

Perform Kriging interpolation using Geostatistical Analyst wizard of ESRI-ArcGIS

Following tutorial document was found at Department of Civil, Architectural and Environmental Engineering, University of Texas, Austin.

Geostatistical Analyst Steps for Kriging Interpolation

Prepared by Parikshit Ranade, Dr Ayse Irmak and David R. Maidment

Spatial Interpolation Methods

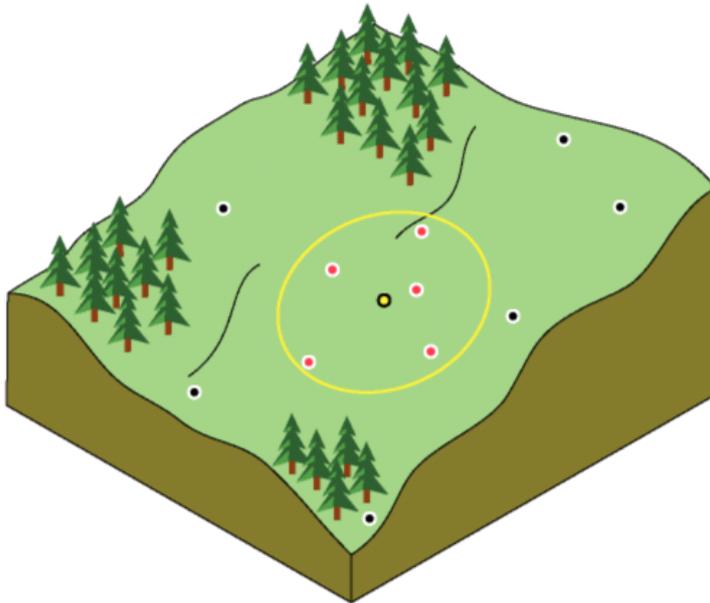


Figure 1 : The interpolated value at the unmeasured yellow point is a function of the neighboring red points (From ArcGIS Help Menu).

A very basic problem in spatial analysis is interpolating a spatially continuous variable from point samples. Many spatially explicit hydrologic/watershed models require continuous surfaces of temperature. Three commonly used interpolation methods to model spatially distribution from point data are Inverse Distance Weighting (IDW), spline and ordinary kriging.

The IDW is simple and intuitive deterministic interpolation method based on principle that sample values closer to the prediction location have more influence on prediction value than sample values farther apart. Using higher power assigns more weight to closer points resulting in less smoother surface. On the other hand, lower power assigns low weight to closer points resulting in smoother surface. We optimized power parameter using ArcGIS. Major disadvantage of IDW is “bull's eye” effect (higher values near observed location) and edgy surface. Spline is deterministic interpolation method which fits mathematical function through input data to create smooth surface. Spline can generate sufficiently accurate surfaces from only a few sampled points and they retain small features (Anderson, 2008). Spline works best for gently varying surfaces like temperature. In ArcGIS Spline is Radial Basis Function.

Unlike IDW and spline, kriging is method based on spatial autocorrelation. It uses semivariogram.

Basics of Kriging

Kriging was developed in the 1960s by the French mathematician Georges Matheron. The motivating application was to estimate gold deposited in a rock from a few random core samples. Kriging has since found its way into the earth sciences and other disciplines. It is an improvement over inverse distance weighting because prediction estimates tend to be less biased and because predictions are accompanied by prediction standard errors (quantification of the uncertainty in the predicted value).

The basic tool of geostatistics and kriging is the semivariogram. The semivariogram captures the spatial dependence between samples by plotting semivariance against separation distance (semivariance will be explained in the next paragraph). The premise of any spatial interpolation is that close samples tend to be more similar than distant samples (this is also called spatial autocorrelation). This property of spatial data is implicitly used in IDW. In kriging, one must model the spatial autocorrelation using a semivariogram instead of assuming a direct, linear relationship with separation distance.

Semivariance equals one-half the squared difference between points separated by a distance $d \pm \Delta d$ (assuming no direction preference). As the distance between samples increases, we expect the semivariance to also increase (again, because near samples are more similar than distant samples). This is true, however, only up to some given separation distance. For this distance and up, points are unrelated. Stated another way, if 50m is this critical separation distance, two points separated by 50m are likely to be just as similar (or dependent on one another) as samples separated by 100, 200, 300, or any distance greater than 50m.

Suppose we have the semivariogram shown in Figure 2. What information does the plot provide? Well, the semivariance between samples separated by no distance is about $1.5E-4$. This is called the nugget. What it says is that if you measure the variable at locations very, very close to one another, the values measured might be quite different. Why would this happen? Suppose you had a gold nugget in the middle of an otherwise gold-free rock. If you sample just on the edge of the nugget you get a high gold estimate. If you sample just outside of the edge, you get no gold in your estimate. The presence of a nugget in the semivariogram therefore tells you that, assuming no measurement error, the variable is not spatially continuous.

The semivariogram also tells us that points separated by 60,000 m are likely to have the same average difference as points separated by 100,000, 150,000, 200,000 m or any distance above 60,000m. 60,000 m is the range of the semivariogram and suggests the area of influence for any given point. An unmeasured location can be predicted based on its neighboring samples closer than 60,000m. A sample collected 61,000 m away from the sample will likely have no influence on the actual value at the unmeasured location.

When you look at the model of a semivariogram, you'll notice that at a certain distance, the model levels out. The distance where the model first flattens out is known as the range. Sample locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not. The value that the semivariogram model attains at the range (the value on the y-axis) is called the sill. The partial sill is the sill minus the nugget.

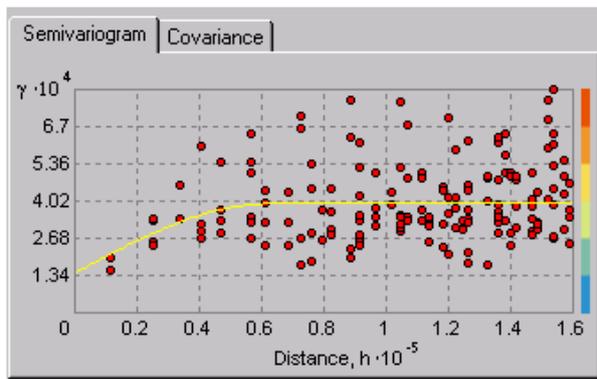


Figure 2 : The semivariogram is used to model the spatial relationships between samples separated by some distance, d

For kriging estimation, the semivariogram model (the yellow line in figure 2) is used to obtain estimates for the weighting parameters of Equation 1. This process is done automatically by the geostatistical analyst once the user is satisfied with the semivariogram. If you are interested in the derivation of the weighting parameters (or any of the other topics discussed here), [Applied Geostatistics](#) by Edward H. Isaaks and R. Mohan Srivastava is an excellent resource. Or for the more mathematical folks, try [Statistics for Spatial Data](#) by Noel A.C. Cressie.

Case study

Now we know the basics of spatial interpolation. Lets use our knowledge to estimate mean annual air temperature for each county of Nebraska.

Study Area

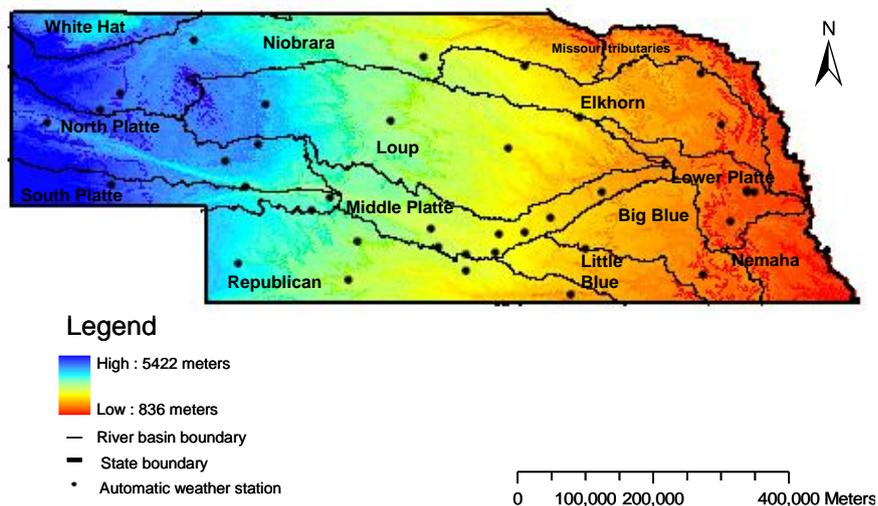


Figure 3 : Location of major river basins in Nebraska.

Nebraska covers a total of 124496 square kilometers area, making it the 16th largest of the 50 states in United States of America. It is the mid-western state between longitude coordinates $95^{\circ}25'W$ and $104^{\circ}W$ (~690km) and latitude coordinates $40^{\circ}N$ and $43^{\circ}N$ (~340 km). The geographic center of the state is located in Custer County with a longitude: $99^{\circ} 51.7'W$; and latitude of $41^{\circ} 31.5'N$ (Figure 1). The state comprises of UTM zones 13, 14 and 15. The highest point is Panorama Point, at 1653 meters

above sea level and the lowest point is 256 meters above sea level at the Missouri River in southeastern Richardson County. The Mean Elevation of the state is 792 meters above sea level. The major basins in the state are Missouri, Niobrara, Platte, and Republican River. State has 93 counties. In this case we have used 215 NWS weather stations in and around Nebraska to model spatial variation of mean annual temperature.

Data Download

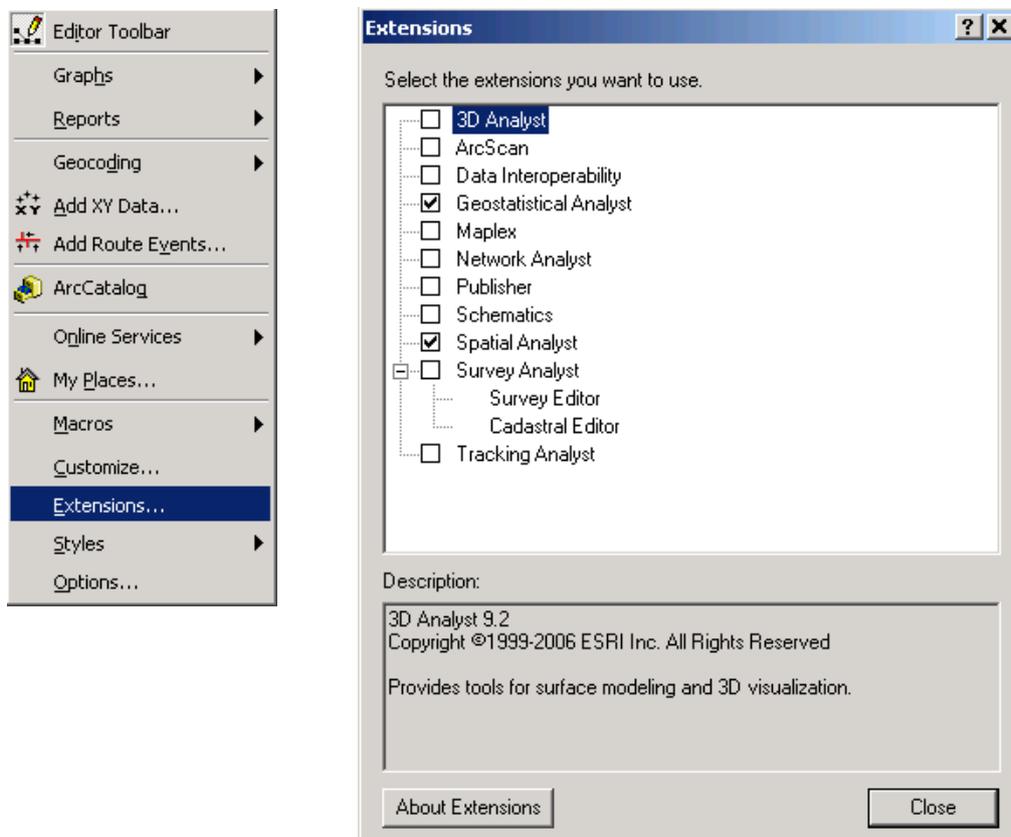
Download the <http://www.ce.utexas.edu/prof/maidment/giswr2008/geostat/Nebraska.zip> folder on your computer. You will see the following subfolders in the folder.

1. Tmean – This is a shpfile of National Weather service and co-operative observer networks weather stations (NWS). The attributes of the file are mean monthly and mean annual air temperature for each station.
2. NE_Boundary – This is shapefile for Nebraska border.
3. NE_County – This is a shpfile of all the counties in Nebraska.
4. NE_Cities – This is shapefile of all the cities in Nebraska.

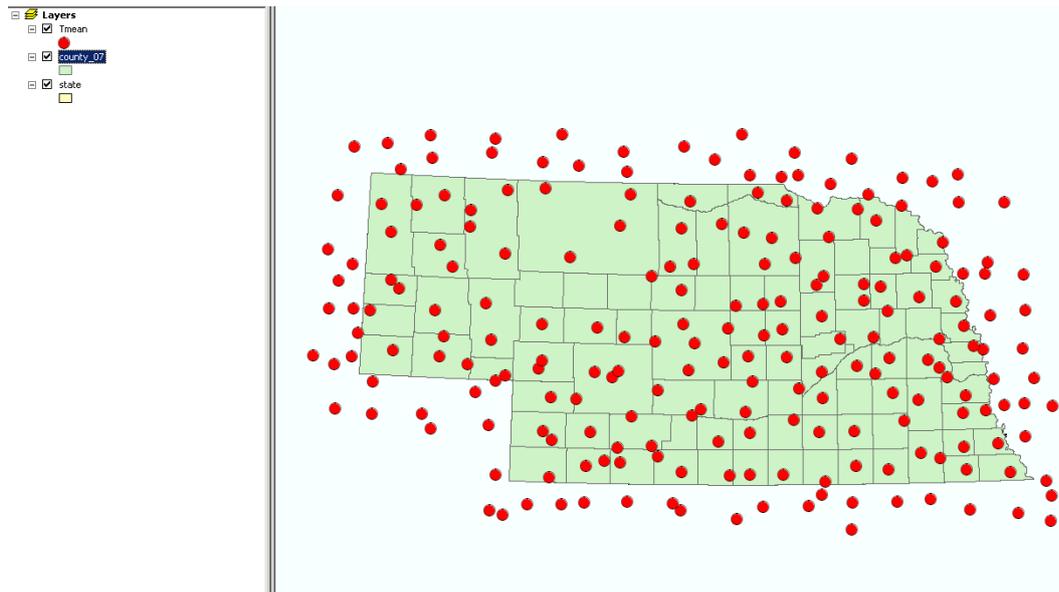
Getting Started

In this exercise we need to use spatial analyst and geospatial analyst toolbar. We need to enable this extension in the first place. Open **Tools** → **Extension**.

Check the ‘Spatial analyst’ and ‘Geospatial analyst’ box. Now we can use all the functions in both toolbars.



Add **State** shapefile from **NE_Boundary** folder to your computer. Add the **Tmean** file to your map. Add **County** shapefile. Following view will appear on your screen.

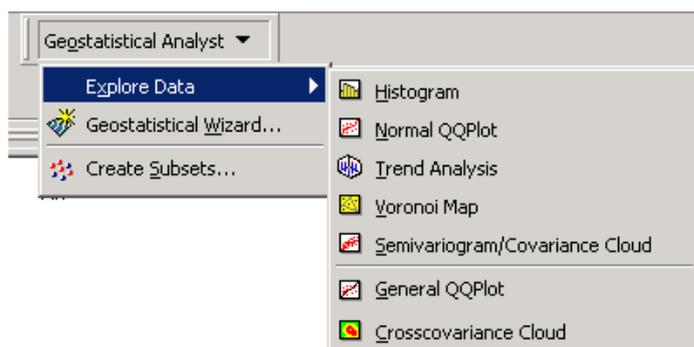


Exploratory Spatial Data Analysis

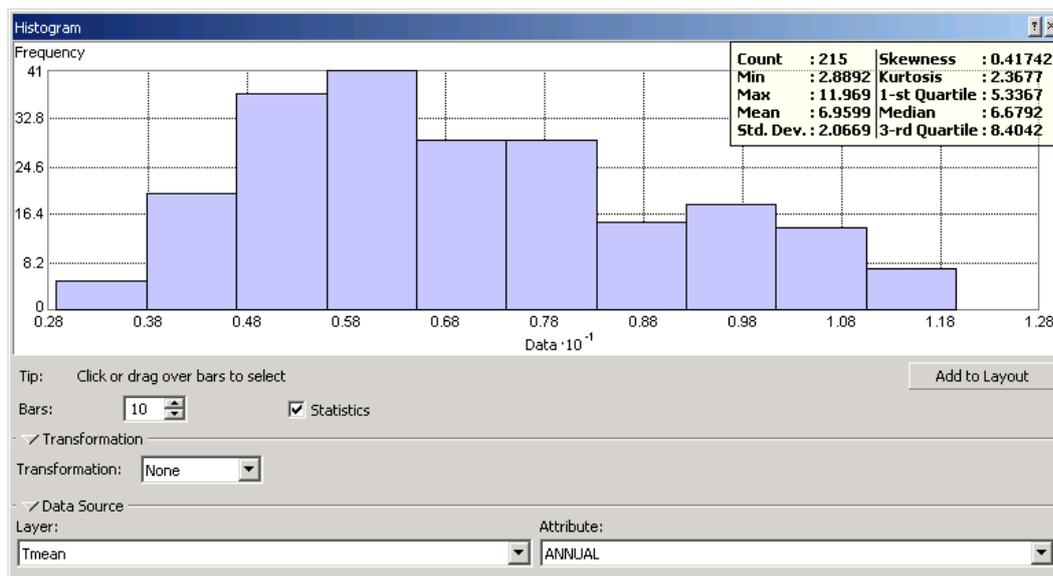
Exploratory Spatial Data Analysis (ESDA) is a process of understanding the properties of a spatial dataset in order to best model the data using geostatistics. The word “Explore” should tell you that ESDA is more of an adventure than a strict – you must follow the path at all times – procedure. In this exercise we show how the Geostatistical Analyst tools can be used to understand the population distribution of the attribute of interest and how to understand the large-scale patterns in the dataset through 3D visualization. This is not a comprehensive list of ESDA procedures, but a good start. Through this process, keep in mind that the better one understands the spatial characteristics of the data, the better kriging model one can build to interpolate the data, and, consequently, the better estimates one will produce.

Histogram

Open **Geospatial analyst** → **Explore Data** → **Histogram**.



Select the 10 bars and check the statistics option to view the statistics of mean annual air temperature. Keep the **transformation** as **None**. Select **layer** as **Tmean** and **Attribute** as **ANNUAL**. Similar window will appear on screen.

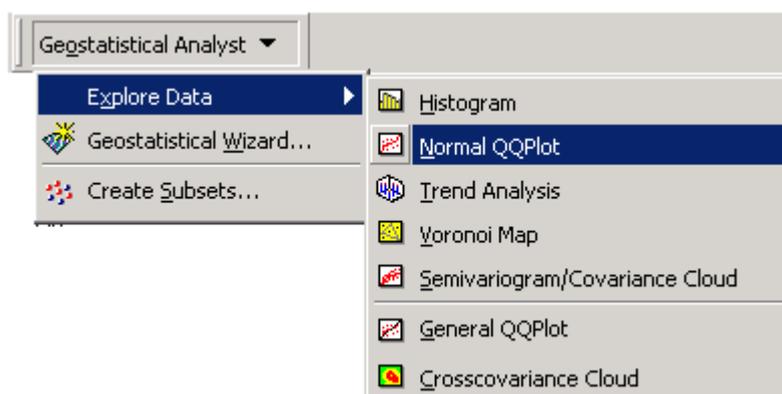


We will now analyze the histogram for mean annual air temperature of Nebraska. Upper right corner shows the statistics of the statistics of mean annual air temperature. Histogram shows that our data is not perfectly normally distributed, but it is fairly normally distributed. One of the crosscheck of normal distribution of data is that mean should be closer to the median. In our case mean is 6.9°C and median is 6.6°C. We can consider our data as normally distributed. User can transform the data into **log** or **box-cox** distribution if it is not normally distributed.

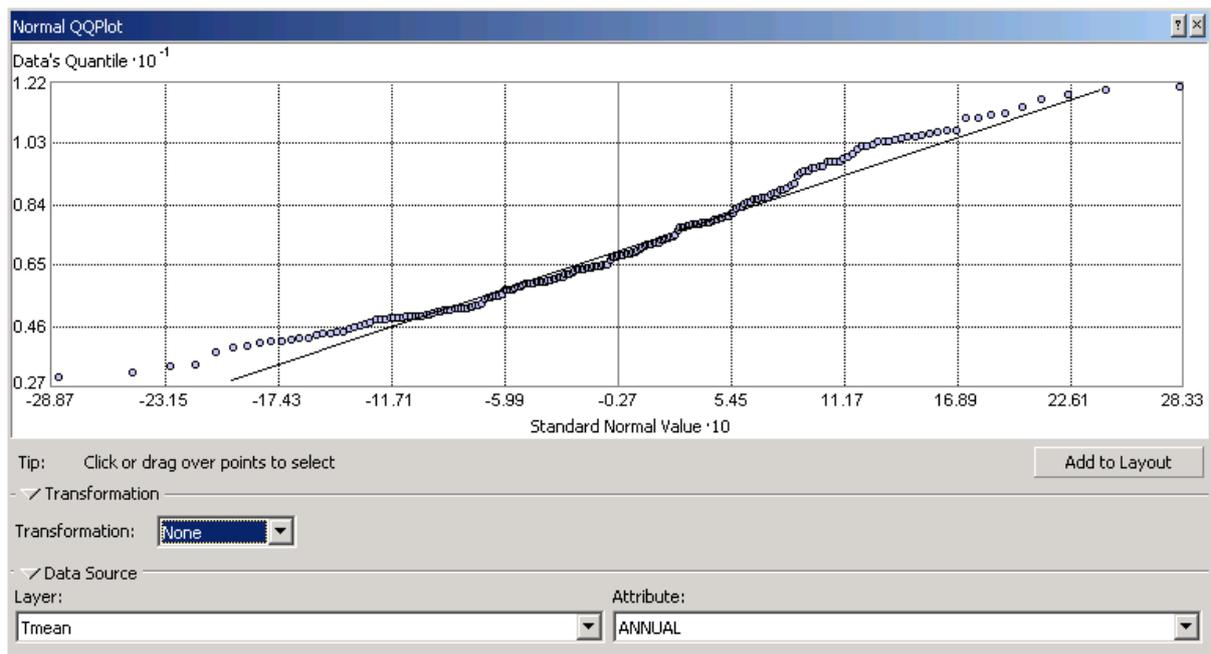
QQ (Quantile-Quantile) plot

Another way to understand the data's distribution is by using the Normal QQ Plot tool. In a QQ (Quantile-Quantile) plot we test whether data are normally distributed by plotting it against a dataset with a known normal distribution. If the plot is linear along the line $Y=X$, then the data follow a normal distribution.

Open **Geospatial analyst** → **Explore data** → **Normal QQ plot**.



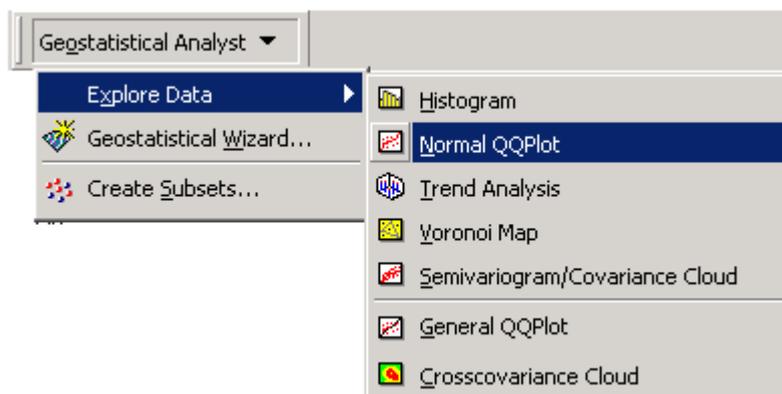
The QQ plot shows the linear relationship between $\log(\text{Tmean})$ and the standard normal distribution. Data is not exactly linear on lower end. But we can say that Tmean is normally distributed.



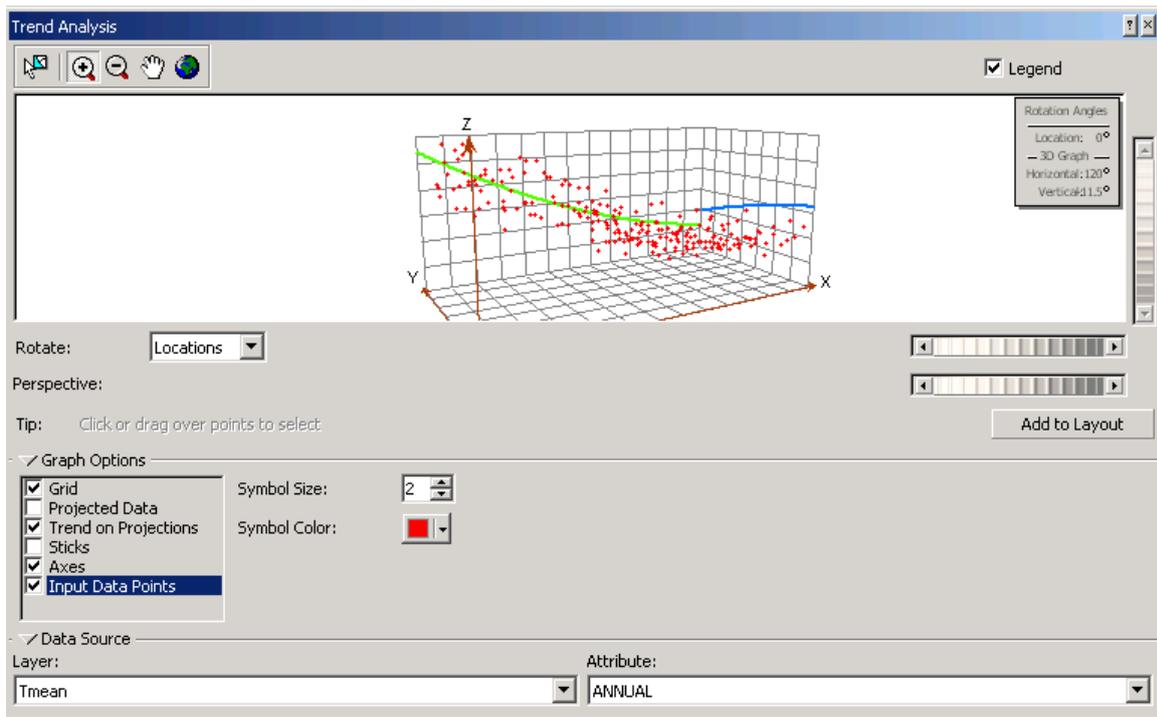
Trend analysis

The trend analysis tool provides a 3D plot of the samples and a regression on the attribute in the XZ and YZ planes. The purpose of the tool is to visualize the data and to observe any large-scale trends that the modeler might want to remove prior to estimation. It is best to keep the kriging model as simple as possible and to only remove a trend if it significantly improves prediction errors.

Open **Geospatial analyst** → **Explore data** → **Trend analysis**.



Screen will look similar to figure We can choose the different graph options and rotate the location to see the trend in the data.



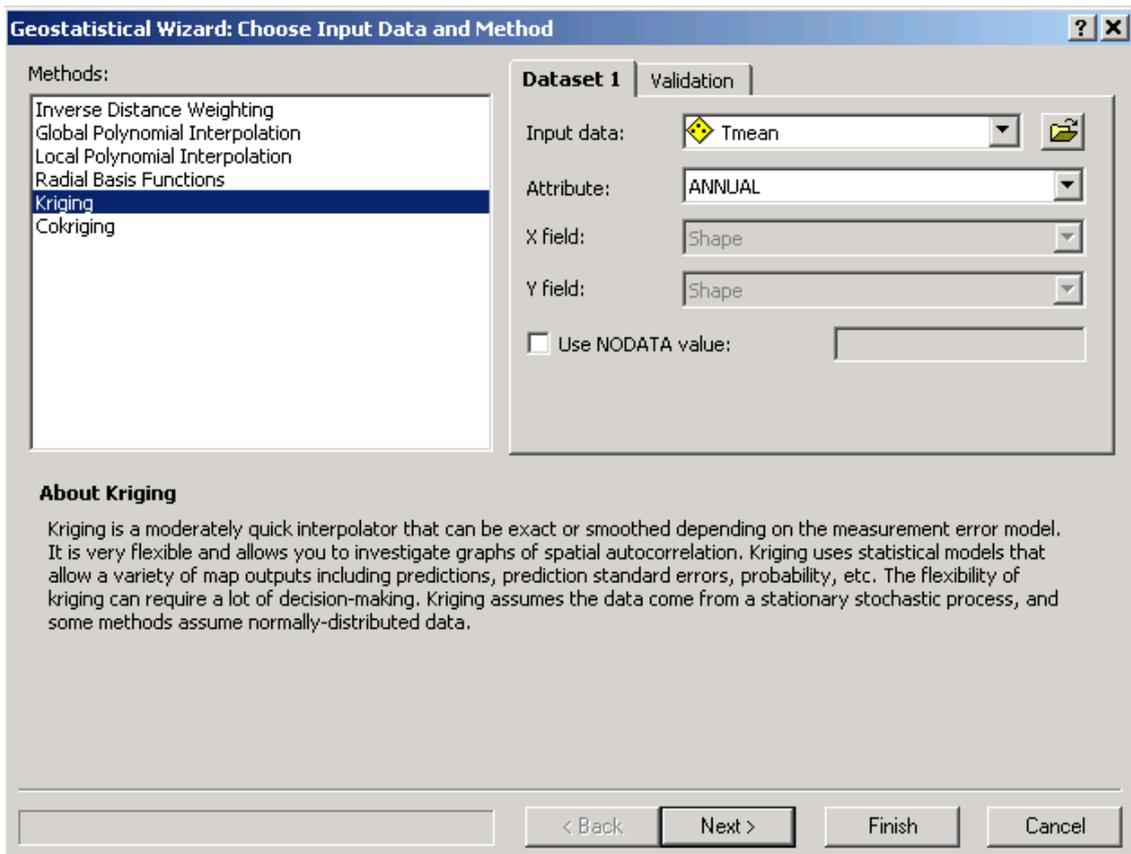
The trend analysis tool provides a 3D plot of the samples and a regression on the attribute in the XZ and YZ planes. The purpose of the tool is to visualize the data and to observe any large-scale trends that the modeler might want to remove prior to estimation. It is best to keep the kriging model as simple as possible and to only remove a trend if it significantly improves prediction errors.

Perform Kriging interpolation

Open Geospatial analyst → Geospatial Wizard

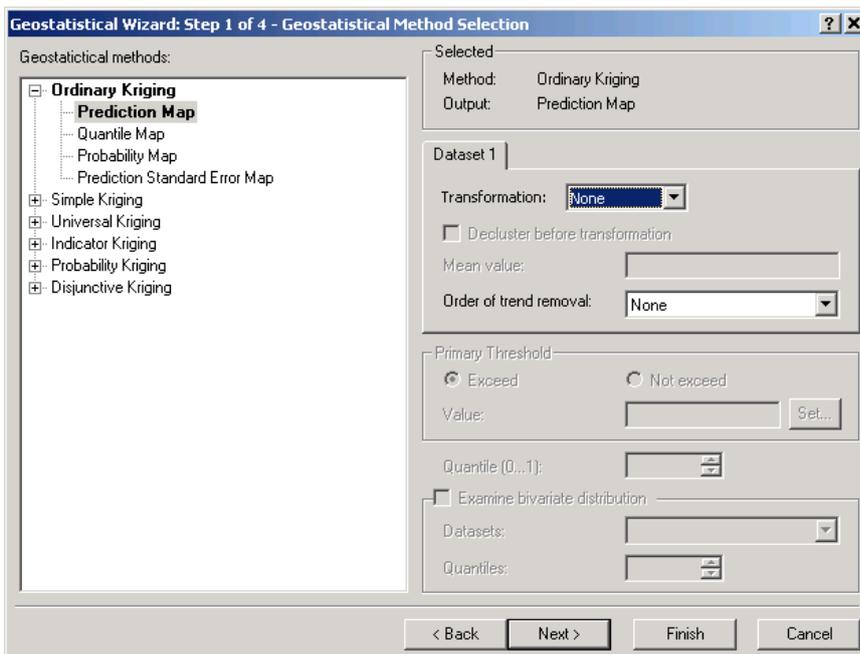


Select the **Kriging** as **method** and **Tmean** as **input data**. Select **attribute** as **ANNUAL**.



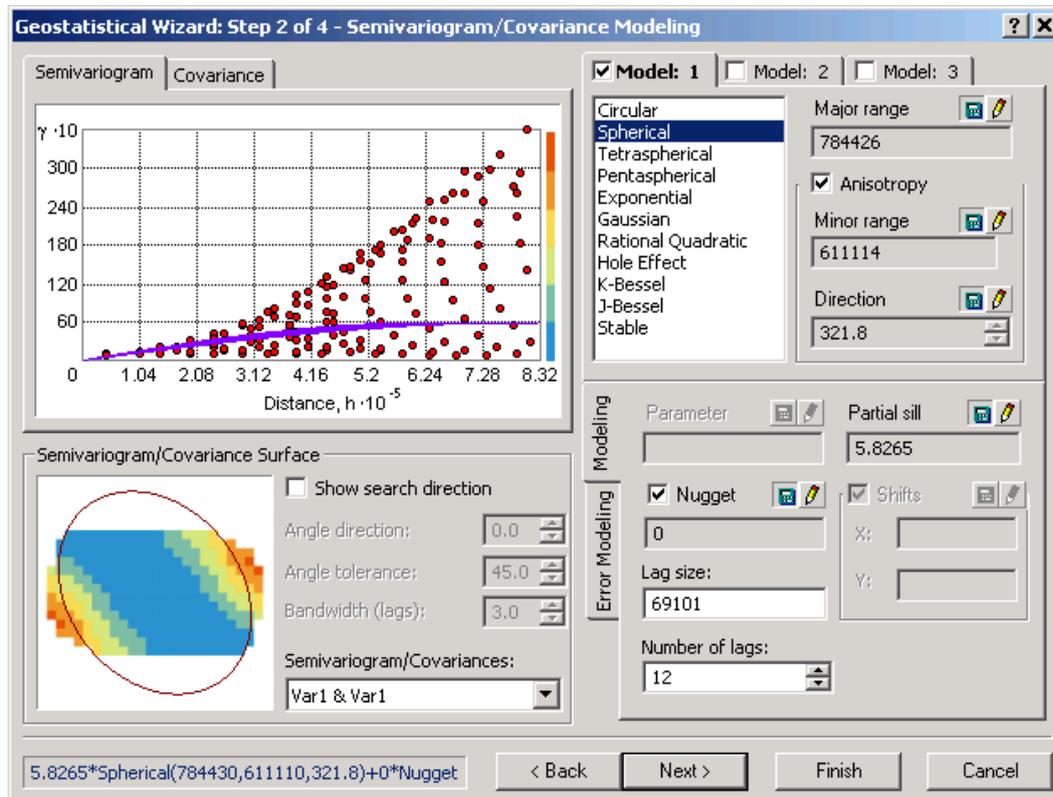
User can assign the **no data value** as zero or any other desired value. Default no data value is -9999.0.

Click **Next** to proceed.



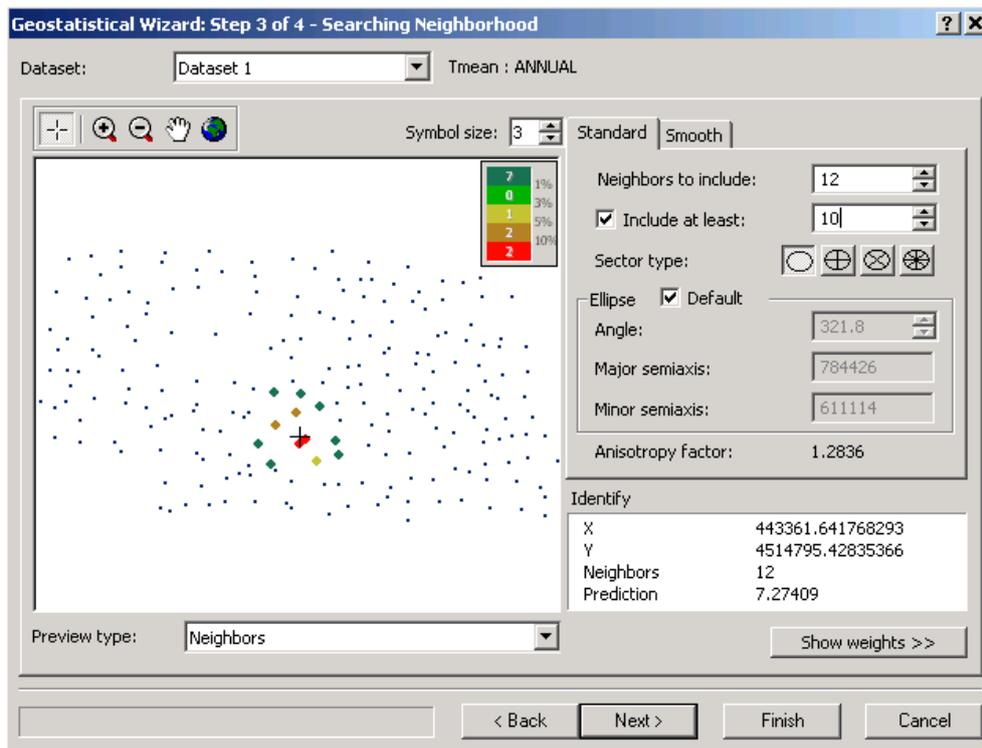
We will choose the most widely used ordinary kriging method and prediction map. As discussed in exploratory spatial data analysis we will keep transformation and order of trend removal as none.

Click Next to proceed.



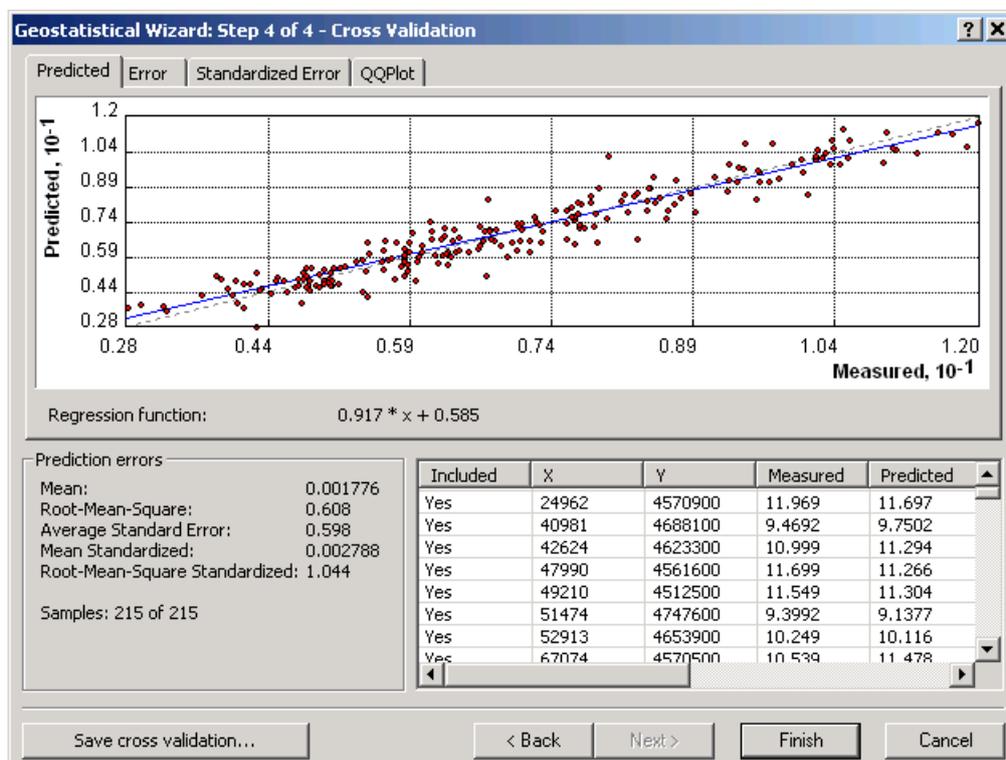
We will choose widely used **spherical** semivariogram model. Other commonly used models are exponential and Gaussian. Lag size and number of lags are the important parameters to be selected. These two parameters are used to group the number of pairs of data. For our analysis we will use the default values. For given lag size and number of lags, ArcGIS automatically calculates the nugget, range and sill. Check the **anisotropy** option. Anisotropy represents the trend in the data in particular direction. We can see that blue value have high covariance and they are oriented in one direction. Semivariogram for particular direction can be investigated using **show search direction** tool.

Click Next to proceed.



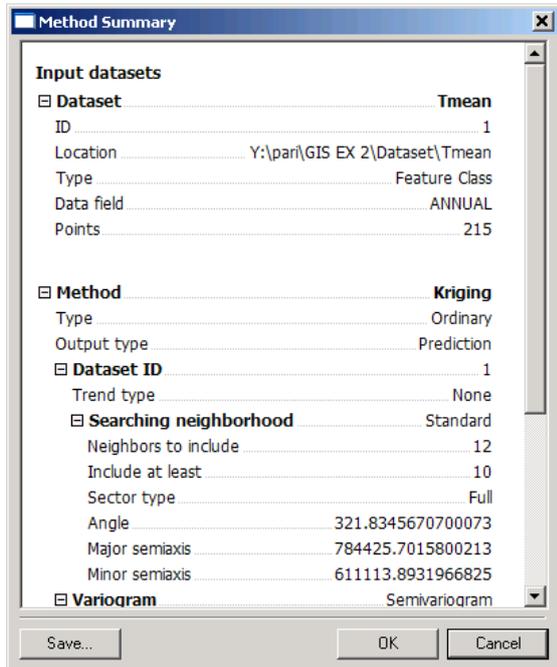
Change **Neighbours to include** to **12** and include at least **10** stations. Since we have the total **215** weather stations, 12 and minimum 10 weather stations is an appropriate number for interpolation. Click on the first circle for **sector type**.

Click **Next** to proceed.



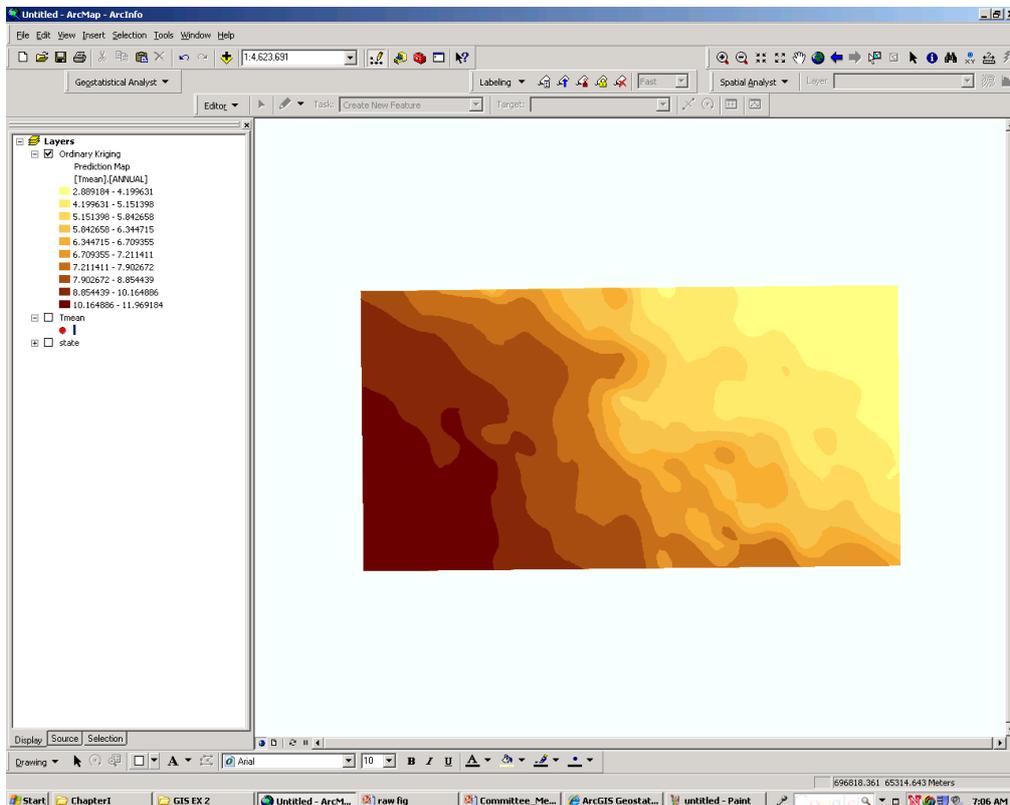
This is the error statistics for the interpolation. This is based on one-leave-out cross validation method. We can save the cross validation table to perform further statistical analysis.

Click **Finish**.

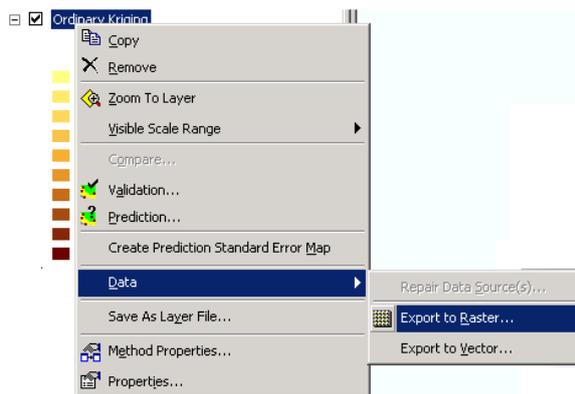


This is the summary of interpolation parameters chosen. Click **Ok**.

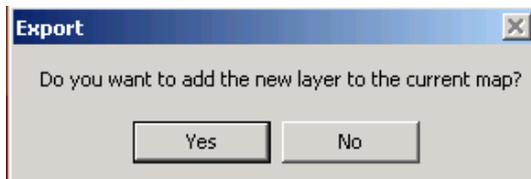
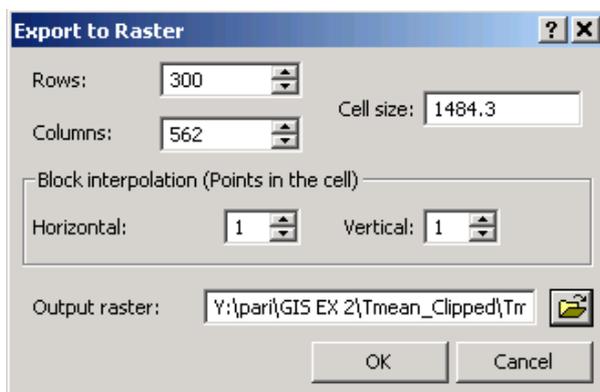
Now you will see the following map of mean annual air temperature.



We will save this as a **raster** dataset. **Right click** on the **Ordinary Kriging** layer. Select **Data** → **Export to raster**.



Select output raster as folder and file name as **Tmean_Clipped**

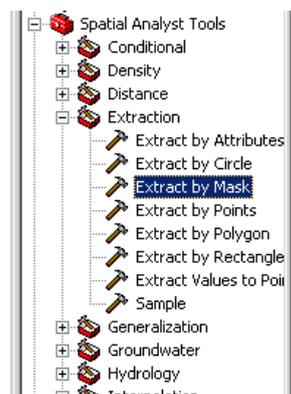


Add this layer to the map.

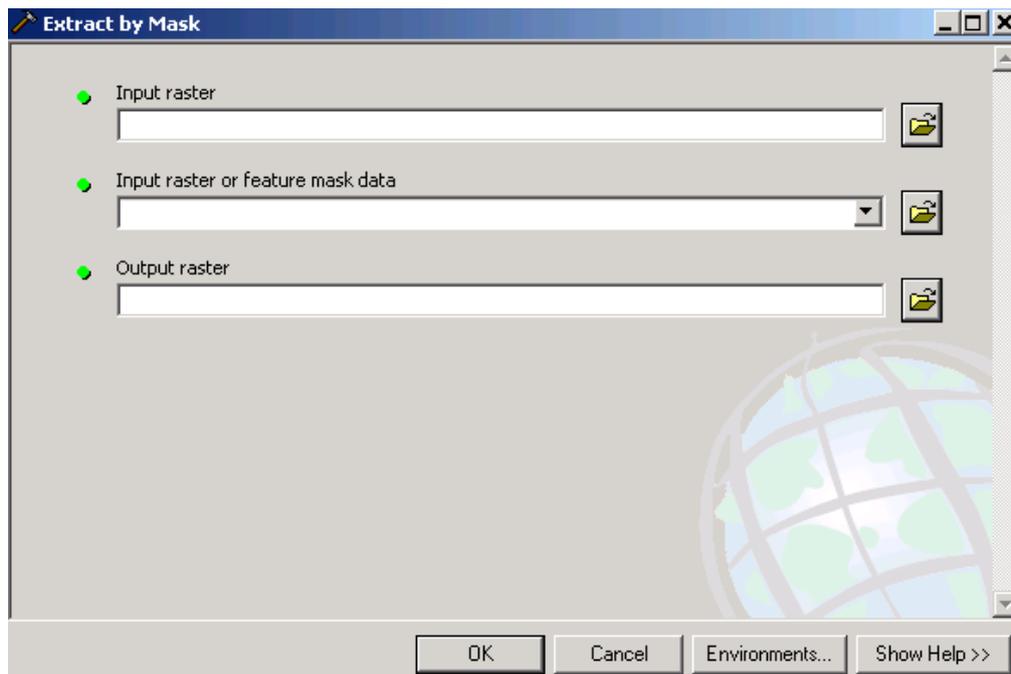
Raster extraction

We will clip this map to the Nebraska border. Open

ArcToolBox → **Spatial analyst tools** → **Extraction** → **Extract by mask**.



Select input raster as **Tmean_Clipped** raster or feature mask data as **State**. Save it on the desired output location with the name **NE_Tmean**.

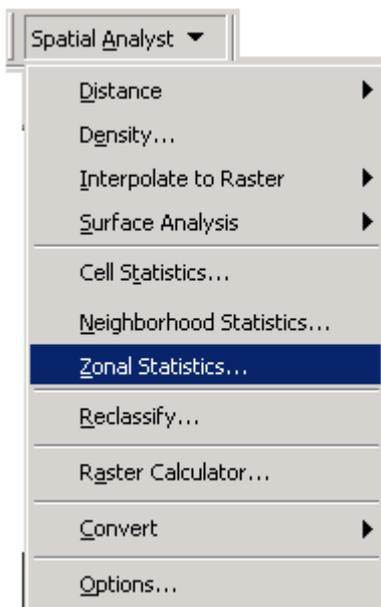


Add this clipped layer to the map.

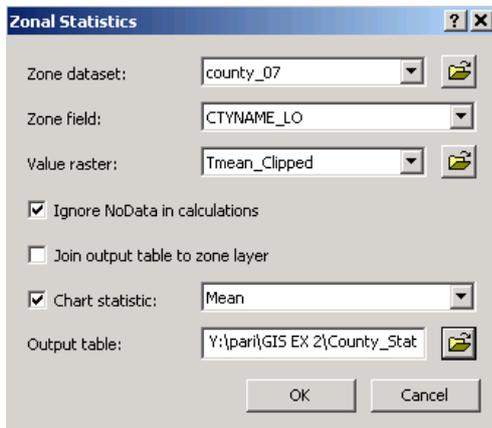
Zonal Statistics

We will compute zonal statistics to know the average rainfall for each county.

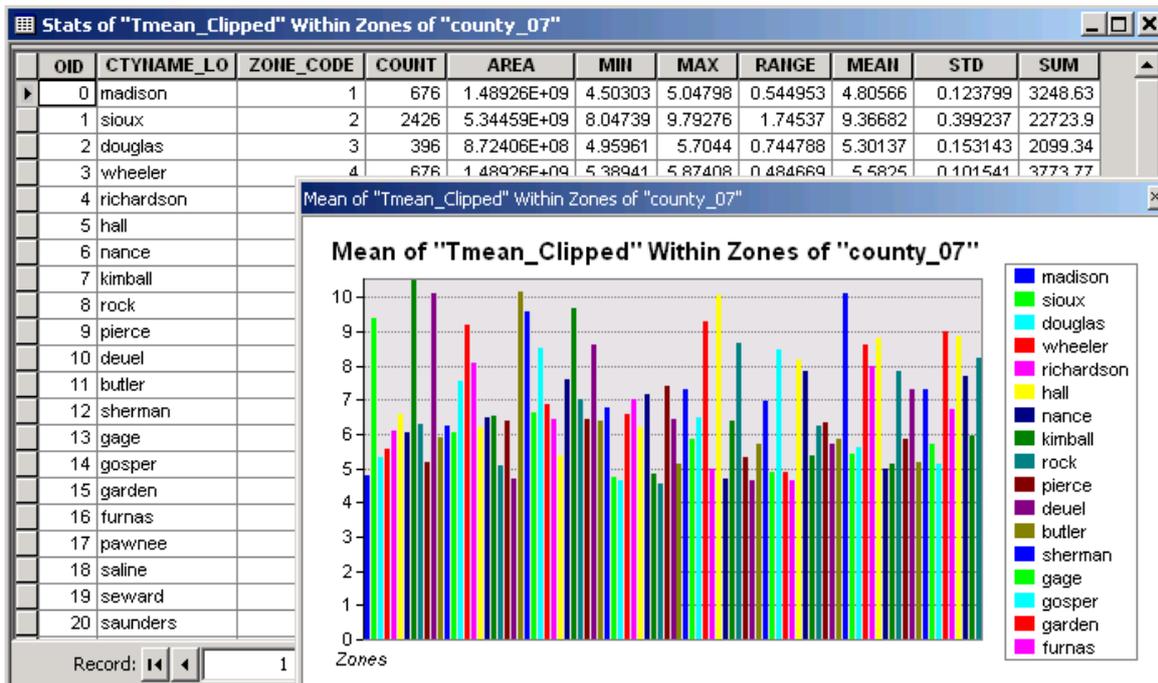
Select **spatial analyst** → **Zonal statistics**



Select **zone dataset** as **County_07**. Select **Zone file** as **CTYNAME_LO** and **value raster** as **NE_Tmean**. Ignore NoData in calculations and select **chart statistics** as **Mean**. Select the output table in desired folder.



Following chart should appear on the screen.



This shows the mean annual temperature for each county.

References:

[1] Understanding Interpolation Analysis, ESRI.

<http://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/understanding-interpolation-analysis.htm>

[2] Comparing Interpolation Methods, ESRI.

<http://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/comparing-interpolation-methods.htm>

[3] Agung Setianto and Tamia Triandini, 2013, Comparison of Kriging And Inverse Distance Weighted (IDW) Interpolation Methods in Lineament Extraction and Analysis. J. SE Asian Appl. Geol., Jan–Jun 2013, Vol. 5(1), pp. 21–29

[4] Parikshit Ranade, Dr Ayse Irmak and David R. Maidment, 2008, Geostatistical Analyst,

<https://www.cae.utexas.edu/prof/maidment/giswr2008/geostat/ExGeostat.doc>