Sept 27th 2016 By: Jordan (Yu-Lin) Wang
TA: Steeve Deschenes V00786970

Geography 418 Lab 1: Point Pattern Analysis

Introduction & Data:

The depletion of glaciers, ice caps, and ice sheets within the Arctic regions has been a major contributor to the increasing of global sea levels. These rising sea levels can have immense impacts on the complex habitats of many living species. Therefore, monitoring ice and glaciers melt durations is a critical process to the maintenance and observation of the sea level budget (Sharp, M. 2011). This can be accomplished through performing point pattern analysis; which is the evaluation of the distribution, or pattern of a set of observed points on surface. Based on the derived information from the analysis, it can inform us about the relationships in space or the possible reasons for the observed patterns (Comber,L. 2013).

The objectives of this lab are to explore the relationship between the meld duration within Canadian Arctic Archipelago and Arctic Regions by using three different methods in ArcGIS: quadrat analysis, nearest neighbor distance and K-function (Figure 1). Lastly, students are asked to explore the melt duration data using GeoDa to determine the local spatial associations between melt duration and regions. The data sets used to analysis the result are provided to us by the instructors within the Lab1 folder. Inside the lab folder, it contains climatology data set for the Canadian Arctic and feature classes such as Land, Melt Duration and Arctic Regions.

Goal:

The goal of this lab assignment is to compare and contrast the results of the different types of point pattern analysis methods; such as quadrat analysis, nearest neighbor distance and K-function.

Method:

Quadrat Analysis

This analysis allows the user to generate information that is used to describe the variability in the number of points per Quadrat. The processes can begin by importing data files of Melt Duration and Arctic Regions Classes provided by the lab folders onto ArcGIS. To observe number of melt duration point falls within each region, a Spatial Join is required to join melt duration observations with each region. By joining these two data sets based on spatial location, it can generate a new join-layer that contains attributes such as Average, Variance and Counts (tells the user how many melt duration observations were in each region). The generated join-layer's attribute table are then exported into Excel in the .dbf format for further calculation purposes. This Excel file contains crucial data to help calculate information like the Variance Mean Ratio, Variance of Frequency, Mean Cell Frequency, and Chi Square. The Mean Cell Frequency was generated by dividing the total count over total region and the Variance of Frequency was derived from the Excel equation of "=VAR(total count)". These two calculated value can be combined to produce the Variance Mean Ratio (Variance of Frequency/Mean Cell Frequency). By dividing the total average duration and total variance duration, we can generate the total Variance Mean Ratio, which can be combined together with degree of freedom to receive the Chi Square value (Total Variance Mean Ratio x Degree of Freedom = Chi Square).

Next, we used the derived values from above to help calculate for the Poisson distribution. This can be generated with the Excel equation, "=POISSON (Frequency, Mean Cell Frequency, FALSE)", and the derived results are then used to create the Poisson Probability graph (Figure 2). Finally, the Expected value is then calculated by multiplying the total regions (50) with the Poisson Probability (Figure 3). With this generated data, we can create a graph that compares the Observed Count VS the Expected Value (Figure 4).

Spatial Randomness Test

For the Average Nearest Neighbor Distance, we first started by calculating the total area of the Study Area (Canadian Arctic Archipelago). This is done by summarizing the total Area from the Region's attribute table with the statistic function and the result is a significant value of 9101471482565.1152 m^2. Next we used the spatial statistics tools provided by the ArcMap to generate the Average Nearest Neighbor distance using the Melt Duration file. Within this tool, we chose the Euclidean Distance for the Distance Method with the area of 9101471476122.110m^2 and selected the Generate Report to create a visual Average Nearest Neighbor Summary (Figure 5).

The next process is to generate a Ripley's K- Function graph from the help of Multi-Distance Spatial Cluster Analysis tool provided by the Arc Toolbox using the Melt Duration file. Within this tool, the distance bands chosen for this study were set to 10 bands and with a 99 permutations (also known as 95% confidence interval). We are also required to select the results graphically and with no weighted field. Lastly, the beginning distances and distance increment can greatly influence the results graph. Therefore, three different sets of distances were chosen for the beginning distances/ increment distance (5000, 10000, and 25000) to generate the best suitable K-Function graph for the given melt duration data.

Sept 27th 2016 By: Jordan (Yu-Lin) Wang
TA: Steeve Deschenes V00786970
<u>Spatial Autocorrelation- GEODA</u>

GeoDa is a set of software tools designed to implement techniques for exploratory spatial data on data such as points and polygons (Fargey. S, 2016). First, we exported the “Melt Duration” from ArcGIS as a Shapefile and then imported it into GeoDa. We then use this imported file to generate Quantile and Percentile maps by using the Map tool. For the Quantile Map, we classified it into 10 classified groups and are interested in the “Duration” value; similar processes were performed for the Percentile Map. Next we used the “Explore” tool to generate Histogram and Box Plot for our Melt Duration Data.

To create the LISA (Local Indicators of Spatial Association) plot, we must first create a spatial weight matrix with the help of Weights Manager from the Tools selection. This allows the user to create the matrix necessary for generating the Univariate LISA. This can be done by selecting the Univariate LISA from the Space bar and select “Duration” for the Weights File ID Variable. The created file contains 3 sets of graphs that will be further discussed in the Result section.

Result:

<u>Quadrat Analysis</u>

The Quadrat Analysis is a method that allows the user to study the spatial arrangement of point locations. In quadrat analysis, a grid of square cells of equal size [Regions] is used as an overlay on top of a map of incidents [Melt Duration] (Rogerson, P. 2001). In this lab, this analysis will be used to generate a static known as the Variance Mean Ratio, which can indicates the tendency either toward clustering in pattern (VMR greater than 1.0) or an evenly spaced arrangement (VMR less than 1.0). But first, the result of Mean Cell Frequency of 24.5800 has to be calculated and by using Excel, we can derive the result. The VMR generated was 1.69002117950, which indicates a tendency toward clustering in the pattern than random pattern (O’Sullivan, D. 2010).

To derive for the Chi Square value we first need to calculate the Variance Mean Ratio, which was 1.6900211 and times it with the degree of freedom (50-1). The confidence interval value was derived from the upper- tail critical values of chi-square distribution with degrees of freedom. The calculated chi square observed value was 82.81103779541.

Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and space (Haight. F, 1967).

Based on our data, the generated Poisson distribution table shows a positive skew trend that is similar to the Expected Value table. This means the mode is larger than the median and mean.

Spatial Randomness Test

The Nearest-Neighbor distance tool measures the distance between each feature centroid and its nearest neighbor's centroid location (ESRI. 2005). The Nearest Neighbor Index is expressed as the ratio of the Observed Mean Distance to the Expected Mean Distance. If the index is less than 1, the pattern exhibits clustering and if the index is greater than 1, the trend is toward dispersion. Based on our result of the Nearest-Neighbor distance, we have a ratio of 0.571109, which indicates a clustering pattern (ArcGIS. 2012). The generated Average Nearest Neighbor Summary also provides information such as the z-score, p-value, observed mean distance: 24563.629 Meters, and expected mean distance: 20960.1589 Meters. The generated z-score of -28.7759727717, means that there is less than 1% likelihood that the clustered pattern could be the result of random chance. This summary also provided us with the p-value of 0.0000, which will be further discussed in the discussion section.

Ripley's K-Function is also another way to analyze the spatial pattern of incident point data. A distinguishing feature of this toolset is that it summarizes spatial dependence over ranges of distances. For this assignment, the selection of an appropriate scale of analysis is required (Esri. 2005). The appropriate scale can influence the outcome for the graph. From the Nearest-Neighbor distance tool, we obtained that the observed mean distance is 24563.629m. With that said the appropriate scale for the beginning distances and distance increment should exceed the observed mean distance. To find the best suitable result of K-function is to generate several graphs by using different distances, such as 5000m (Figure 6), 10000m (Figure 7) and 25000m (Figure 8). These graphs all represent the same data sets; however, there is one graph that is the most suitable illustration of the graph. This will be further discussed in the discussion section.

Spatial Autocorrelation

LISA, short for Local Indicators of Spatial Association, is a set of tools designed for visualizing spatial association. It includes several tools such as Local Moran's I, Boxplots, Histograms. The Histogram shows that there are high frequency values from the duration ranging from 62.3 to 166 (Figure 9). From the Boxplots, we can observe that there is significant clustering from duration range between 90 and 130 (Figure 10). These two charts are different representations that are representing a similar trend based on the melt duration data. In the Quantile map, we can observed that there are higher duration value as it moves its way down south and vice versa, the duration value decreases as it moves towards North (Figure 11). The Percentile Map suggest that there are high percentage of melt duration closer towards the South East/West and has a low percentage of melt duration closer towards North (Figure 12).

Moran's I is the classic/common ways of measuring the degrees of spatial autocorrelation. By using the GeoDa, the derived Moran's I value is 0.874085 and this indicates that there is a trend of clustering in space (Figure 13). The LISA Cluster Map suggests that there are very high clustering of melt duration towards East, and South-West direction. Based on the LISA Significant Map, it is hard to derive much information from it due to the poor color choice (Figure 13).

Discussion:

Quadrat analysis allows the user to identify the number of melt duration points per region. However due to the nature of given data, most of the melt duration points are clustered and fall within few of the regions. This resulted with clustering within shorter distance, and can be observed from the K-Function graphs and based on the calculated VMR (1.69 greater than 1, therefore clustering). The size of the Quadrat can also influence the VMR result due to the melt duration points being covered within each region. The given data can greatly influence the result of Quadrat analysis and can lead to false observation. The Nearest Neighbor Analysis has the advantage of being simple to implement and can be very flexible. On the other hand, the Nearest Neighbor Analysis does not take into consideration the size of the regions when comparing the points and can influence the result (Cornelus, B. 2016). The Spatial Autocorrelation is a statistical measure of spatial dependence that has the strength of quantitative assessment of sign and value. However, the weakness of this spatial analysis is that there are no causal explanations for the spatial process that is occurring (Fargey, S. 2016).

One of the tasks is to identify the best suitable distance for the K-Function graph, because the selected beginning distance and increment distance will greatly influence our results. Based on Nearest Neighbor ratio and VMR, it suggested that there are clustering pattern and has an observed distance of 24563.629m. This information suggested that our total distance had to be greater than 25000m to include the entire range of the study area. With that said, several beginning/incremental distances were used; 5000m (total distance of 50000m), 10000m (total distance of 100000m) and 25000m (total distance of 250000m). The distance of 5000m and 10000m derived a very similar K-Function graph with a dispersed pattern at shorter distance and becomes much clustered as the distance gets further. By using the 25000m distance, the derived graph shows no disperse and has a steady trend of clustering as the distance gets further. This is due to the large scale chosen, the distance increments are so large that the dispersion at shorter distance cannot be identified.

Can null hypothesis of complete spatial randomness be rejected? As mentioned in the result section, the calculated chi square value was 82.81103779541. Based on the p-value and degree of freedom, the confidence interval value can be retrieved from the Chi Square table. The calculated chi square value is greater than the confidence interval of 66.339, and as a result we reject the null hypothesis. The Poisson distribution shows a positive skew trend and the observed count shows a random trend that has very high frequency at the start and rapidly drop down. The expected value of the observed count is also a positive skew trend, but the observed counts does not reflect a trend pattern.

Boundary problem in spatial analysis refers to a situation in which geographical patterns are governed by the shape and arrangements of boundaries that are drawn for measurement purposes. These boundary shape/size can influence the results of Nearest Neighbor and K-Function due to the melt duration points that fall within each area, a particular type of boundary problem is the Edge Effect. These arise when an artificial boundary is imposed on the study. Issues can arise, such as when sites in the center of the study area have nearby observations in all directions, and the sites at edges of the study area only have neighbors toward the center of the study area (O’Sullivan, D. 2010).

Conclusion:

In this assignment, several methods of spatial analysis have been compared to understand the spatial relationship of the melt duration point in the Canadian Arctic Archipelago. Based on the derived results from this assignment, we can observe a significant clustering of melt duration towards South West and South East. From the derived results of VMR and Nearest Neighbor Analysis, it suggested that there are clustering trend for the melt duration. The derived information suggest a significant clustering pattern. This can be the result of several factors such as higher temperature towards lower latitude, or human interactions that cause a higher concentration of melt duration. The three types of spatial analysis performed in this assignment give information regarding the relationship between the melt duration data and the trend. However, it does not provide explanations for several important questions, such as what is causing the trend and why the trend is occuring.

References:

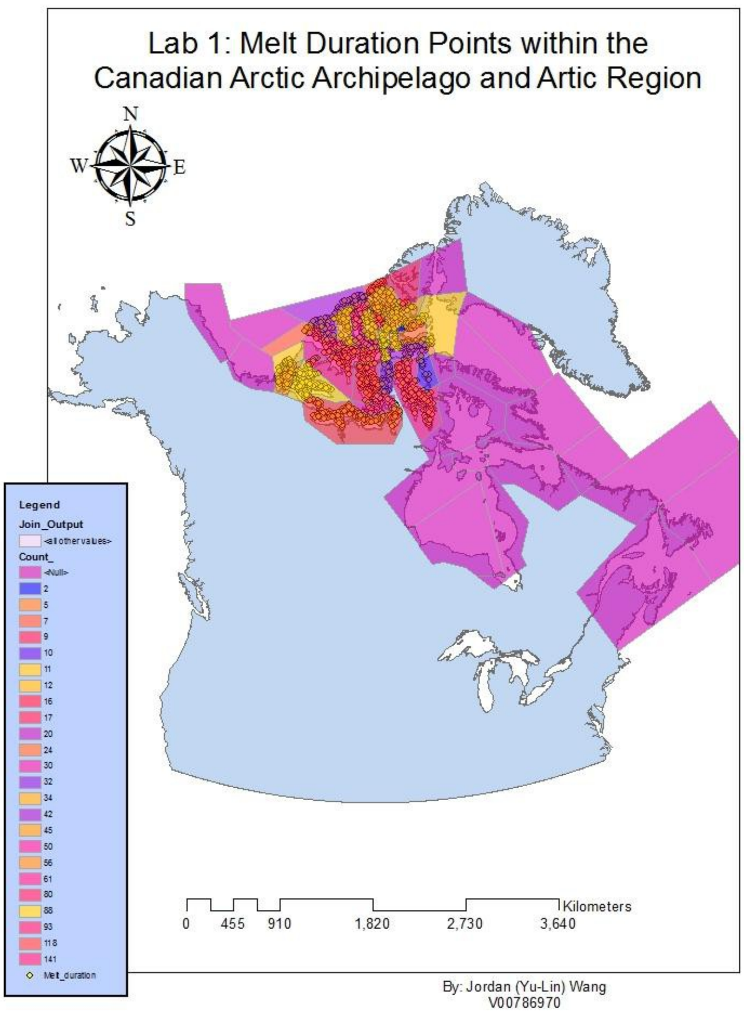
Sept 27th 2016 By: Jordan (Yu-Lin) Wang
TA: Steeve Deschenes V00786970
http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p00000008000000

lab instructions

http://www.academia.edu/16080548/Nearest_Neighbour_Analysis

http://alexei.nfshost.com/PopEcol/tables/chisq.html Chi square table

Figures:



Sept 27th 2016 By: Jordan (Yu-Lin) Wang
TA: Steeve Deschenes V00786970

Figure 1. Melt Duration Points within the Canadian Artic Archipelago

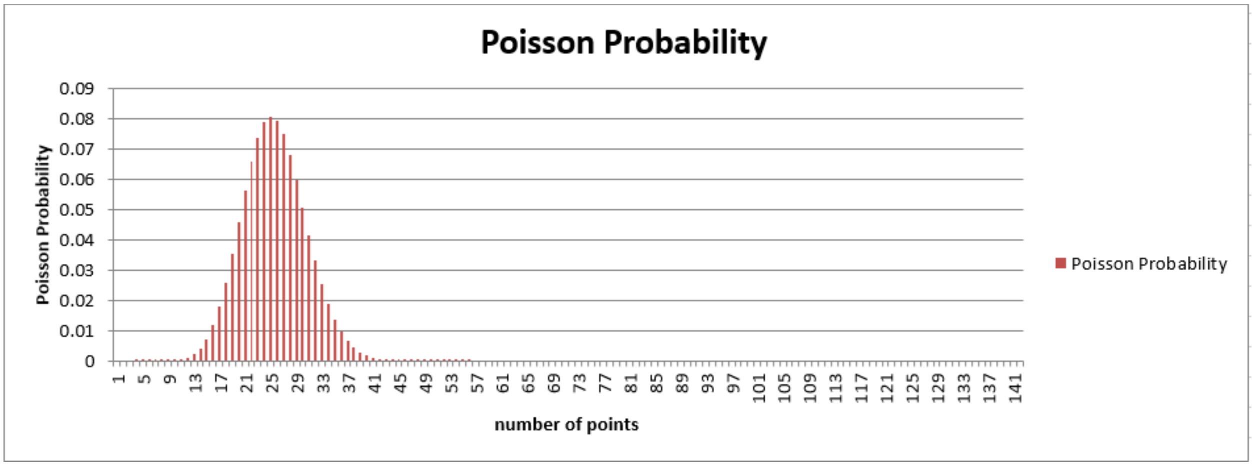


Figure 2. Poisson Probability

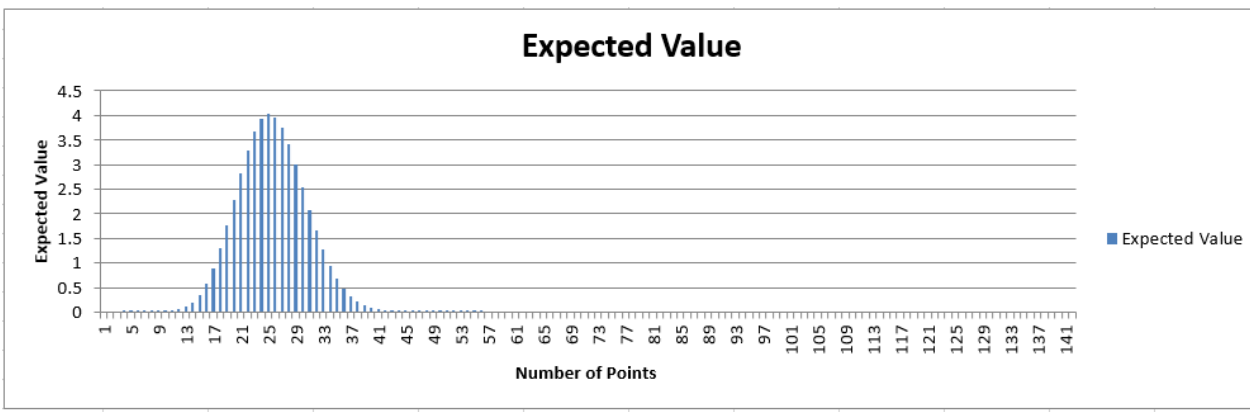


Figure 3. Expected Value

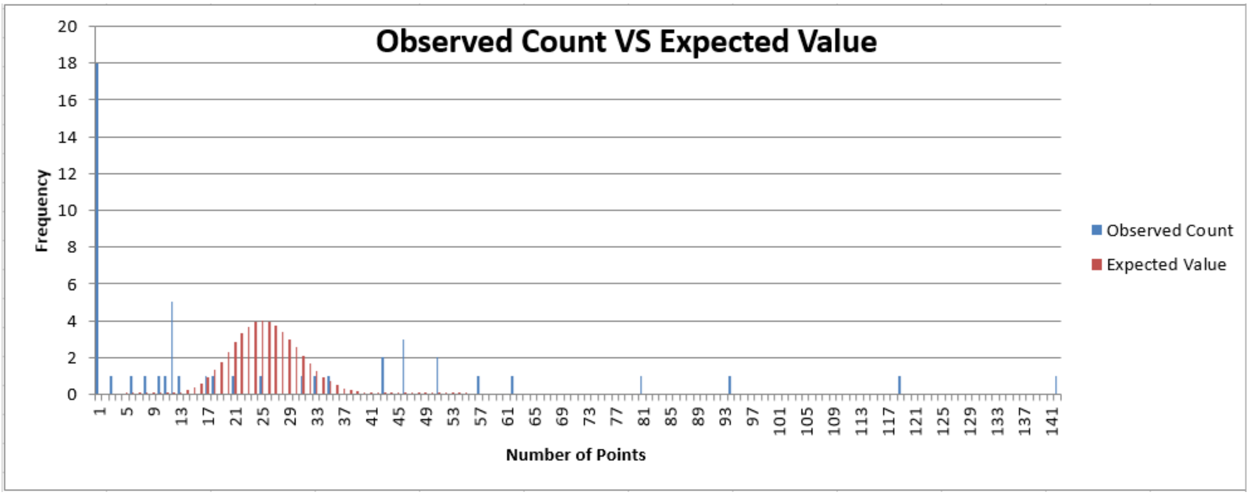
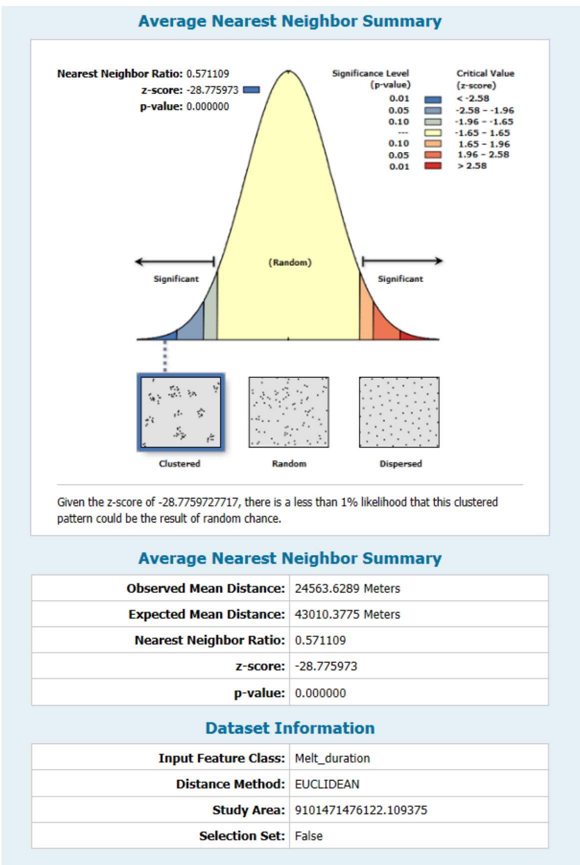


Figure 4. Observed Count VS Expected Value



Average Nearest Neighbor Summary

Given the z-score of -28.7759727717, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

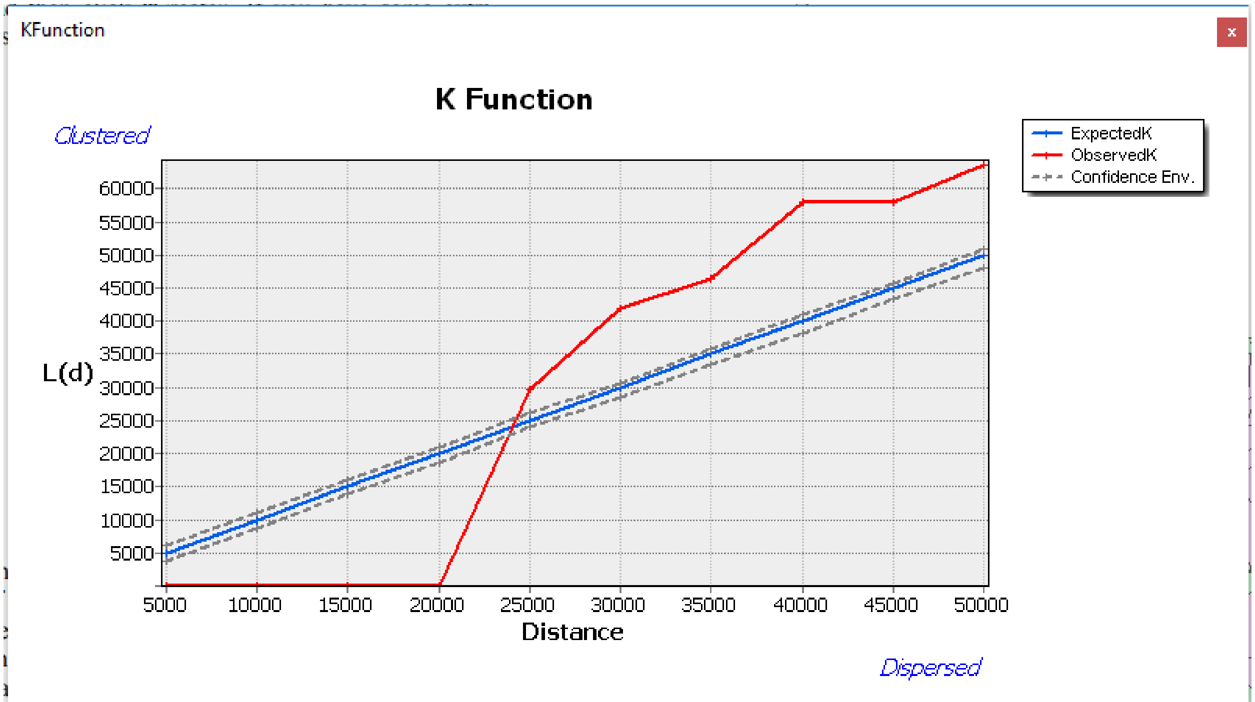
Average Nearest Neighbor Summary

Observed Mean Distance:	24563.6289 Meters
Expected Mean Distance:	43010.3775 Meters
Nearest Neighbor Ratio:	0.571109
z-score:	-28.775973
p-value:	0.000000

Dataset Information

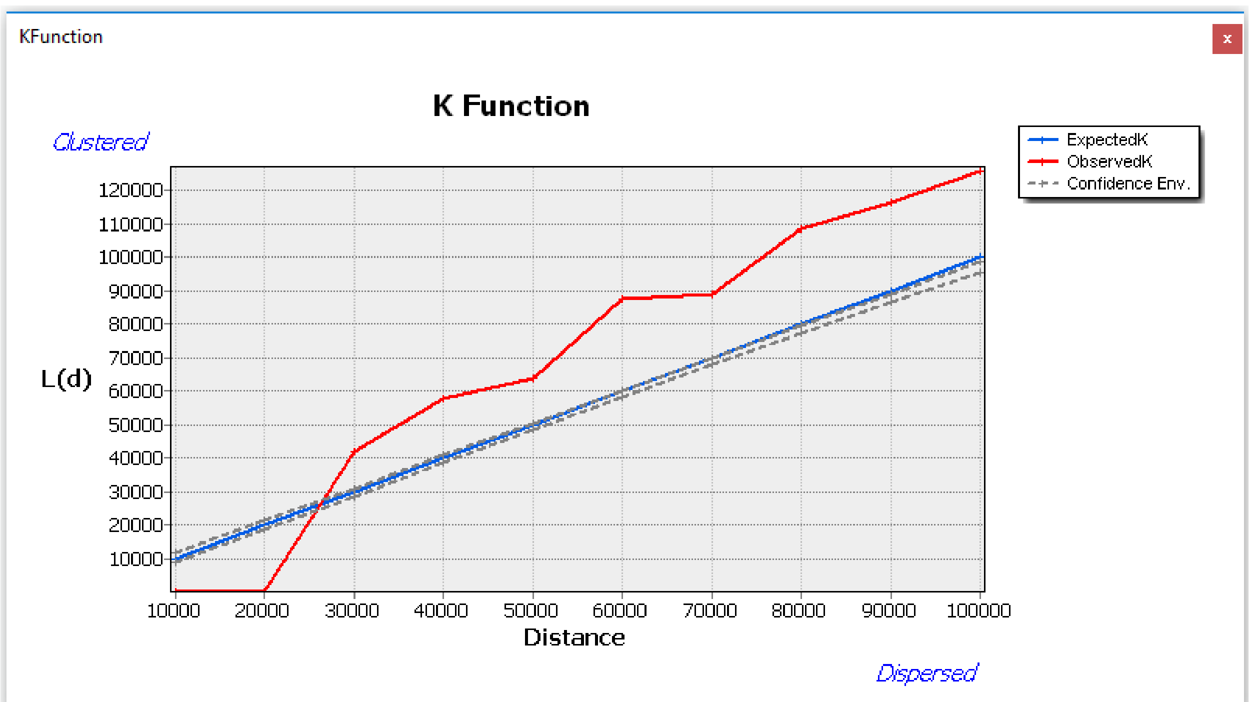
Input Feature Class:	Melt_duration
Distance Method:	EUCLIDEAN
Study Area:	9101471476122.109375
Selection Set:	False

Figure 5. Average Nearest Neighbor Summary



F

igure 6. K-Function 5000 meter beginning/increment



Sept 27th 2016 By: Jordan (Yu-Lin) Wang
TA: Steeve Deschenes V00786970
Figure 7. K-Function 10000 meter beginning/increment

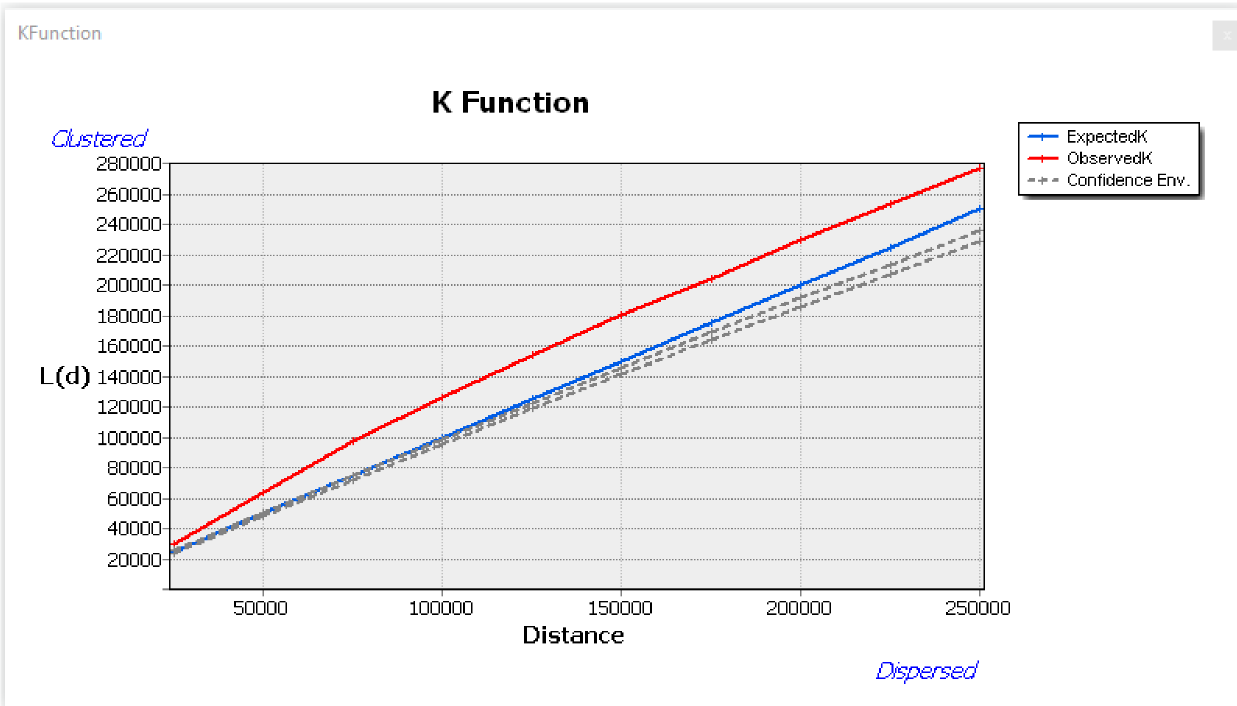


Figure 8. K-Function 25000 meter beginning/increment

By: Jordan (Yu-Lin) Wang
V00786970

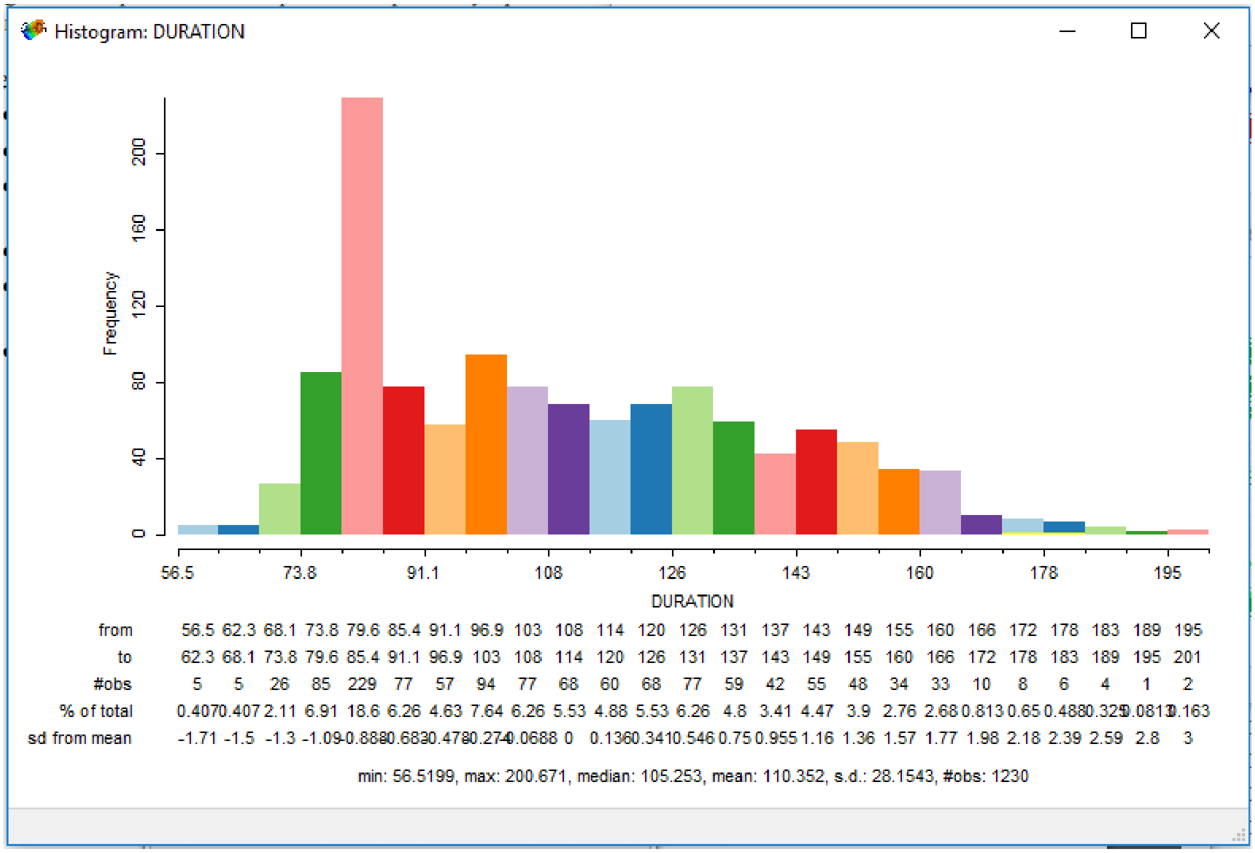


Figure 9. Histogram

Sept 27th 2016
TA: Steeve Deschenes

By: Jordan (Yu-Lin) Wang
V00786970

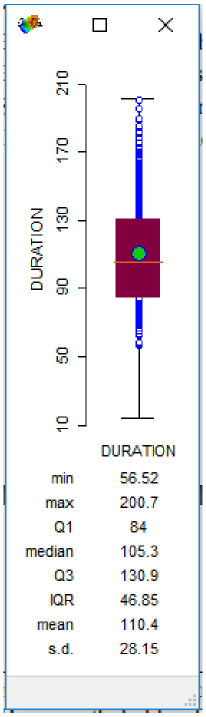


Figure 10. Boxplot

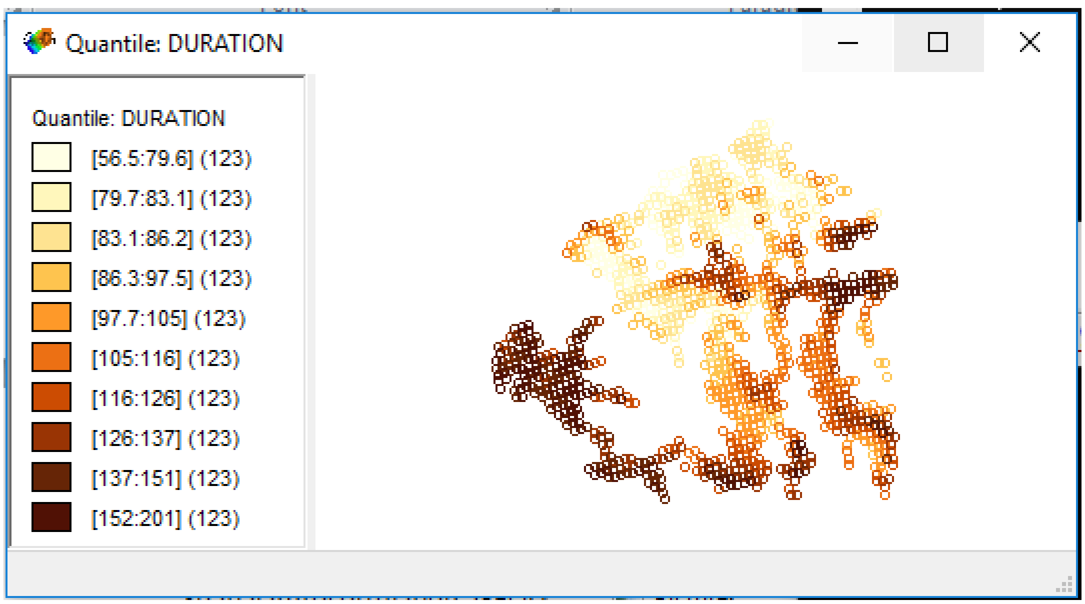


Figure 11. Quantile Map

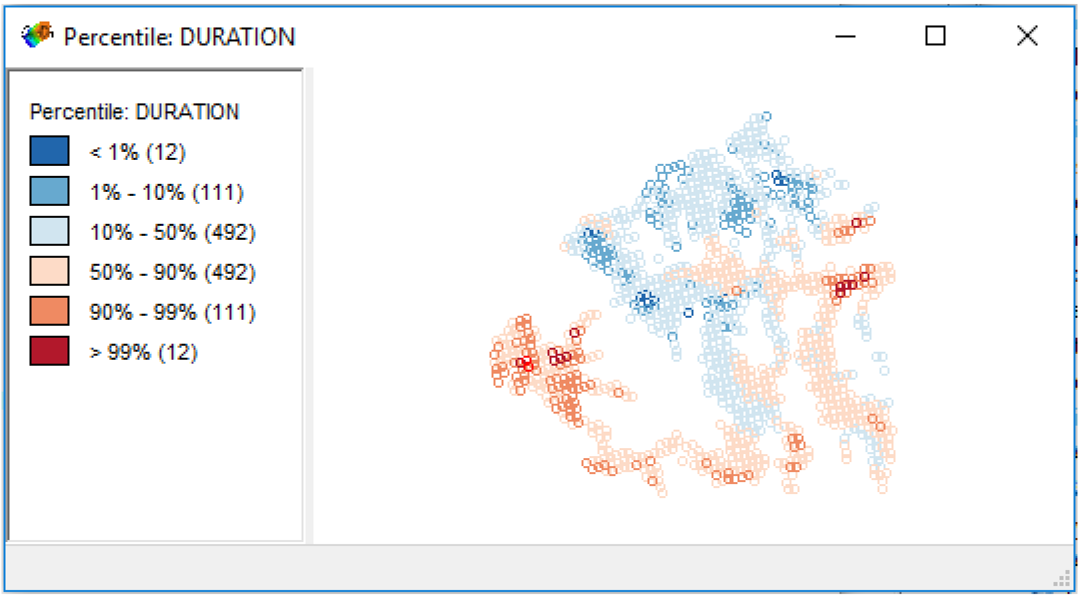


Figure 12. Percentile Map

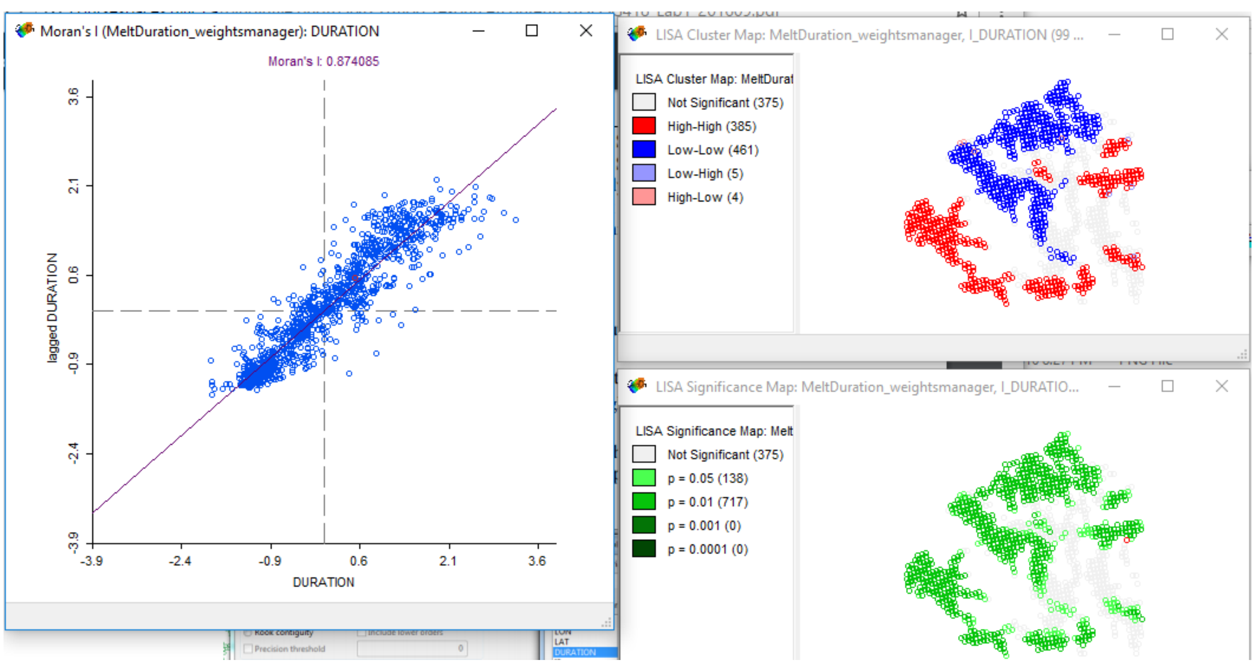


Figure 13. Univariate Local Moran's I, Cluster Map, and Significance Map