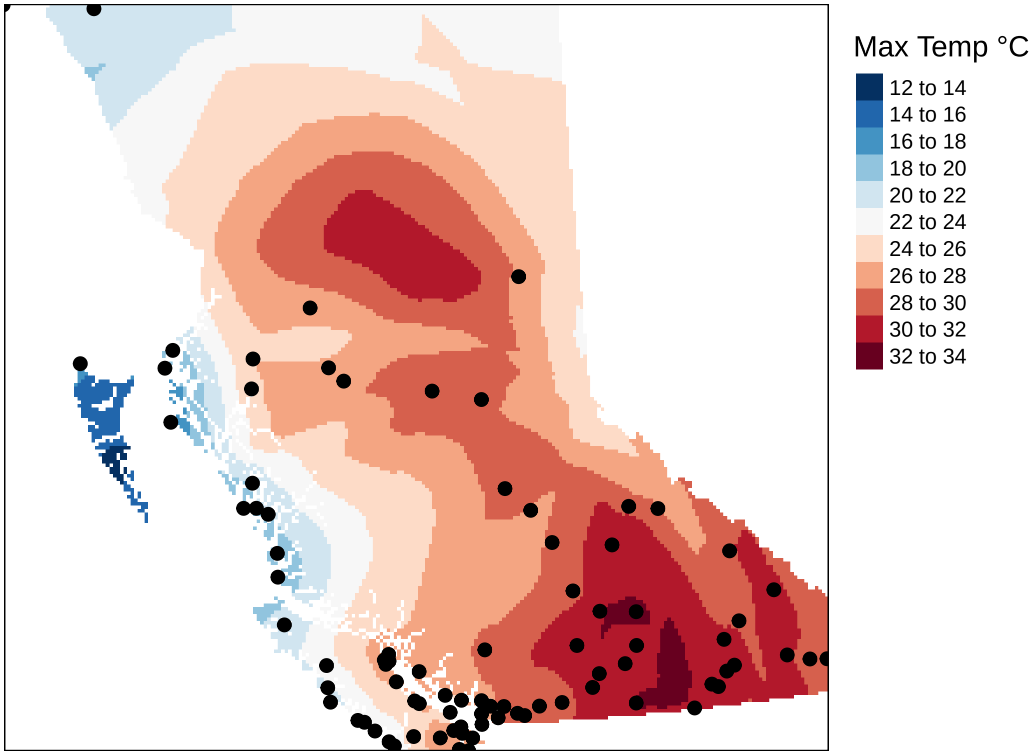

Figure 9. Kriged Interpolation Map

Looking at the results from the krig interpolation we again see a similar trend in maximum temperature to the other methods, particularly within the lower interior, still exhibiting very warm temperatures. Though other trends like the previously mentioned proximity to the Pacific Ocean also seem to be occurring as well.

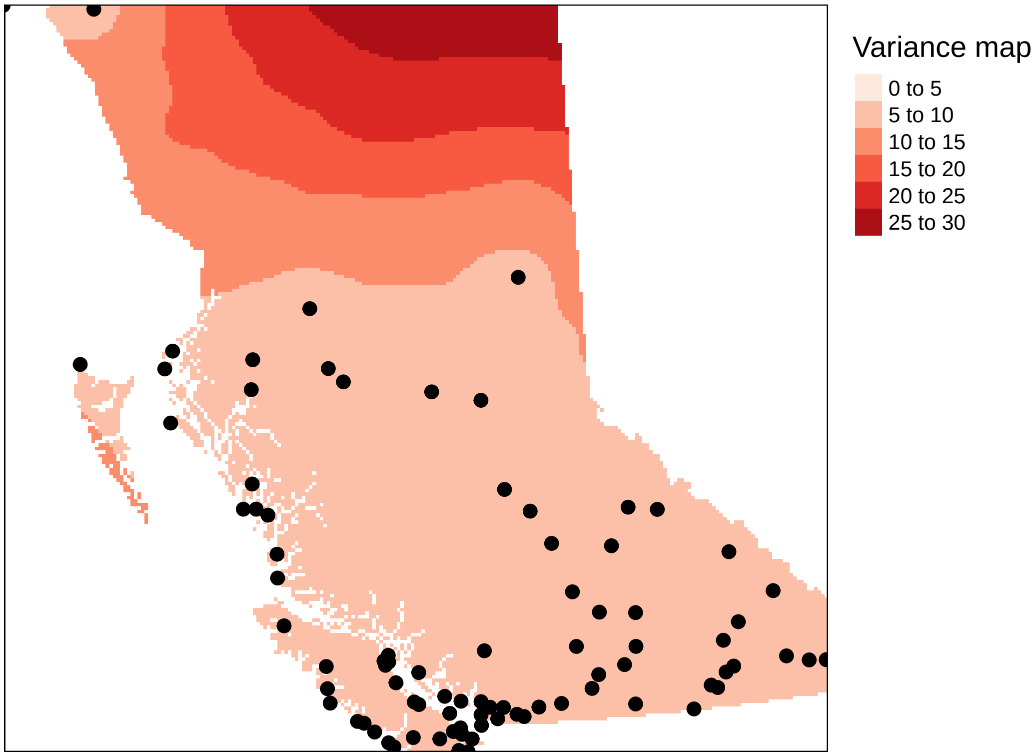

Figure 10. Kriging Variance Map

As to be expected regions which feature a greater volume of sample locations are less susceptible to noticeable variations

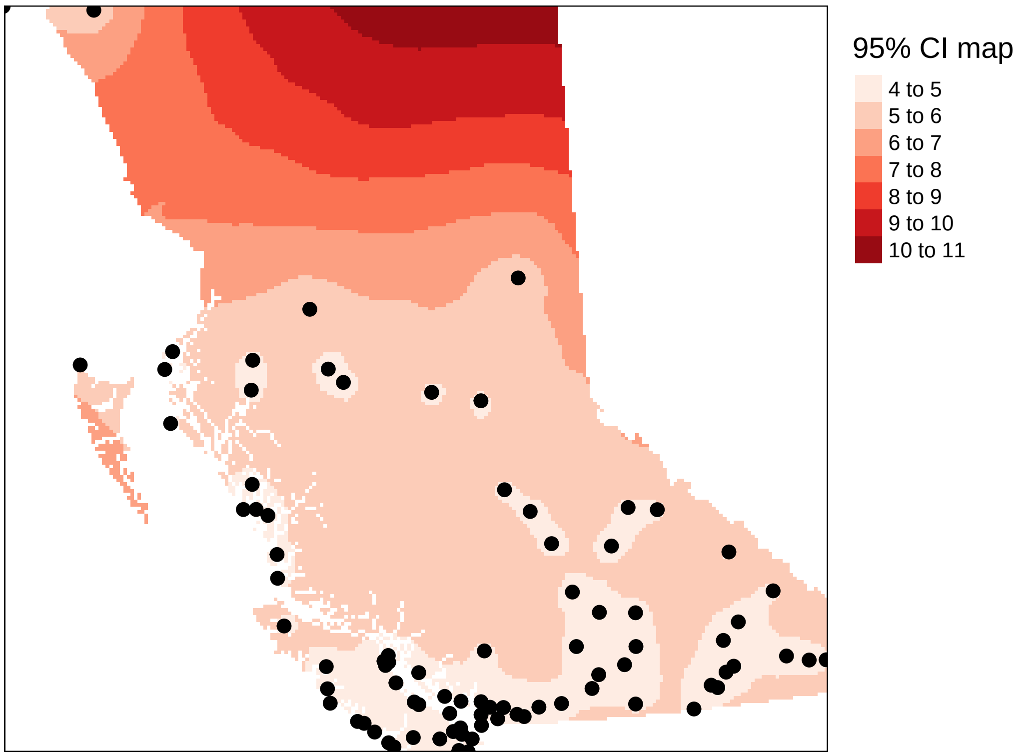

Figure 11. Kriging Confidence Interval Map

Similarly to Figure 10, the confidence of the kriging layer is still dependent on the prevalence of nearby samples to ensure confident interpolation.

Discussion

To compare each method of interpolation we can see that Thiessen polygons is a very rudimentary though straight forward method of interpolation. It does not feature any adjustable parameters beyond the sample data itself. Its best application would likely coincide with determining regions of influence (Naoum, 2004) or having a very quick interpolated surface as it does not require much computational power comparatively.

Inverse Distance Weighting on the other hand provided a much more robust method for developing an interpolated surface as well as a route to validate how well user generated parameterization met the real data. Though looking at figure 6s jackknife results we can still see there are regions of uncertainty at locations where strong deviations in temperature were seen. Though this is not unheard of given other studies have also seen extreme values providing roughness to their interpolation as well(Mitas, 1999).

Looking Kriging we see the greatest amount of control found given adaptions made to the semivariogram and the model used, providing a smooth interpolation. Its high level of control allowed for greater fitting, however and it should be noted that this increases chances of human error and potential other issues around over fitting (Jarvis & Stuart, 2001). That said this method also allowed a chance to identify regions of lower confidence and higher variance with the adaption from the generated variogram. Many locations were again noted to be uncertain in accuracy from interpolation, reinforcing the issue found in the lacking variety of sample space.

Overall, none of these techniques are perfect, some require more computational power, others are straight forward but rather inaccurate, but they all could benefit from a more diverse and uniform coverage of the province. IDW and Kriging handle the lack of diversity better than Thiessen polygons, which critically constrained in its accuracy by the presence of good sampling. Though IDW and Kriging methods also could benefit greatly from this as well. To get a more definitive answer regarding the effectiveness of interpolation methods on the province of BCs maximum temperature it would be wise to have a greater volume of stations for analysis as well combining other weather data such as humidity (Johnson et al., 2000) and other spatial variables such as elevation (DeGaetano & Belcher, 2007).

References

DeGaetano, A. T., & Belcher, B. N. (2007). Spatial Interpolation of Daily Maximum and Minimum Air Temperature Based on Meteorological Model Analyses and Independent Observations. Journal of Applied Meteorology and Climatology, 46(11), 1981–1992. https://doi.org/10.1175/2007JAMC1536.1

http://gisgeography.com. (2017, February 4). Kriging Interpolation - The Prediction Is Strong in this One - GIS Geography. GIS Geography. https://gisgeography.com/kriging-interpolation-prediction/

Jarvis, C. H., & Stuart, N. (2001). A Comparison among Strategies for Interpolating Maximum and Minimum Daily Air Temperatures. Part II: The Interaction between Number of Guiding Variables and the Type of Interpolation Method. Journal of Applied Meteorology and Climatology, 40(6), 1075–1084. https://doi.org/2.0.CO;2“>10.1175/1520-0450(2001)040<1075:ACASFI>2.0.CO;2

Johnson, G. L., Daly, C., Taylor, G. H., & Hanson, C. L. (2000). Spatial Variability and Interpolation of Stochastic Weather Simulation Model Parameters. Journal of Applied Meteorology and Climatology, 39(6), 778–796. https://doi.org/2.0.CO;2“>10.1175/1520-0450(2000)039<0778:SVAIOS>2.0.CO;2

Leroux, C., & Tisseyre, B. (2018). How to measure and report within-field variability: a review of common indicators and their sensitivity. Precision Agriculture, 20(3), 562–590. https://doi.org/10.1007/s11119-018-9598-x

NAOUM, S., & TSANIS, I. K. (2013). Ranking spatial interpolation techniques using a GIS-based DSS. Issue 1, 6(1), 1–20. https://doi.org/10.30955/gnj.000224

O’Sullivan, D., & Unwin, D. (2003). Geographic Information Analysis. In Google Books. John Wiley & Sons. https://books.google.ca/books?hl=en&lr=&id=K_kQN-GO7uUC&oi=fnd&pg=PR9&dq=david+o%27sullivan+and+david+unwin+kriging&ots=

b_Ht4V5xVr&sig=j0hTYRwOyFJV04I0adXi2Ri6dBQ#v=onepage&q=kriging&f=false

Simpson, G., & Wu, Y. (2014). Accuracy and Effort of Interpolation and Sampling: Can GIS Help Lower Field Costs? ISPRS International Journal of Geo-Information, 3(4), 1317–1333. https://doi.org/10.3390/ijgi3041317

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., & McKendry, I. G. (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and Forest Meteorology, 139(3-4), 224–236. https://doi.org/10.1016/j.agrformet.2006.07.004