# Statistical Analysis

1. Introduction

Whether you are concerned with the amount of rainfall, concentrations of pollution, or the differences in elevation, it is impossible to measure these phenomena at every point within a geographic area. You can, however, obtain a sample of measurements from various locations within the study area, then, using those samples, make inferences about the entire geographic area.

Interpolation is the process of estimating unknown values that fall between known values. The primary assumption of spatial interpolation is that points near each other are more alike than those farther away; therefore, any location's values should be estimated based on the values of points nearby.
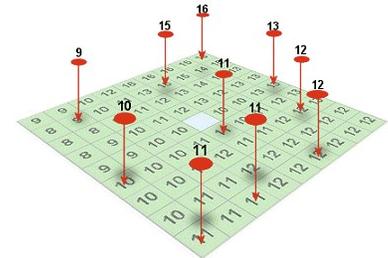
Interpolating the sample points' values creates a surface. With all of the cells, unknown value of the cell in the center will be estimated based on values of the surrounding sample points.

With spatial interpolation, you can create an entire surface from just a small number of sample points; however, more sample points are better if you want a detailed surface.

Type of interpolation

IDW (Inverse distance weighted interpolation)
IDW determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The surface being interpolated should be that of a locationally dependent variable. IDW allows you to control the significance of known points upon the interpolated values, based upon their distance from the output point. The interpolation can be shifted from local to global by changing the {power} option. A larger power will result in less influence from surrounding points, i.e., nearby data will have the most influence, and the surface will have more detail (be less smooth). The power is a positive, real number. A common value is 2.

SPLINE
A spline is a two-dimensional minimum curvature interpolation resulting in a smooth surface that passes exactly through the input points. The basic minimum-curvature technique is also referred to as thin plate interpolation. It ensures a smooth (continuous and differentiable) surface together with continuous first-derivative surfaces.

The SPLINE function has two options, REGULARIZED and TENSION. These options, combined with a weight function are used to control the curvature of the output surface. The REGULARIZED option yields a smooth surface and smooth first derivatives. The TENSION option tunes the stiffness of the interplant according to the character of the modelled phenomenon.

KRIGING
Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z values. Unlike the other interpolation methods supported, KRIGING involves an interactive investigation of the spatial behavior of the phenomenon represented by the z values before the user selects the best estimation method for generating the output surface.
Kriging is based on the regionalized variable theory that assumes that the spatial variation in the phenomenon represented by the z values is statistically homogeneous throughout the surface. This hypothesis of spatial homogeneity is fundamental to the regionalized variable theory.

2. Data Preparation

Summary on data

The climate data were gathered from [Canadian Climate and Water Information](). These data contain information on Temperature, Rainfall, and Precipitation from 10 northern climate monitoring stations (Upper Fraser River Basin) from 1971 to 1990. The data were extracted for every 5 years on Mean Temperature, Maximum Temperature, Minimum Temperature and Total Rainfall, Total Snowfall and Total Precipitation for each station. Missing values for some stations on some years are replaced by overall mean value. These data are stored in the table **climate_dat**. The station number (**station_no**) was included in the file to provide a link to the spatial data (*station*).

Datasets

All the spatial and non-spatial datasets required for this lab are located at L:\labs\geog413\data\climate

Datasets	Description
station	Weather Station locations
study_bndy	Study boundary
roads	Major roads
rivers	Major rivers
climate_data (table)	Climate data

Set Analysis Environment

- Start ArcMap and add all data listed above.
- Set working environment as the followings.

Current / Scratch Workspace	your local folder
Processing Extent:	*study_bndy*
Raster Cell Size:	100
Analysis Mask:	*study_bndy*

3. Interpolating Surface

The surface functions create a raster surface from the point locations with z-values stored in the attribute table. The Z value could be any type of information such as elevation, population, temperature, pollutions etc. The functions use several different interpolation methods and allow for weights to be applied to the point locations.

Function SPLINE will be used for creating a raster surface with *station* dataset as the point locations

Generate a surface showing the mean temperature of the study region

You will generate a mean temperature surface using the station point locations (**station**) and the mean temperature in the attribute table **climate_data**.

- Join the table **climate_data** to *station* attribute table using data field **station_no**
- Examine the attributes of *station*. You should have information on temperature, rainfall, snowfall, precipitation etc. from 1971 to 1990.

To test the interpolation methods, two different surfaces will be generated.

First, use the SPLINE function with REGULARIZED option. Use the data field **MEAN71_75** (the mean temperature for 1971 - 1975) for the Z Value field. Name the output raster ***mean_75reg***

The tools used for interpolation are located in **ArcToolbox->Spatial Analyst->Interpolation**

SPLINE function interpolates a surface from a point feature class using a minimum curvature spline technique.

ArcToolbox ->Spatial Analyst -> Interpolation->Spline
      Input points: ***station***
      Value field: **mean_71_75**
      Spline Type: **Regularized**
      Output Raster: ***men_75reg***

- Symbolize ***mean_75ten*** to 10 classes with break value at 1, 2, 3, 4, 5, 6, 7, 8, 9, 30. Select a proper color ramp. Save the symbology as a layer file ***mean.lyr***

The ***mean_75reg*** is a raster surface showing the mean temperature from 1971 - 1975. Each pixel value indicates the mean temperature at the location.

- Repeat the same steps above but set **Tension** for Spline Type to create a surface ***mean_75ten***

The REGULARIZED option of Spline Type usually produces smoother surfaces than those created with the TENSION option. With the REGULARIZED option, higher values used for the weight parameter produce smoother surfaces. The values entered for this parameter must be equal to or greater than zero. Typical values used are 0, 0.001, 0.01, 0.1, and 0.5.

With the TENSION option, higher values entered for the weight parameter result in somewhat coarser surfaces, but surfaces that closely conform to the control points. The values entered must be equal to or greater than zero. Typical values are 0, 1, 5, and 10. The Weight is the square of the parameter referred to in the literature as pi (π).

The greater the value of Number of Points is, the smoother the surface of the output raster is.

Use SPLINE function with REGULARIZED option for the rest of surface creation

*Follow the same procedure generate a surface on mean temperature for each five-year using the SPLINE function with REGULARIZED option. Save the output raster as **mean_75reg**, **mean_80reg, mean_85reg, mean_90reg** respectively. Symbolize each surface using the **mean.lyr** file. Compare these layers and discuss the changes.*

---

4. Statistical Functions

The statistical functions of raster allowing for statistical calculations across two or more raster datasets include:

Functions	Description
MEAN	uses multiple input rasters to determine the mean value on a cell-by-cell basis
MAX	uses multiple input rasters to determine the maximum value on a cell-by-cell basis.
MIN	uses multiple input rasters to determine the minimum value on a cell-by-cell basis.
RANGE	uses multiple input rasters to determine the range of values on a cell-by-cell basis.
STD	uses multiple input rasters to determine the standard deviation on a cell-by-cell basis.
SUM	calculates, on a cell by cell basis, the sum of a set of input raster

Generate a raster surface showing the overall mean temperature from 1971 to 1990 for the entire study region.
      ArcToolbox ->Spatial Analyst ->Local->Cell Statistics

      Input Raster: ***mean_75reg, mean_80reg, mean_85reg, mean90reg***
      Overlay Statistics: **MEAN**
      Output raster: ***overall_mean***

- Check the properties of the layer for highest/lowest mean temperature

*With MAX, MIN, RANGE functions, generate a raster surface showing the overall maximum, minimum and range temperature from 1971 to 1990 for the study region respectively. Name the output as **overall_max, overall_min, overall_range**. Symbolize each surface in 10 classes and natural break method. Produce a map to show four*

*surfaces (**overall_mean, overall_max, overall_min, overall_range**) on one page with an appropriate tile, legend, scale*

overall_range map shows where temperature varies the most over the twenty years of the region. Once the new raster has been generated, query the raster with **i** button. Consider carefully what these values represent. The values represent the temperature difference between the minimum and maximum mean temperature values. For instance, if the cell value is 2.3, this means that the mean temperature by five-year varied by 2.3 degrees over the 20 years.

5. Map Algebra

The raster can be used to perform mathematical operations with Map-algebra.

Map algebra is a language specifically designed for geographic cell-based systems. The language provides the rules and syntax needed for user to communicate with the computer and other users. Extensions have been added to the algebra that explicitly addresses spatial data manipulation of cell-based data.

Operations:

Addition, Subtraction, Multiplication, Division, Boolean operators, Exponents, Logs, Square root, Absolute value, Rounding, Trigonometry.

These powerful functions allow the execution of complex modeling procedures with spatial data as input and output for predictive model development and can be the foundation for more analysis in GIS.

Generate a raster surface showing the extreme range on temperature from 1971 to 1990 for the study region using the overall minimum and maximum temperature data.

From the overall maximum and minimum raster generated in previous sections, extreme range values can be calculated by subtracting the ***overall_min*** raster from the ***overall_max*** raster. The output raster dataset can be generated using Raster Calculator.

- With Raster Calculator generate an extreme range raster with the following expression

    ```
    "overal_max" - "overal_min"
    ```

- Save the output raster as ***extreme_range***
- The statistics information can be found from the Properties -> Source.

Generate a raster surface showing each five-year mean temperature varied from the historical mean temperature calculated for the study region from 1971 to 1990.

The historical mean temperature for the entire study region from 1971 to 1990 is 2.5 degrees. Climatologists would like to see how each five-year period varied spatially from this historical mean.

In order to do this, you first need a raster that represents the historical mean of 2.5. This would simply be a raster for the study region where all cells have the value of 2.5.

ArcToolbox ->Spatial Analyst-> Raster Creation->Create Constant Raster

- Set Output raster as ***hist_mean***, Constant Value to 2.5 and Data Type to Float
- Set Cell size to 100 and Extent to ***study_bndy***

With the historic mean temperature, we can get how the mean temperature changes from the historic mean temperature by subtracting the historic mean from each mean temperature raster

- With the Raster Calculator, create a raster surface to show the mean temperature varied from historic mean value for each five year.

- Save the output raster as **vary_hist75, vary_hist80, vary_hist85, vayy_hist90** respectively in your local folder
- Symbolize **vary+hist75** to 10 classes with break value at –8, -6, -4, -2, 0, 1, 2, 3. Select a proper color ramp. Save the symbology as a layer file **vary_hist.lyr**
- Apply **very_hist. lyr** to the rest of three rasters for the symbology

Produce a map showing each five year mean temperature varied from historic mean temperature and discuss the changes

With the raster calculator, create a raster to show the overall mean temperature varied from historic mean temperature

- Save the output raster as **overall_hist**

Consider how this analysis provides more of information than that provided by a simple statistical calculation.

Produce the following maps. Symbolize each raster surface in 10 classes and natural break method. The map should come with an appropriate tile, legend, scale and your name and date

- Create a new field to store the percentage of rainfall and snowfall over precipitation respectively for each five-year.
- Calculate the percentage of rainfall over the precipitation and snowfall over precipitation for each five-year

— *Produce a map to show the percentage of rainfall over precipitation for each five-year*

— *Produce a map to show the percentage of snowfall over precipitation for each five-year*

— *Produce a map to show the percentage of snowfall over precipitation for each five-year*

— *Produce a map showing overall mean temperature, rainfall, snowfall and precipitation from 1971 – 1990 for the study region*