



Ted MacKinnon
AGRG
Dec 12 2003

ESRI Geostatistical Analyst

- The **geostatistical analyst extension** for ArcGIS 9 provides a number of interpolation techniques that use sample points to produce surfaces of the phenomena of interest

- Two main types of techniques:

Deterministic techniques are based on parameters that control either the extent of similarity (IDW) of the values or the degree of smoothing (radial basis) in the surface. These techniques do not use a model of spatial processes

Geostatistic techniques assume that at least some of the spatial variation of the natural phenomena can be modeled by random processes with spatial autocorrelation

- Describe and model spatial patterns (variography)
- Predict values at unmeasured locations (kriging)
- Assess the uncertainty associated with a predicted value at the unmeasured locations (kriging)

- This project dealt with exploring the geostatistical analyst extension to produce surface models to aid in the AGRG climate study

Brief Procedure:

- An excel spread sheet was provided containing historical weather data from Environment Canada weather stations
- Average Precipitation, average 5GDD, average 10GDD, and the lowest possible temperature attributes were examined
- A layer file was created of points representing geographical locations of the weather stations
- There were 15 stations that were located within the study area
- Data was converted into a dbf file and then joined to the layer file
- An ArcMap project was set up containing all of the spatial data (weather station point data, clipped polygon of the study area, and a shaded relief model of the Annapolis Valley)
- The following 4 models were created using the default options (Ordinary Kriging – Prediction method) of the geostatistical analyst (before statistically exploring the data)

Lowest Temperatures

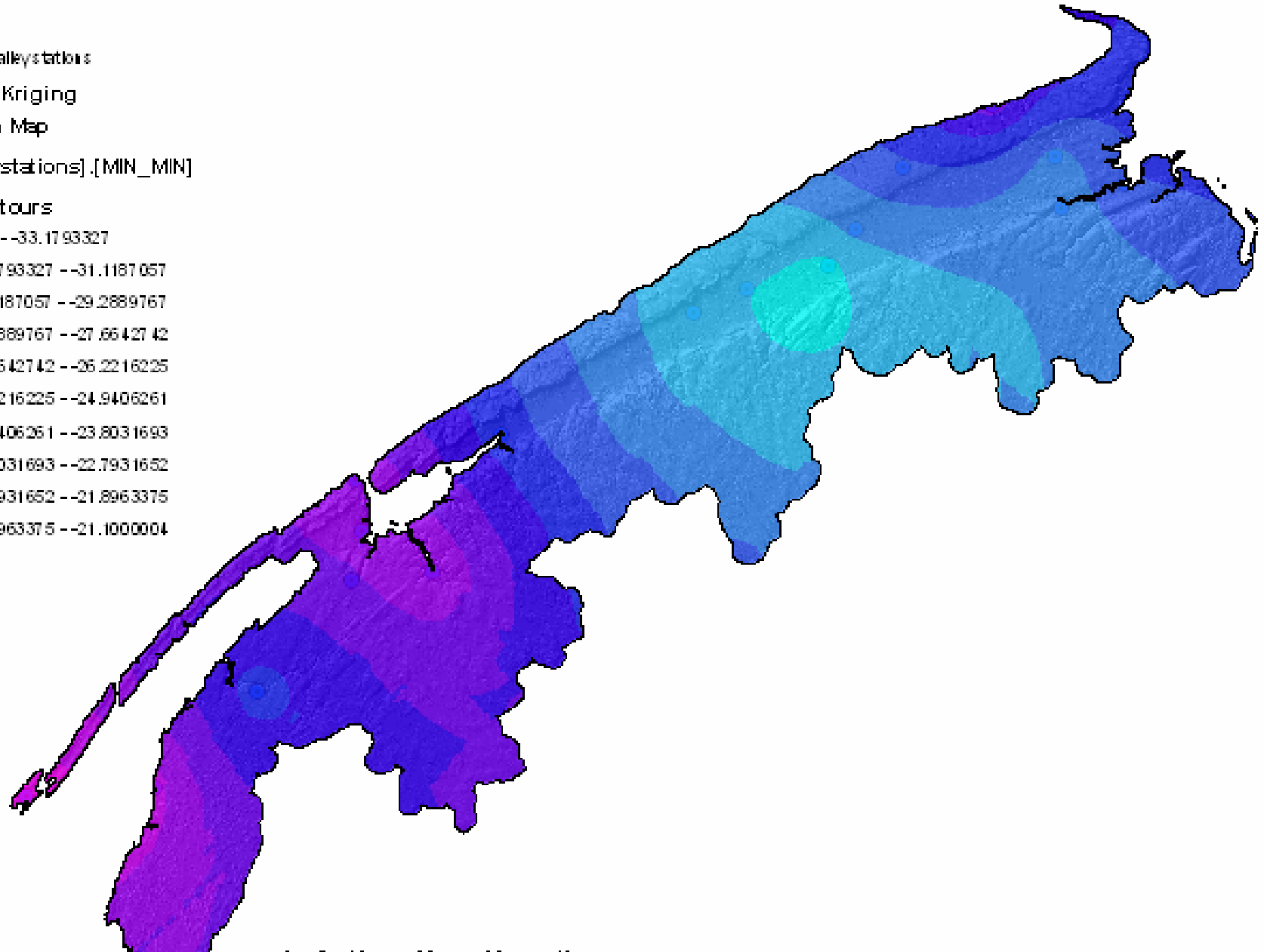
● EC-valleystations

Ordinary Kriging

Prediction Map

[EC-valleystations].[MIN_MIN]

Filled Contours



Average 5GDD

● EC-valleystations

Ordinary Kriging_2

Prediction Map

[EC-valleystations].[AVG_5GDD]

Filled Contours

1293.900024 - 1429.738892

1429.738892 - 1530.662231

1530.662231 - 1605.644653

1605.644653 - 1661.353882

1661.353882 - 1702.743774

1702.743774 - 1733.494995

1733.494995 - 1774.885010

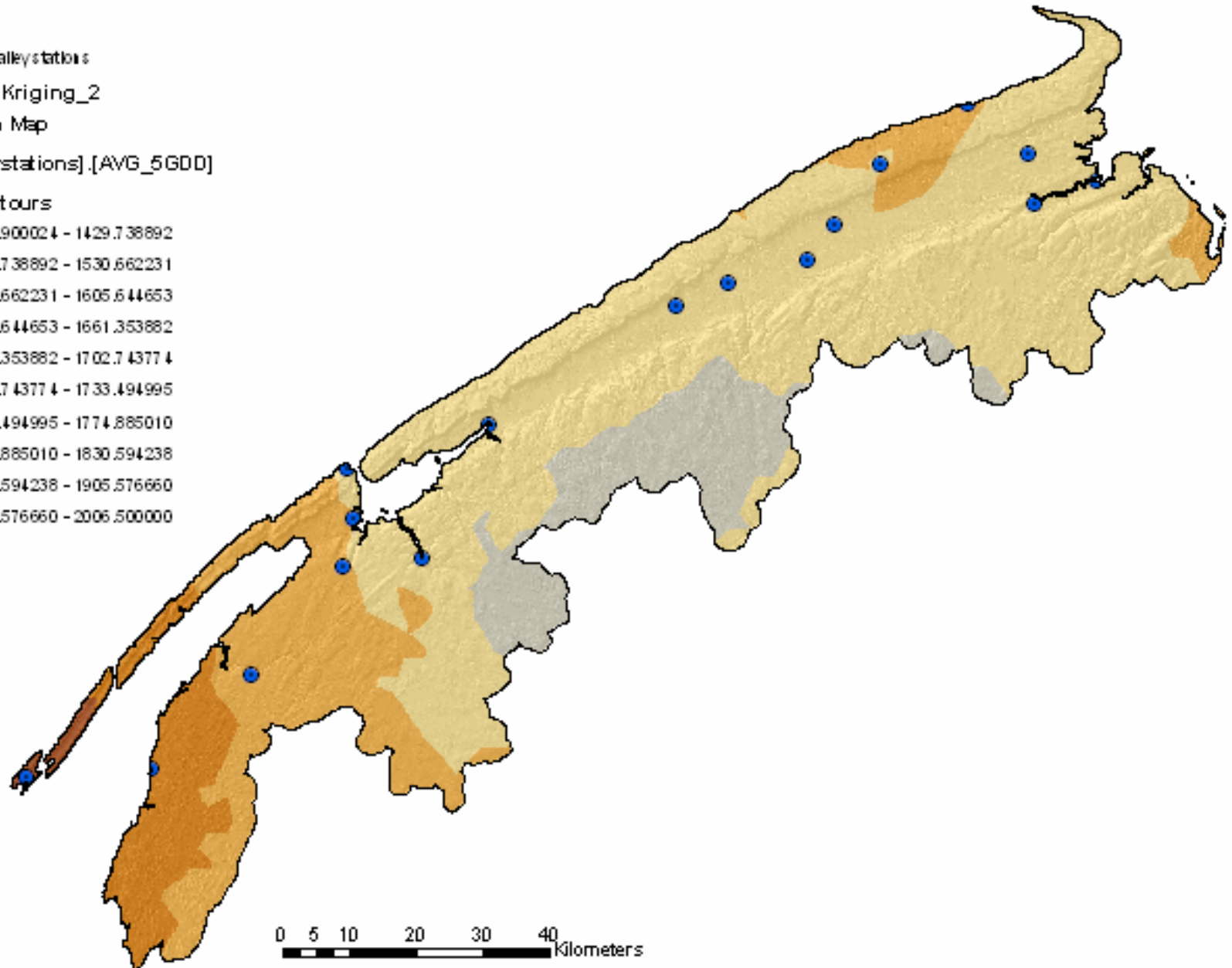
1774.885010 - 1830.594238

1830.594238 - 1905.576660

1905.576660 - 2006.500000



0 5 10 20 30 40 Kilometers



Average 10GDD

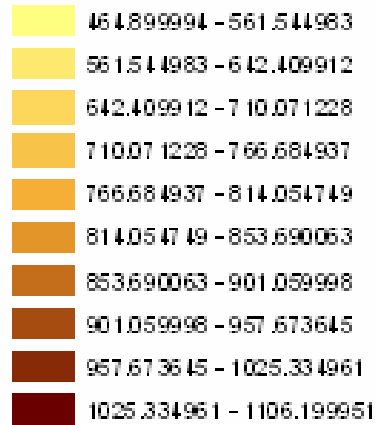
● EC-valleystations

Ordinary Kriging_3

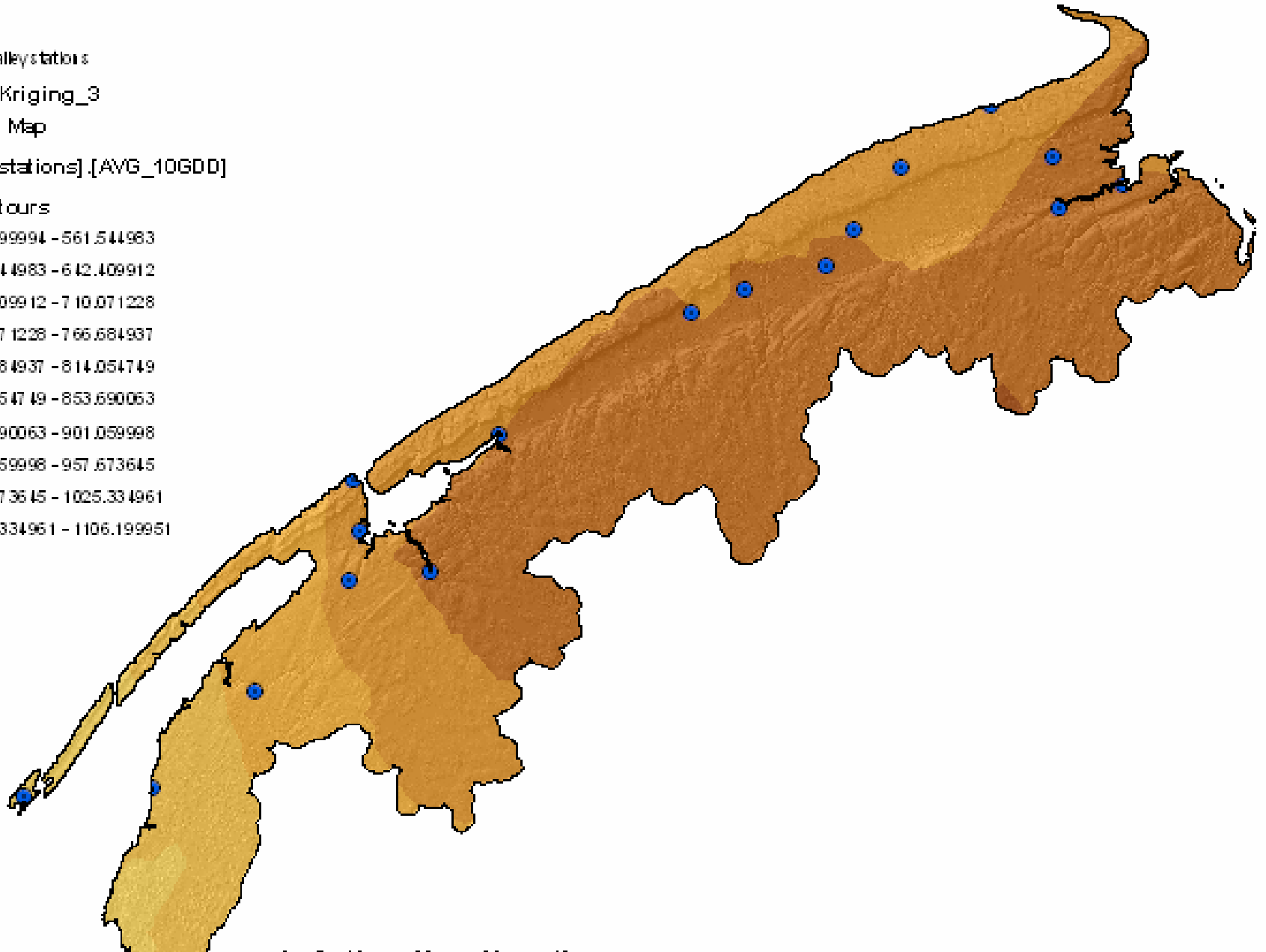
Prediction Map

[EC-valleystations].[AVG_10GDD]

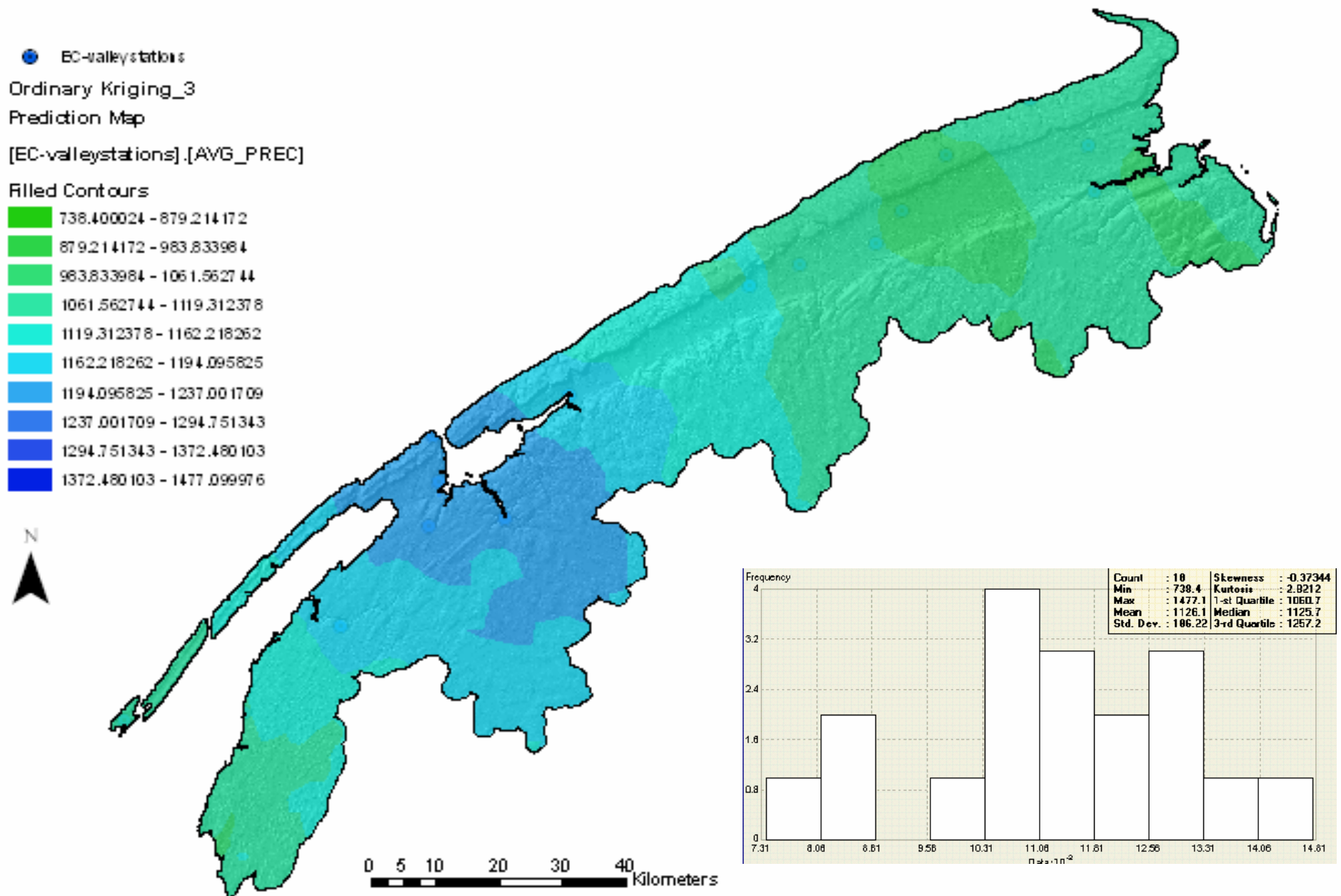
Filled Contours



N



Average Annual Precipitation



Average Annual Precipitation

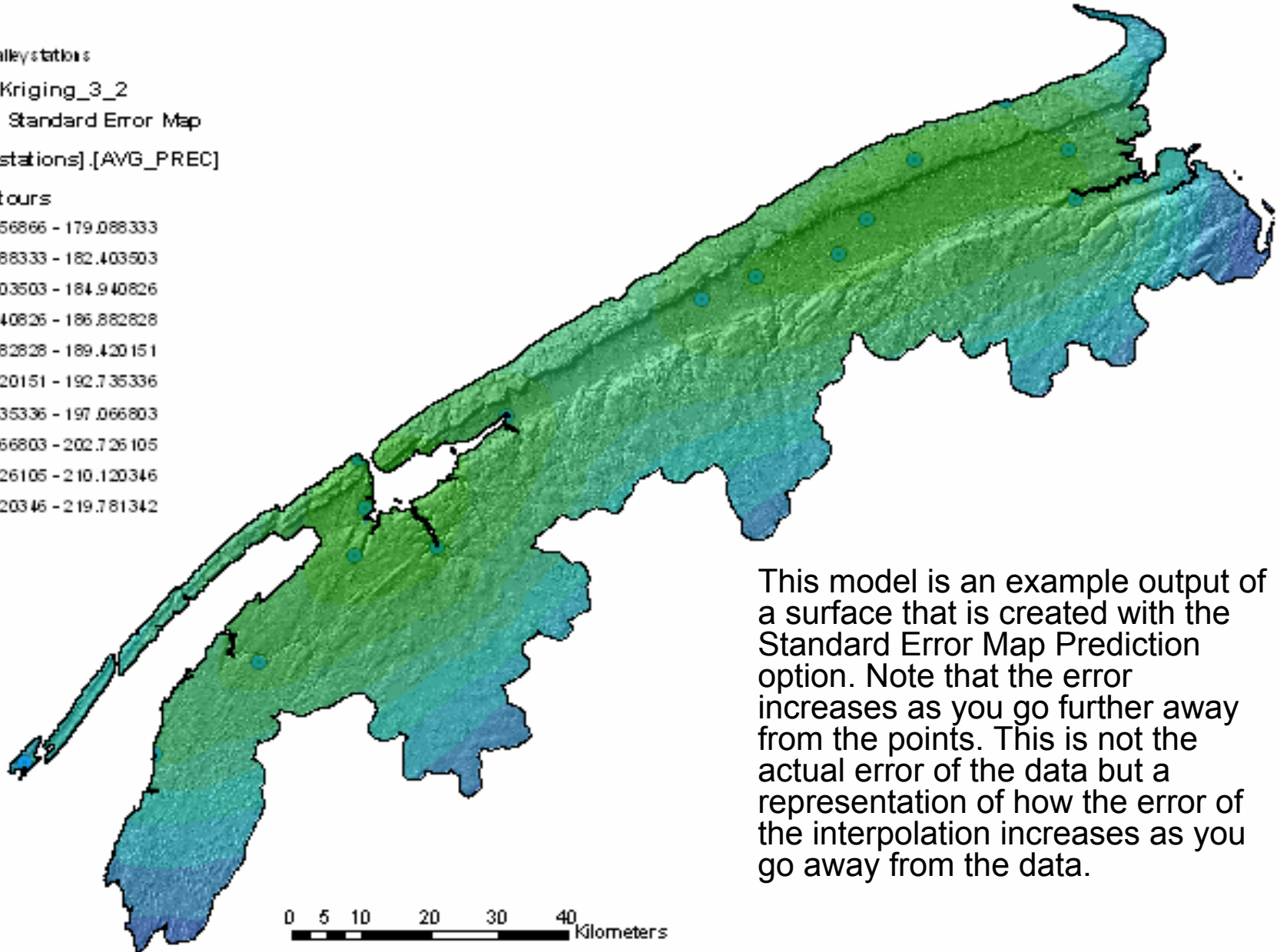
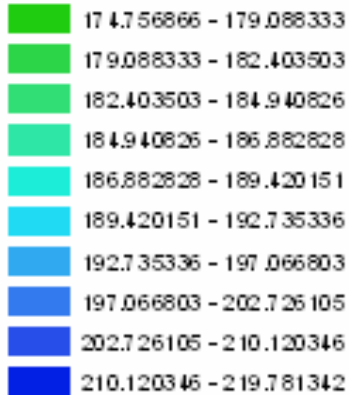
● EC-valleystations

Ordinary Kriging_3_2

Prediction Standard Error Map

[EC-valleystations].[AVG_PREC]

Filled Contours



This model is an example output of a surface that is created with the Standard Error Map Prediction option. Note that the error increases as you go further away from the points. This is not the actual error of the data but a representation of how the error of the interpolation increases as you go away from the data.

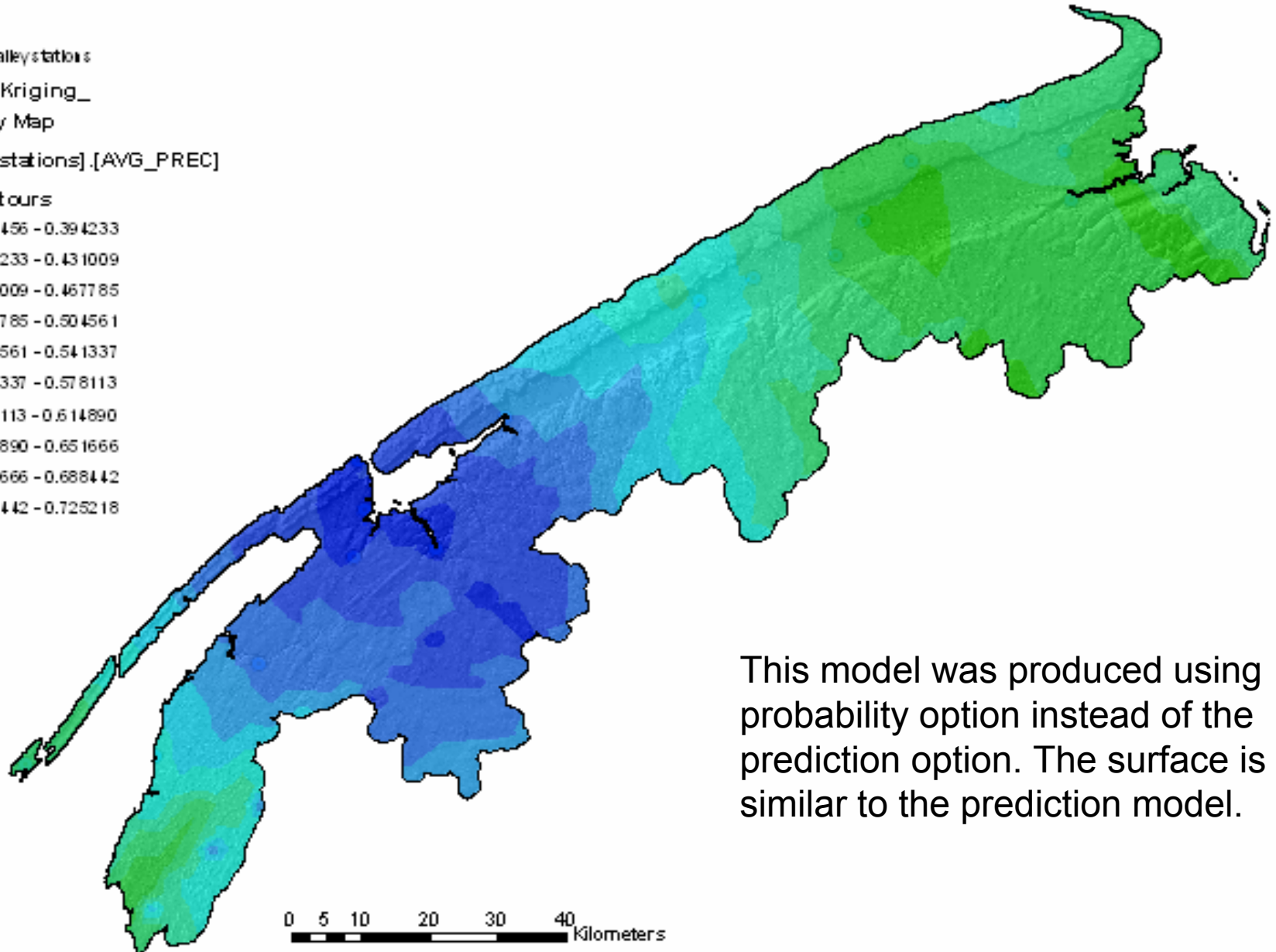
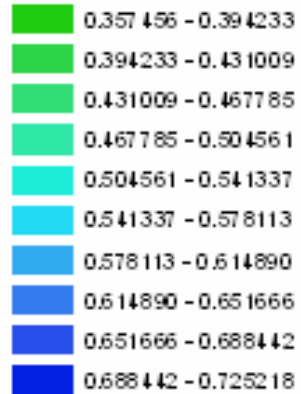
Average Annual Precipitation

● EC-valleystations

Ordinary Kriging_
Probability Map

[EC-valleystations].[AVG_PREC]

Filled Contours



This model was produced using probability option instead of the prediction option. The surface is similar to the prediction model.

Exploring the Data

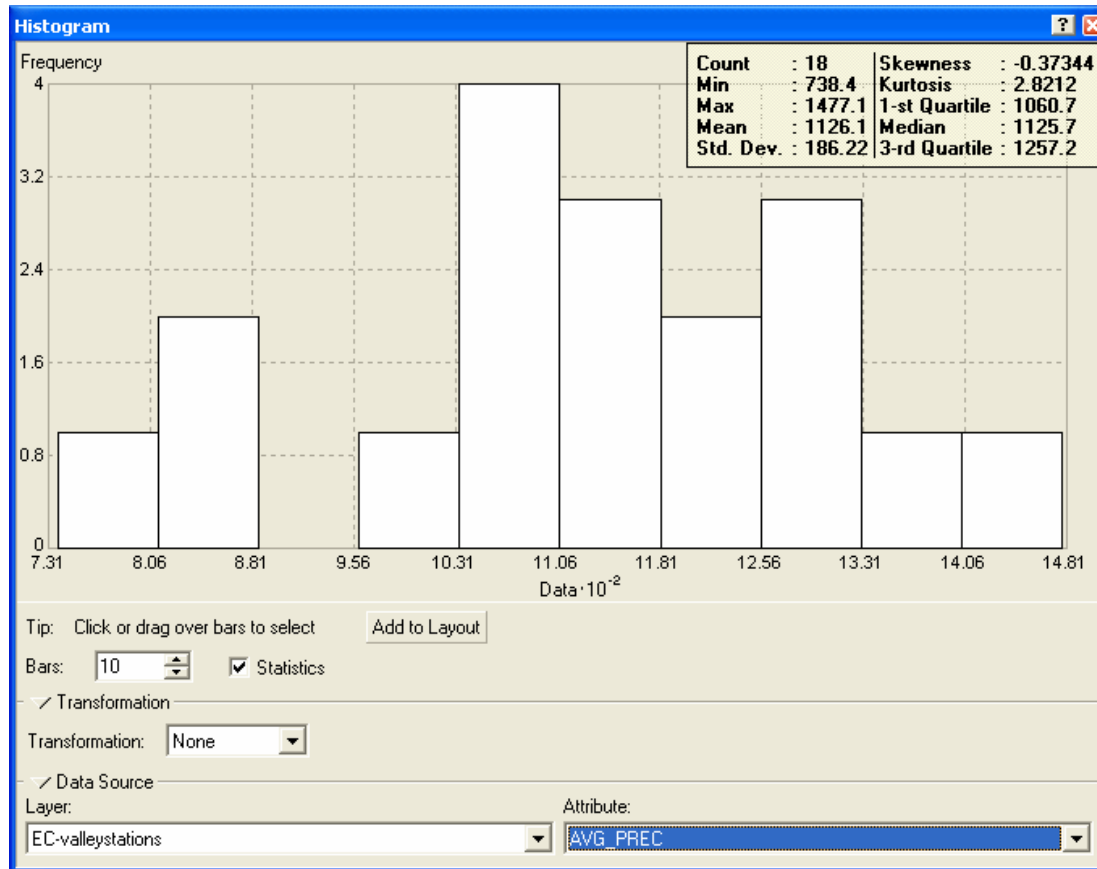
- Before using the interpolation techniques you can explore your data with the ESDA tool set
- This will allow you to select the appropriate parameters for the interpolation model
- Normal distribution is assumed when using the defaults

- Some of the available tools are:
 - Histogram
 - QQPlot
 - Trend Analysis
 - Semivariogram/Covariance Cloud

Histograms

- Interpolation methods that are used to generate modeled surfaces provide the best results if the data is normally distributed (bell shape curve)
- If the data is skewed you may choose to transform the data to make it normal
- Important to understand the distribution of the data before creating a surface
- The histogram tool plots frequency for the attributes in the dataset enabling you to examine the univariate distribution of the dataset for each attribute
- The distribution of the attribute is depicted by a histogram with the range of values separated into classes
- The relative proportion of data within each class is represented by the height of each bar.
- Important features of the distribution are its central value, its spread, and its symmetry.
- If the mean and median are approximately the same value, you have one piece of evidence that the data may be normally distributed.
- Histogram is unimodal (one hump) and fairly symmetric and appears to be close to a normal distribution.

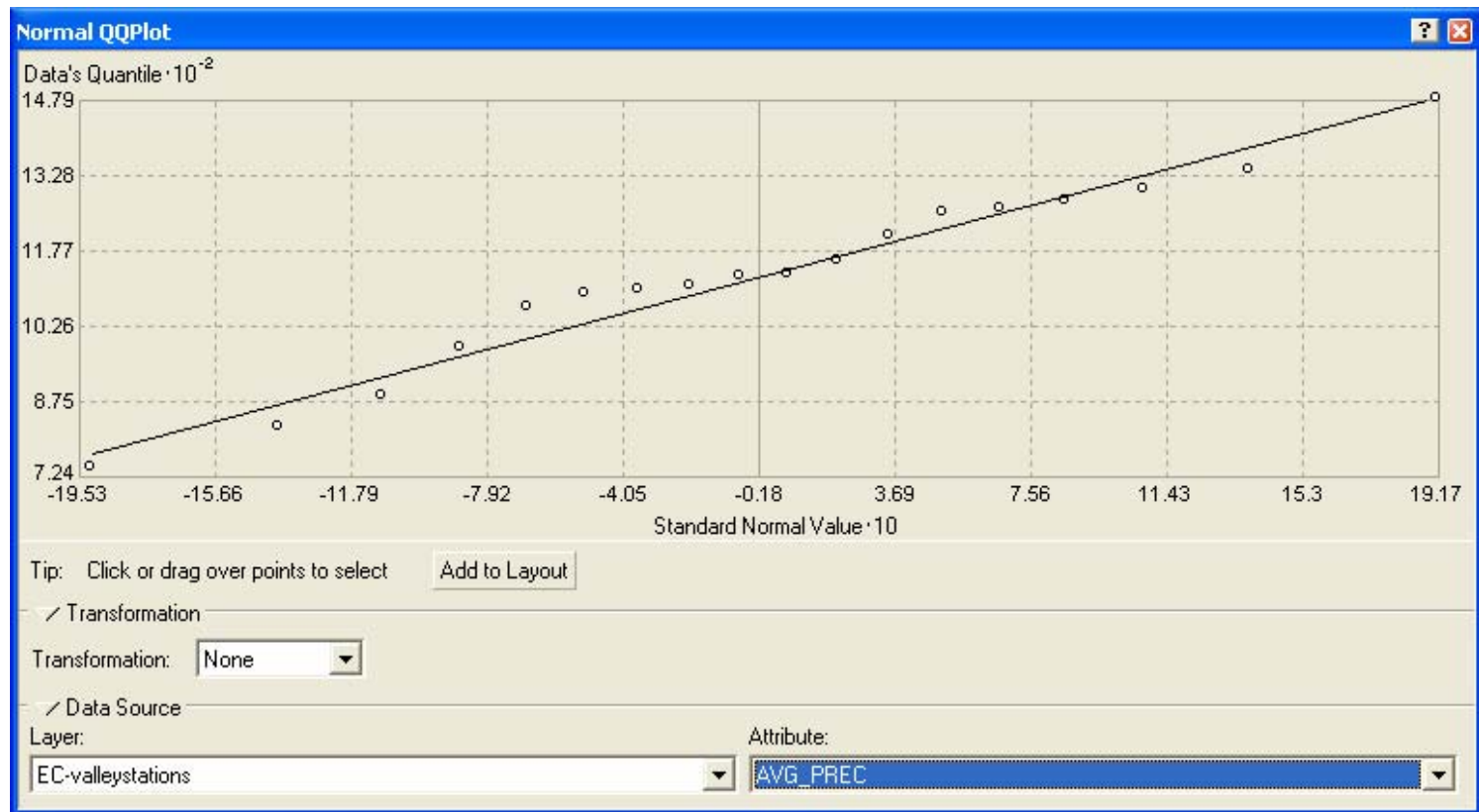
Example Histogram output



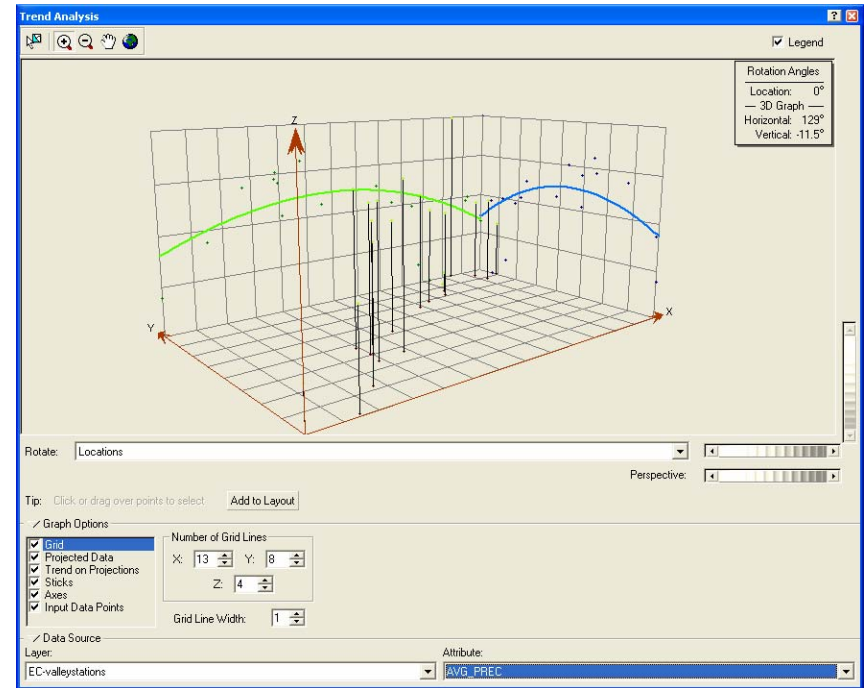
QQ Plots

- Compare the distribution of the data to a standard normal distribution, providing yet another measure of the normality of the data.
- The closer the points are to a straight line, the closer the distribution is to being normally distributed.
- A general QQPlot is a graph on which the quantiles from two distributions are plotted verses each other
- For two identical distributions, the QQPlot should be a straight line.
- If the data does not exhibit a normal distribution (same with histogram) it may be necessary to transform the data to make it conform to a normal distribution before using interpolation techniques

Example QQPlot output

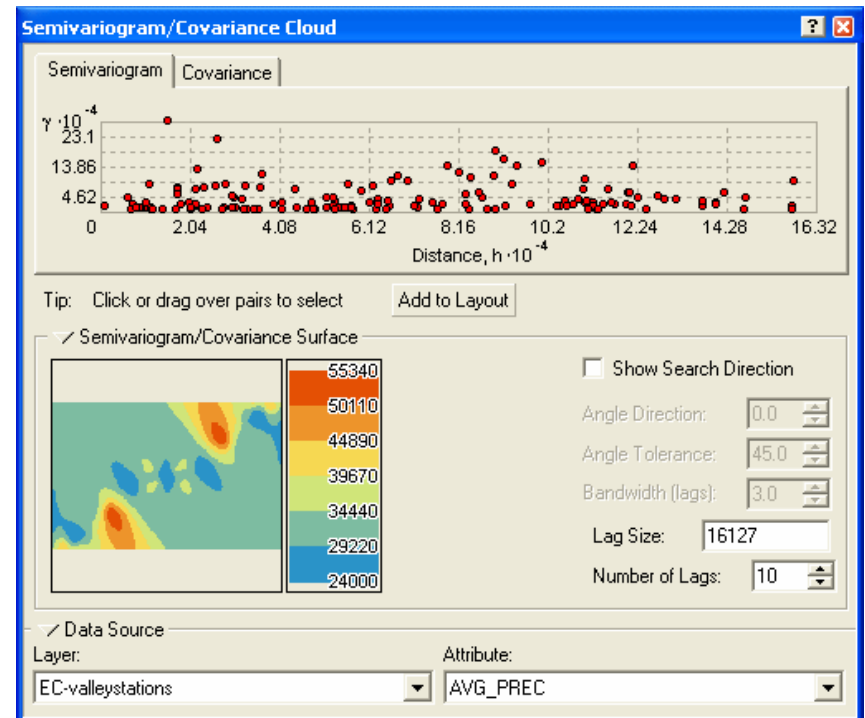


- You can Identify global trends in your data with the trend analysis tool
- Trend analysis tool provides a 3D perspective of the data



- If a trend exists in the data, it is the non random component of a surface that can be represented by some mathematical formula.
- If the trend surface does not adequately portray your surface for your particular need, you may want to remove and continue with your analysis, modelling the residuals, which is what remains after the trend is removed.
- The trend analysis tool enables you to identify the presence/absence of trends in the input data set.
- A vertical stick in the trend analysis plot represents the location and value of each point
- The points are projected onto the perpendicular planes, an east-west and a north-south plane.
- A best fit line (polynomial) is drawn through the projected points, which model trends in specific directions
- If the line were flat, this would indicate that there would be no trend

- Allows you to examine the spatial autocorrelation (assumed that things are close to one another are more alike) between the measured sample points
- A semivariogram value, which is the difference squared between the values of each pair of locations, is plotted on the y axis relative to the distance separating each pair on the x axis
- Each red dot represents a pair of locations

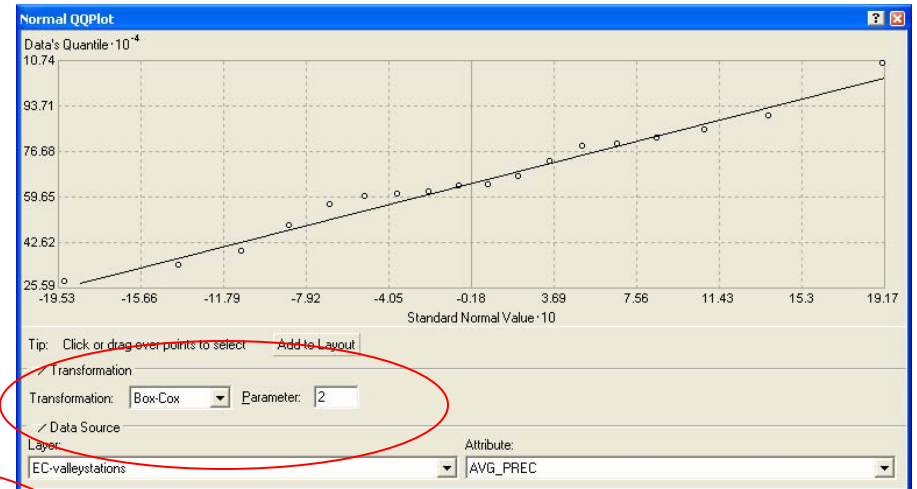
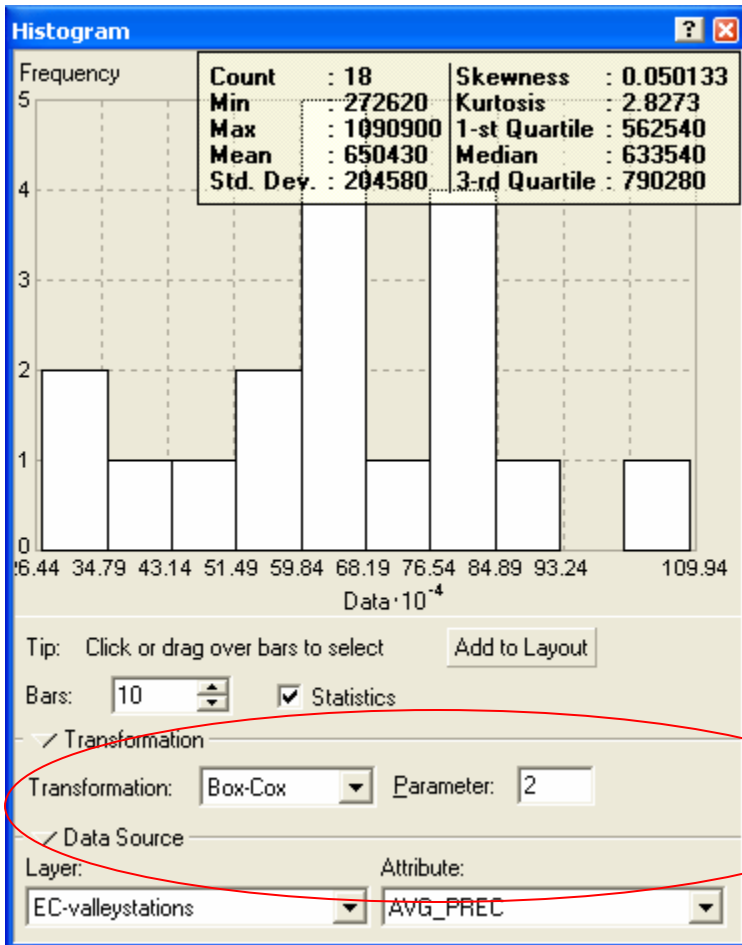


A Normal Distribution of your data will be apparent if the:

- Mean and median are similar
- Skewness should be near zero
- Kurtosis should be near 3

- If data is lightly skewed you may chose to transform it to see if you can make it a normal distrubution

The following transformations were applied (red circle) to the average precipitation data in order to achieve a normal distribution. (The next page shows the model after the transformations were applied)



Average Annual Precipitation

● EC-valleystations

Ordinary Kriging

Prediction Map

[EC-valleystations].[AVG_PREC]

Hillshade

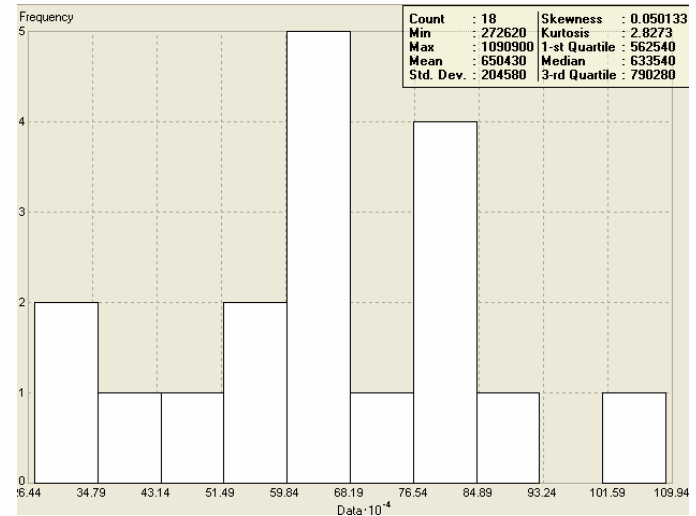
Filled Contours



N



0 5 10 20 30 40 kilometers



Average Annual Precipitation

● EC-valleystations

Ordinary Kriging_2

Prediction Standard Error Map

[EC-valleystations].[AVG_PREC]

Hillshade

Filled Contours

174.647827 - 179.808273

179.808273 - 183.999725

183.999725 - 187.404144

187.404144 - 191.595596

191.595596 - 196.756042

196.756042 - 203.109497

203.109497 - 210.931763

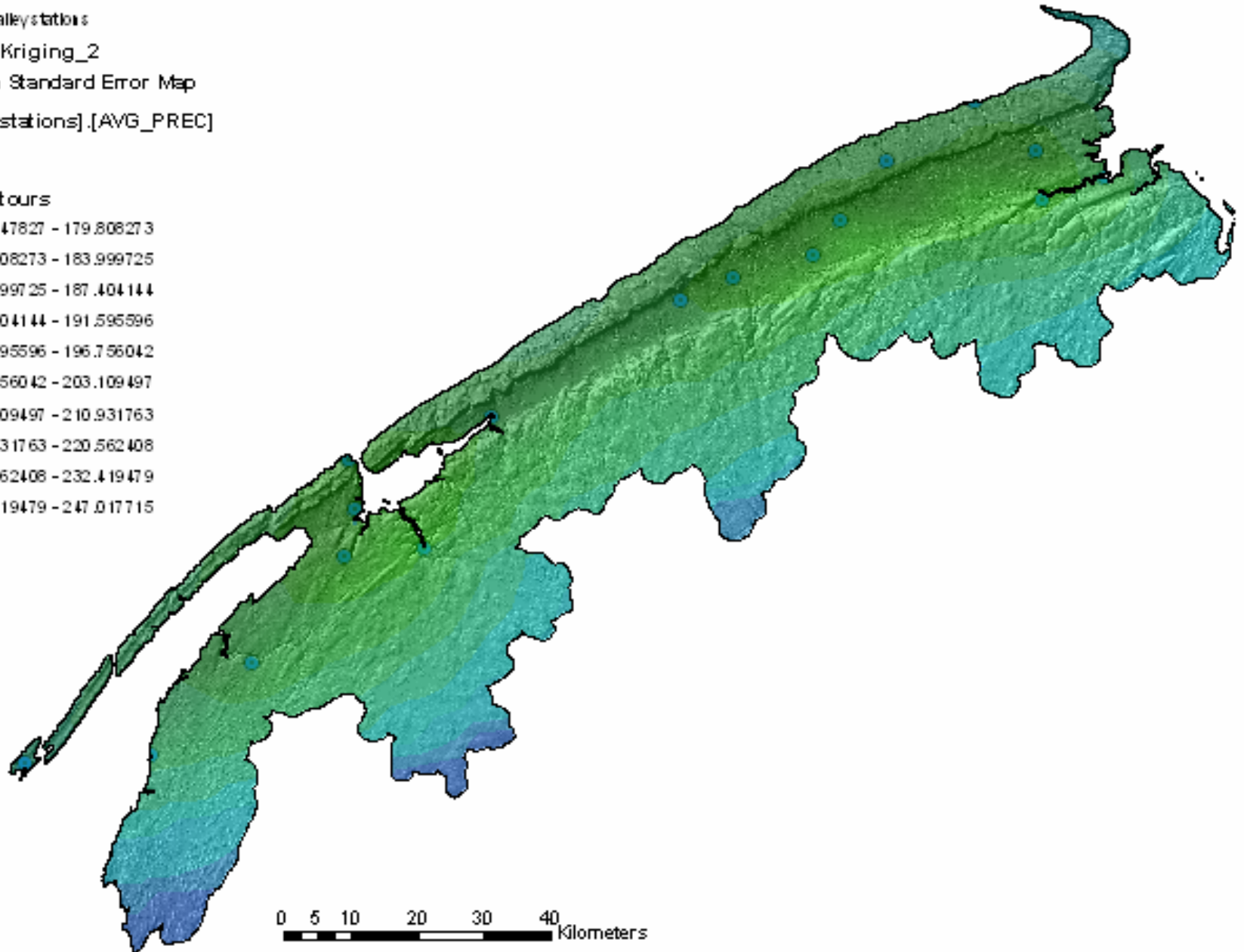
210.931763 - 220.562408

220.562408 - 232.419479

232.419479 - 247.017715



0 5 10 20 30 40 Kilometers



- There are many tools with the Geostatistical Analyst extension that can assist with the production of a geostatistical prediction model
- Geostatistical analyst contains a number of unique tools to help improve prediction, including tools for transformation, interpolation, error modeling and much more. Being new to this extension has shown me that you do not need to know everything about the analyst to get immediate results, but a thorough understanding of statistics would be beneficial to ensure that your model will spatially represent your data.

- Further work learning this extension would prove beneficial

References

- www.esri.com
- ArcGIS 9 – Using Geostatistical Analyst