PDHonline Course C423 (15 PDH)

GIS Applications in Water Resources and Environmental Engineering

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

COURSE CONTENT – Lecture 4

Course C423 is a practical GIS course with particular reference to applications in Water Resources and Environmental Engineering, addressing some of the basic, yet common GIS needs in these two fields. The course describes the availability of free databases and public domain GIS software, and their use in performing common GIS processing tasks that are relevant to Water Resources and Environmental Engineering. The ability to search, download and make use of such data sets and software is presented within the context of this course. The course consists of 5 lectures, which start off with introductory level material in Lecture 1, and progressively builds up through the remaining four lectures. A basic awareness of the concept of GIS is desirable but not required as a pre-requisite for this course.

.

The following topics are covered in this 4th lecture of Course C423:

- ❖ GIS modeling
- ❖ Common GIS Tools
- ❖ Surface creation
- ❖ Hydrological modeling

1. GIS Modeling

GIS technologies facilitate input of data sets to simulation models. This may take the form of one- or multiple-variable or GIS 'layer' inputs to models. However, users still need to develop or identify models in application areas that are already amenable to the entry of spatially distributed data.

Models are typically formulated to receive one or more spatial data sets directly from the GIS. A typical data model using ArcGIS software for a site suitability GIS application is shown in Figure 1. It takes a set of spatial data sets (Elevation, Study Area, and Freeway), and uses a set of tools (Calculate Slope, Clip, Extract by attributes, Euclidean Distance, Con), to derive the final results which is the suitable area.

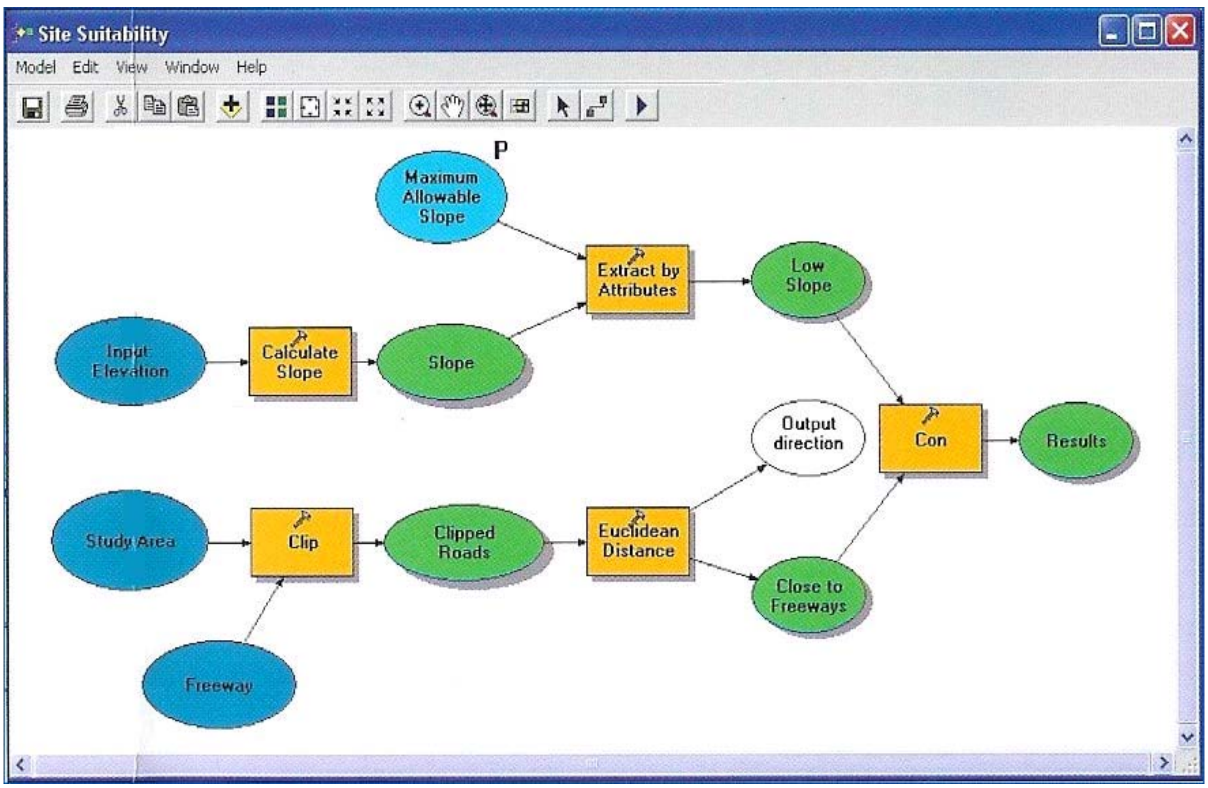


Figure 1: Site Suitability data model

Many current and traditional model procedures use summary information in the form of model coefficients. These types of models can also be improved by GIS or remote sensor technologies by conducting tests so that model coefficients can be further defined or refined through improved measurements of their characteristics. A more 'natural' coefficient better defines the behavior of model variables, and allows the modeler to achieve high fidelity between natural systems and their model simulations.

There are four aspects of modeling spatial information that can be identified: cartographic modeling, simulation/deterministic modeling, statistical/predictive modeling, and model calibration:

Cartographic Modeling: When the user of spatial information is presented with a problem, the response should be a careful plan of what should be done. A more common response is to rush to the computer and start to work. Instead, cartographic modeling suggests detailed flow charts and careful planning to decide what data are important and how they should be used.

Simulation Modeling: Another aspect of modeling is the simulation approach. In this case, the user tries to simulate some complex phenomenon using a combination of spatial and nonspatial information. This approach typically requires an expert who is knowledgeable enough to build such a simulation or model. It should be noted that rarely in these cases do any two experts agree on exactly how the model should be built. A good example of this type of modeling is evaluating wildlife habitat suitability. In this example, one might use the following spatial information layers: vegetation, elevation, aspect, slope, ownership, roads, and streams. This information could then be combined in some model with weights used to prioritize important layers. In addition, calculations of distances (i.e., distance from roads, distance to streams) and measures of diversity may be included in the model. This model is then used to evaluate areas of good habitat and determine where the habitat can be improved.

Predictive Modeling: In this approach, statistical techniques are used to build a model that will be able to predict using the spatial information. The statistical tool used for building such models is most commonly regression analysis. The first step in this process is to collect information about the phenomena one wishes to model. A subset of this information is then used to statistically build the model. This model building is performed by looking at each layer of spatial information and each component of nonspatial information to see which are correlated to the phenomena one wishes to predict. Once the model is built, the model is tested using the remaining information.

An example will elucidate this explanation. Suppose one wants to predict the amount of snowmelt runoff from a forested watershed. These predictions currently are being made by point samples taken throughout the watershed. The predictions can be compared to the actual runoff statistics collected by stream gauges. One might hypothesize that using spatial data that completely covers the area should lead to better predictions than point samples. Therefore, one would put together a

GIS with the necessary layers to predict runoff. These layers might include: vegetation, slope, aspect, snow extent, elevation, and soil type. In addition, the point sample data which includes snow depth and the amount of water in the snow may also be included. Some of this information remains constant over time while some changes daily. Therefore, a variety of conditions (i.e., years) should be represented in the information collected. In this example, one would collect runoff data for dry years, wet years, and average years. A subset of this runoff data would then be used to develop the model and select the necessary spatial and point sample data to be included in it. The remainder of the data would then be used to test the model. Once the model has been proved effective, it can be used to predict snowmelt runoff in future years.

Statistical models can also be applied later to predict other conditions at other places and times if the model represents the behavior of the variable. When statistical models of physical, chemical, or biological systems truly predict variables, users can apply the model to new situations. Good applications include: the evaluation of suspended sediments in freshwater and coastal ecosystems, temperature of water bodies, crop residue and tillage practices for evaluation of nonpoint pollution, concentrations of chlorophyll in water and plants, and a variety of other detailed applications. The use of statistical analyses has proven of great value in water resources studies over large areas.

Model Calibration: Another valuable use of GIS and related technologies is the calibration of model coefficients in statistical models. Both statistical and deterministic models often consist of a number of submodel units. Coefficients used in either approach reflect the characteristics of nature, and they will adjust the contribution of variables or submodels to the overall model results.

To optimize the model simulation of natural phenomena, the coefficients need to reflect the reality of the situation. As a given model begins to approximate nature, its further development often takes the path of improving the quality of coefficients. Many times, a number of experiments will be executed to better measure the level of a coefficient and thus better mimic nature and help supply better model predictions.

In GIS and remote sensor experiments, using statistical models can be greatly facilitated by the analysis of individual coefficients. These analyses are driven to find the 'sensitivity' of the overall model result or simulation to a given variable.

Sensitivity analyses are part of a good modeling strategy because it is very desirable to understand the contributions of model coefficients to the overall results, and to ensure that each variable and/or submodel contribution is appropriate or similar to that of nature.

On many occasions, deterministic or statistical models are 'run' engineers that often have no model inputs from GIS or remote sensor technologies. However, the model *results* are usually some variable that can be measured or evaluated using GIS and remote sensors. The use of separate results from GIS and/or remote sensor technologies supplies an independent verification of modeling results. A number of verification checks between model results and those of GIS or remote sensors will allow a validation of the model. Verification of results is paramount and is used to check the validity of the model as it relates to the real world.

2. Common GIS Tools

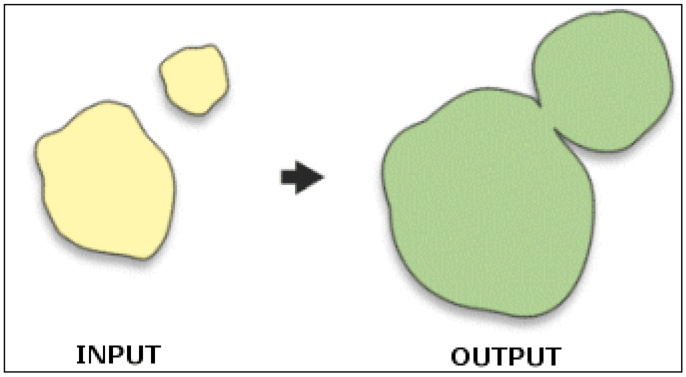
GIS capabilities are extensive but dependent upon the application under consideration. There are however, some basic GIS tools that are commonly used in all GIS applications. Some of these tools are standalone programs, and are explained briefly as follows:

Georeference: The Georeferencing Tool basically sets a real world coordinate system for a given data set. All of the GIS software has built into their coding the georeferencing capability. A free downloadable georeference tool[1] is available. The tool can click three points on the image and provide the real-world coordinates for those points to georeference an image. It can optionally be rectified such that north is straight up. The appropriate world file will then be written.

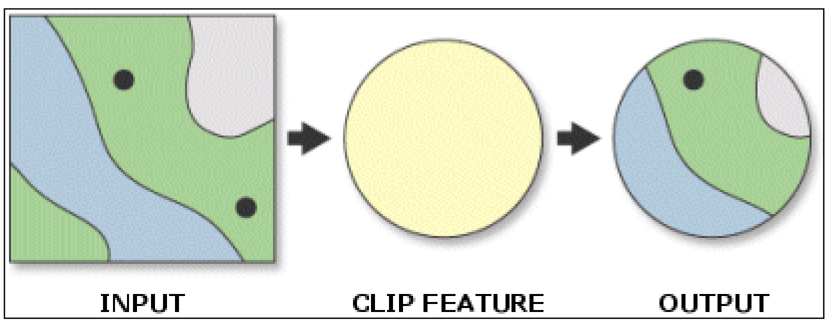
Buffering – This is a technique by which a boundary of known width is drawn around a point, linear or polygon feature. Some examples of point buffers may be a zone around a hazardous waste site or around a tree that is a nest for a particular endangered bird. Examples of linear buffers may be an area around a stream to prevent logging or an area around a utility pipeline to prevent digging. A related function is generally referred to as "proximity searches." Proximity searches can be used to identify adjacencies between particular features or data classifications. Examples might include identifying all groundwater pollution cases within a specific

[1] http://www.softpedia.com/get/Multimedia/Graphic/Graphic-Others/Georeferencing-Tool.shtml

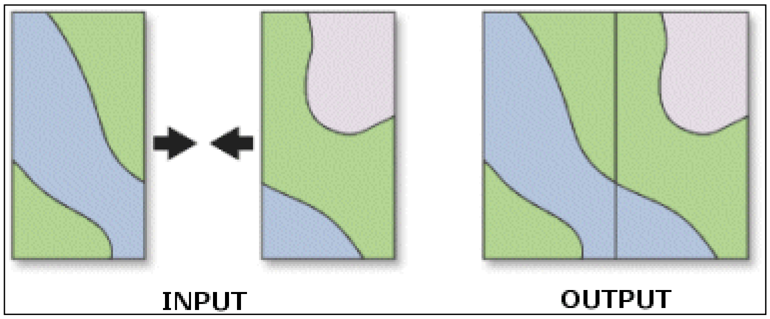
radius of a given drinking supply well. Proximity searches can be used to locate potential sources of contamination, to schedule inspections, to identify monitoring wells or sampling points which might provide relevant data from nearby sites, to identify specific residents or the size exposed populations, or to generate mailing lists for further investigation.



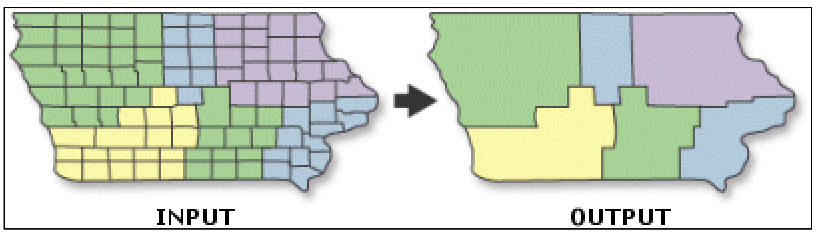
Clip – A tool which extracts only the area of interest for the given GIS application. It can be applied to both raster and vector data sets. Based on an internet search, there is no standalone clip tool. It is bundled into the various GIS software capabilities.



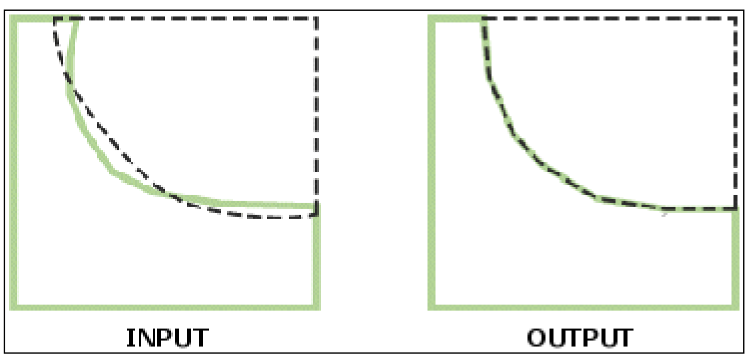
Append – This is one of the Data Management tools for manipulating feature classes. This tool merges multiple feature classes together to create a single feature class.



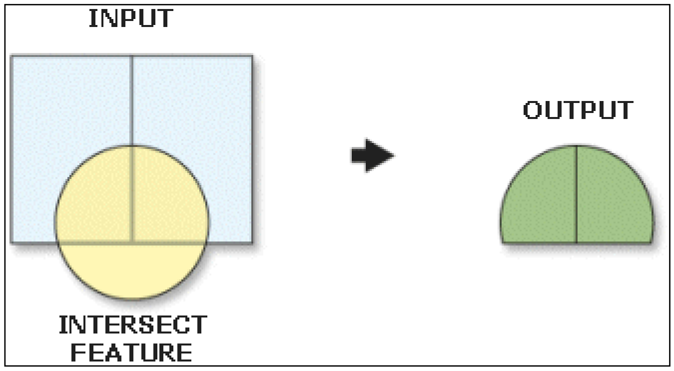
Dissolve – This is one of the Data Management tools used for generalizing features. This tool combines similar features based on a specified attribute or attributes.



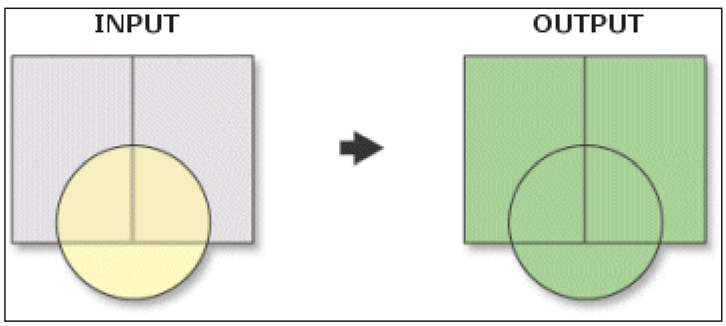
Integrate – This is one of the Data Management tools for manipulating feature classes. This tool compares feature classes and makes identical or coincident any lines or vertices that are within a certain distance of one another.



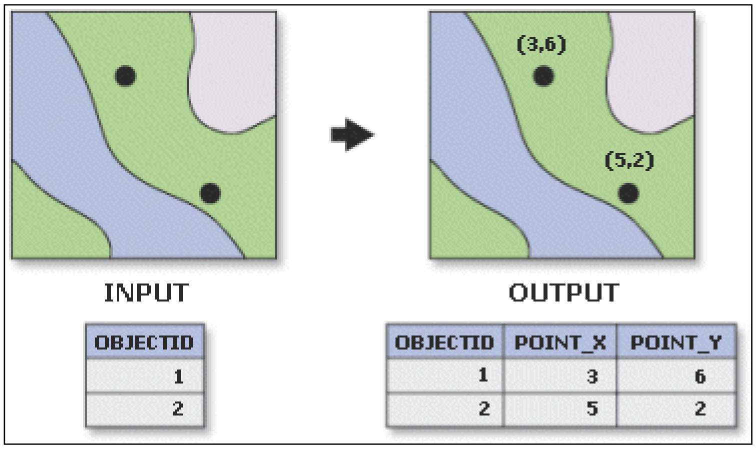
Intersect – Intersect is one of the Analysis tools used to perform overlay analysis on feature classes. This tool builds a new feature class from the intersecting features common in both feature classes.



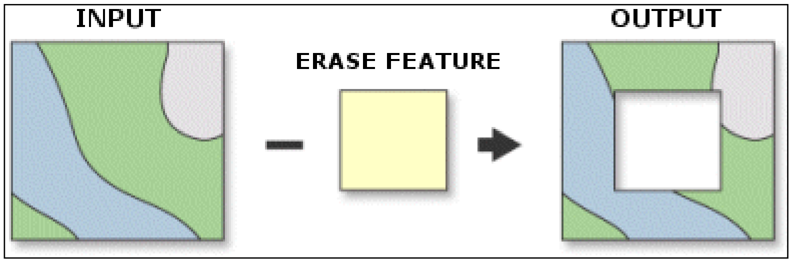
<u>Union</u> – This is another Analysis tool used to perform overlay analysis on feature classes. This tool builds a new feature class by combining the features and attributes of each feature class.



Add XY Coordinates – This is one of the Data Management tools for managing features and their attributes. This tool adds the POINT_X and POINT_Y fields to the feature's attribute table and calculates the x,y location for the fields for labels, points, tics, or nodes.



Erase – This is one of the Analysis tools used to perform overlay analysis on feature classes. This tool creates a feature class from those features or portions of features outside the erase feature class.



Network Analysis – This is a technique by which a linear path is identified that represents the flow of some object through the area. Network analysis is especially useful in hydrology, water resources, transportation, and other disciplines that study the flow of an object. This flow is not limited to water but can also be used for vehicles, utility and communication lines, and animals.

Using the basic tools together with specific tools as they relate to the various GIS applications developed by users, there exist a wide variety of GIS applications in today's professional world.

3. Surface creation

Surfaces in GIS are continuous representations of the GIS phenomena, which are typically raster data types. The creation of surfaces is done using different interpolation methods. Before discussing different interpolation techniques, the differences in the methods used for surface representation need to be discussed. Each representation is useful for specific situations. This discussion concentrates on the creation of a surface in the native ArcGIS grid format.

A grid representation (raster based) of a surface is considered to be a functional surface because for any given *x,y* location, it stores only a single *z* value as opposed to multiple *z* values. Functional surfaces are continuous because an *x,y* location has one and only one *z* value regardless of the direction from which the *x,y* point is approached.

Functional surfaces can be used to represent terrestrial surfaces that depict the earth's surface, statistical surfaces that describe demographic and other types of data, and mathematical surfaces that are based on arithmetic expressions. Surface representation in its simplest form is done by storing *x,y* and *z* values that define the location of a sample and the change characteristic represented by the *z* value.

Contours or isolines are used to define a common characteristic along a line. Technically, contours join locations of equal value to each other. In the case of a contour line representing height, it is a line drawn on a map that connects points of equal elevation above a datum that usually represents mean sea level.

A triangulated irregular network (TIN) is a vector data structure used to store and display surface models. TIN partitions geographic space using a set of irregularly spaced data points, each of which has *x-, y-,* and z-values. These points are connected by edges that form contiguous, non-overlapping triangles and create a continuous surface that represents the terrain.
A grid is a spatial data structure that defines space as an array of cells of equal size that are arranged in rows and columns. In the case of a grid that represents a surface, each cell contains an attribute value that represents a change in *z* value. The location of the cell in geographic space is obtained from its position relative to the grid's origin. Figure 1 shows the four methods of representations that are used to represent the elevation of an area.

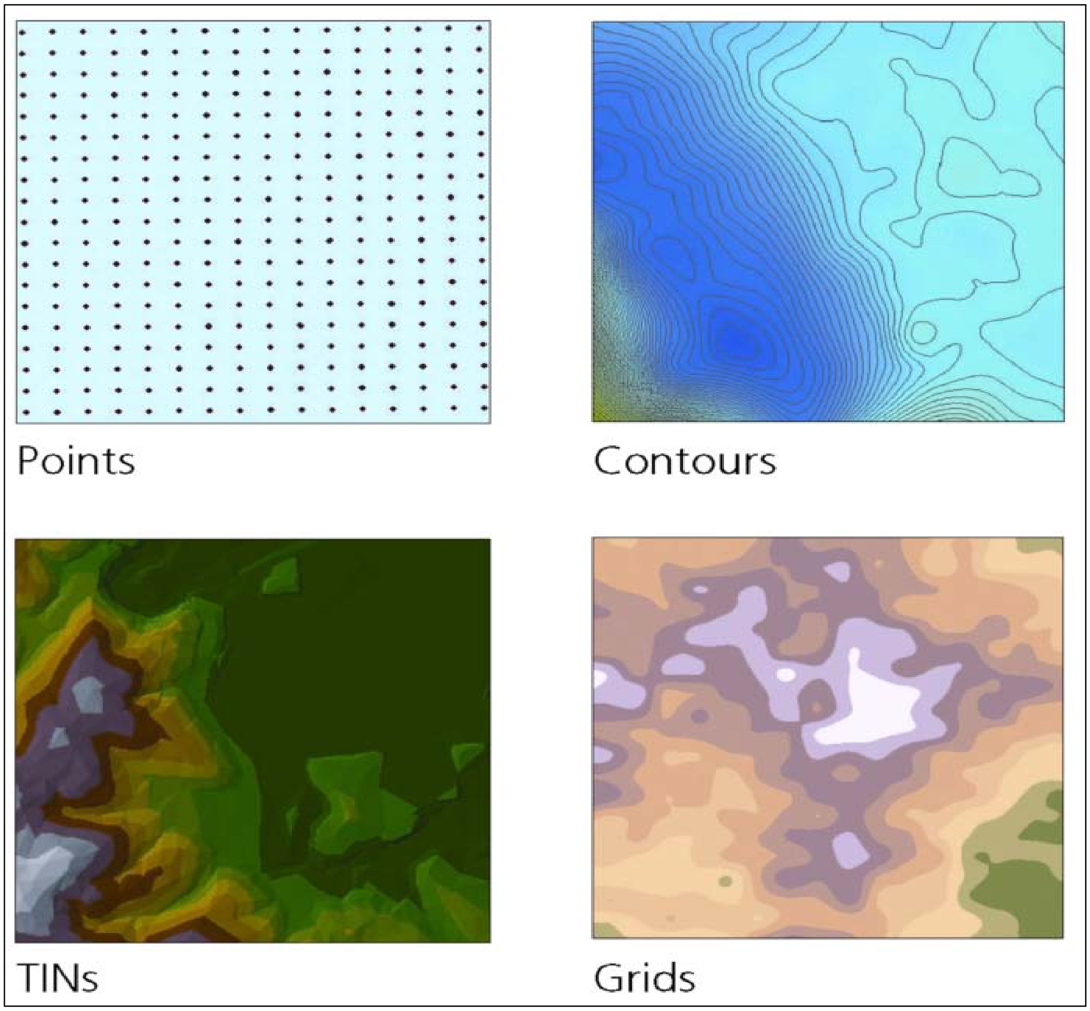


Figure 1: Methods of Surface Representation

There are two categories of interpolation techniques:

- Deterministic interpolation techniques create surfaces based on measured points or mathematical formulas. Methods such as Inverse Distance Weight (IDW) are based on the extent of similarity of cells while methods such as Trend fit a smooth surface defined by a mathematical function.
- Geostatistical interpolation techniques such as Kriging are based on statistics and are used for more advanced prediction surface modeling that also includes some measure of the certainty or accuracy of predictions.

The characteristics of an interpolated surface can be controlled by limiting the input points used in the calculation of output cell values. This can be done by limiting the number of points sampled or the area from which sampled points are taken. Specifying the maximum number of points to be sampled will return the points closest to the output cell location until the maximum number is reached. Alternatively, specifying a fixed radius in map units will select only input points within the radius distance from the center of the output cell unless there are not enough points within that radius.

Many interpolation tools incorporate barriers that define and control surface behavior in terms of smoothness and continuity. Some text refer to the barriers as breaklines/breakpoints. Barriers are needed because sometimes interpolation operations should not be performed across features, such as fault lines, levees, cliffs, and streams that create a linear discontinuity in the surface. By using barriers, changes in the behavior of the surface can be described and enforced.

Available Interpolation Options in ArcGIS

ArcGIS Spatial Analyst extension offers several interpolation tools for generating surface grids from point data. IDW, Spline, Kriging, PointInterp, Natural Neighbors, Trend methods, and Topo to Raster interpolation methods are all available in ArcGIS Spatial Analyst. Each method uses a different approach for determining the output cell values. The most appropriate method will depend on the distribution of sample points and the phenomenon being studied.

- **IDW**

  The IDW (Inverse Distance Weighted) function should be used when the set of points is dense enough to capture the extent of local surface variation needed for analysis. IDW determines cell values using a linear-weighted combination set of sample points. The weight assigned is a function of the distance of an input point from the output cell location. The greater the distance, the less influence the cell has on the output value. Figure 2 shows the use of the IDW interpolation method.

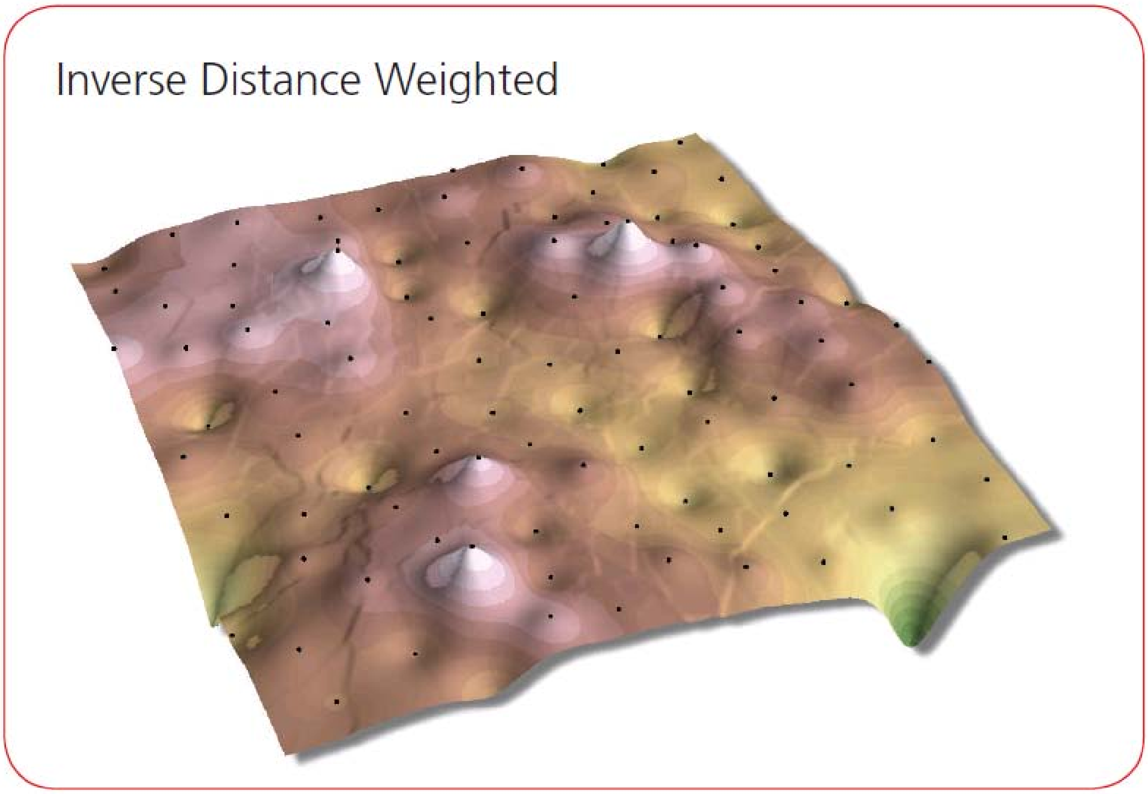


Figure 2: IDW interpolation method

❖ **Spline**

Spline estimates values using a mathematical function that minimizes overall surface curvature. This results in a smooth surface that passes exactly through the input points. Conceptually, it is like bending a sheet of rubber so that it passes through the points while minimizing the total curvature of the surface. It can predict ridges and valleys in the data and is the best method for representing the smoothly varying surfaces of phenomena such as temperature.

There are two variations of spline—regularized and tension. A regularized spline incorporates the first derivative (slope), second derivative (rate of change in slope), and third derivative (rate of change in the second derivative) into its minimization calculations. Although a tension spline uses only first and second derivatives, it includes more points in the spline calculations, which usually creates smoother surfaces but increases computation time. Figure 3 shows the use of the Spline interpolation method.

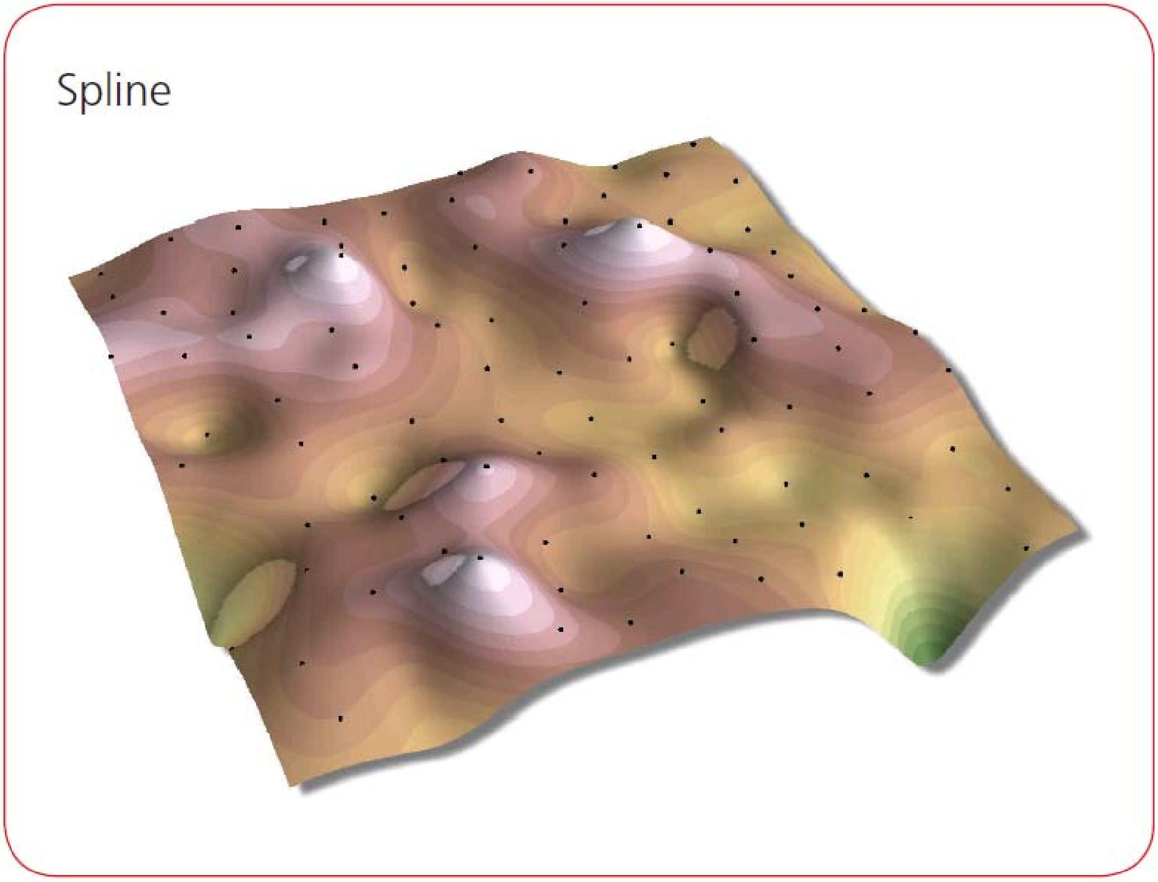


Figure 3: Spline interpolation method

❖ **Kriging**

A powerful statistical interpolation method used for diverse applications such as health sciences, geochemistry, and pollution modeling, Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. It fits a function to a specified number of points or all points within a specified radius to determine the output value for each location. Kriging is most appropriate when a spatially correlated distance or directional bias in the data is known and is often used for applications in soil science and geology.

The predicted values are derived from the measure of relationship in samples using sophisticated weighted average techniques. It uses a search radius that can be fixed or variable. The generated cell values can exceed value range of samples, and the surface does not pass through samples.

There are several types of Kriging. Ordinary Kriging, the most common method, assumes that there is no constant mean for the data over an area mean (i.e., no trend). Universal Kriging does assume that an overriding trend exists in the data and that it can be modeled. Figure 4 shows the use of the Kriging interpolation method.

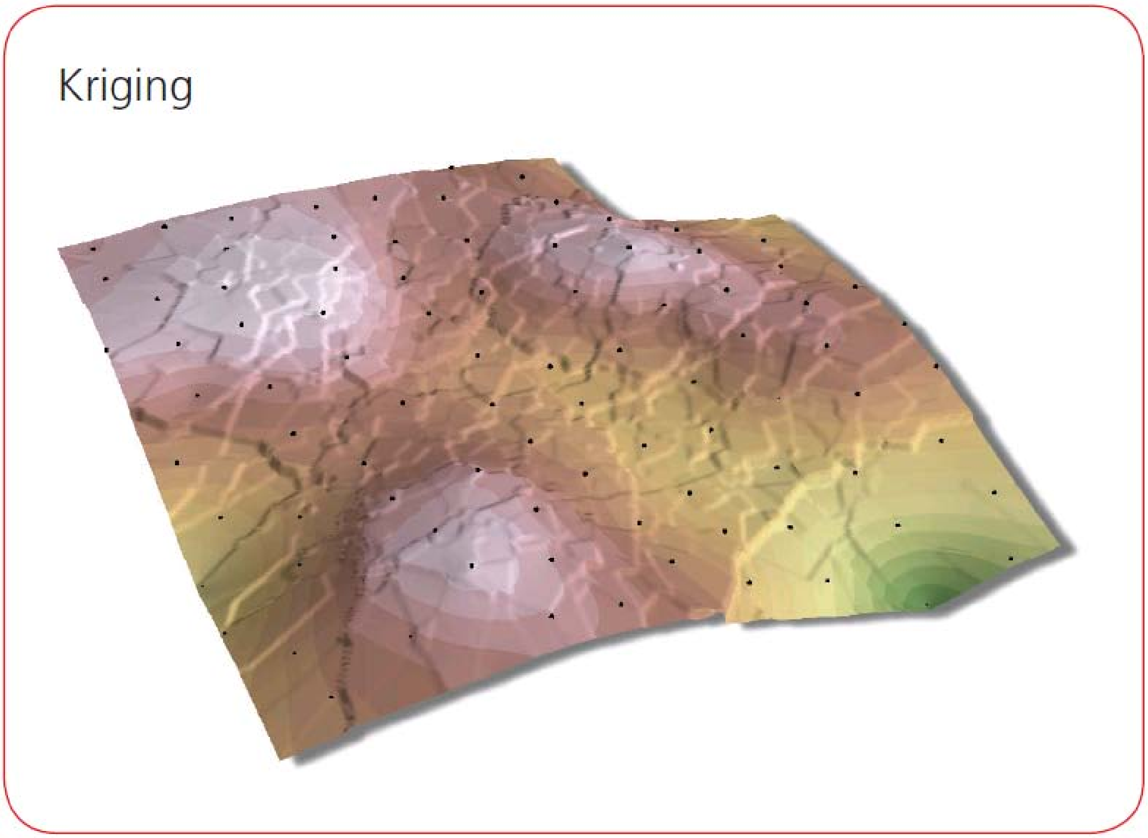


Figure 4: Kriging interpolation method

❖ **<u>PointInterp</u>**

A method that is similar to IDW, the PointInterp function allows more control over the sampling neighborhood. The influence of a particular sample on the interpolated grid cell value depends on whether the sample point is in the cell's neighborhood and how far from the cell being interpolated it is located. Points outside the neighborhood have no influence.

The weighted value of points inside the neighborhood is calculated using an inverse distance weighted interpolation or inverse exponential distance interpolation. This method interpolates a raster using point features but allows for different types of neighborhoods. Neighborhoods can have shapes such as circles, rectangles, irregular polygons, annuluses, or wedges. Figure 5 shows the use of the PointInterp interpolation method.

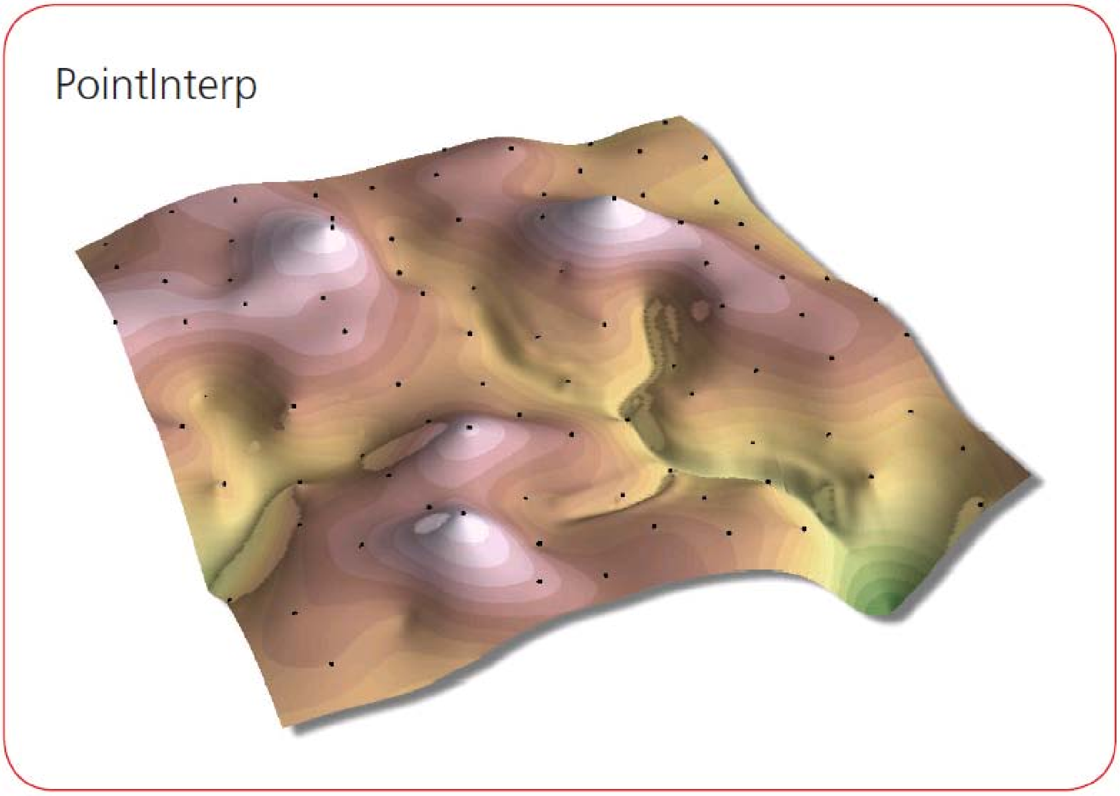


Figure 5: PointInterp interpolation method

❖ **Natural Neighbor**

Natural Neighbor interpolation has many positive features, can be used for both interpolation and extrapolation, and generally works well with clustered scatter points. Another weighted-average method, the basic equation used in natural neighbor interpolation is identical to the one used in IDW interpolation. This method can efficiently handle large input point datasets. When using the Natural Neighbor method, local coordinates define the amount of influence any scatter point will have on output cells. Figure 6 shows the use of the Natural Neighbor interpolation method.

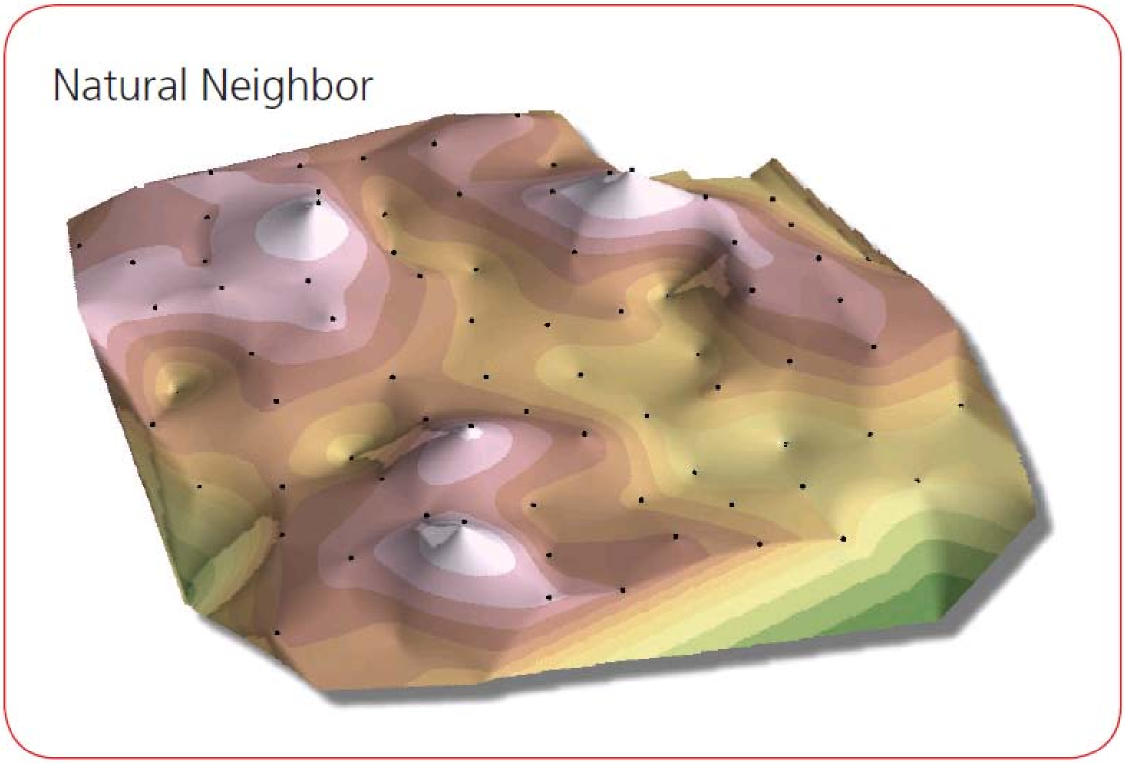


Figure 6: Natural Neighbor interpolation method

❖ **Trend**

Trend is a statistical method that finds the surface that fits the sample points using a least-squares regression relationship. It fits one polynomial equation to the entire surface. This results in a surface that minimizes surface variance in relation to the input values. The surface is constructed so that for every input point, the total of the differences between the actual values and the estimated values (i.e., the variance) will be as small as possible. It is an inexact interpolator, and the resulting surface rarely passes through the input points. However, this method detects trends in the sample data and is similar to natural phenomena that typically vary smoothly. Figure 7 shows the use of the Trend interpolation method.

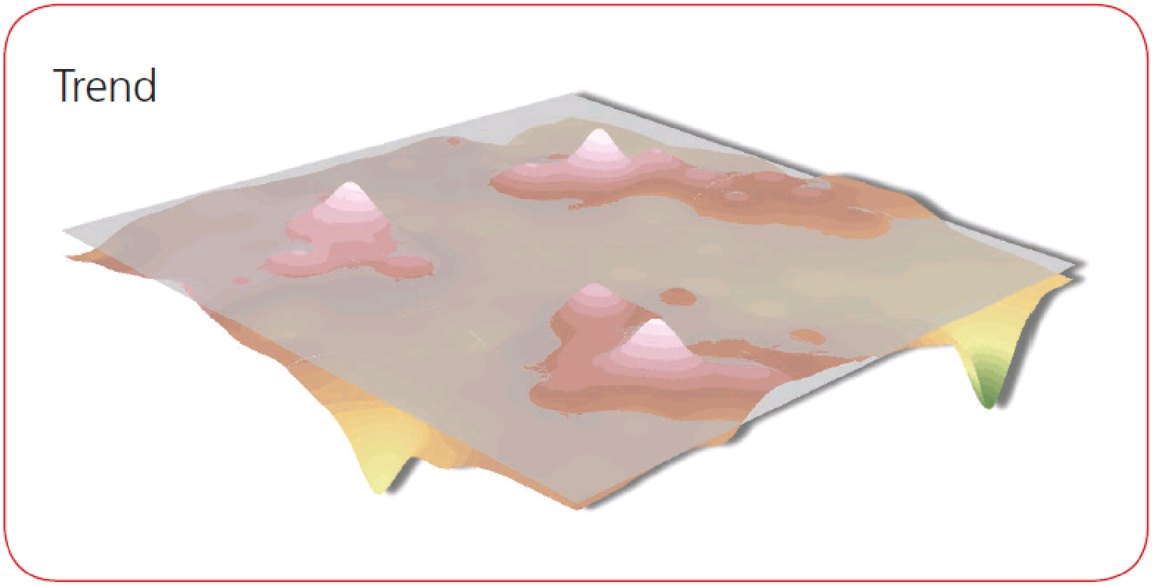


Figure 7: Trend interpolation method

❖ Topo to Raster

By interpolating elevation values for a raster, the Topo to Raster method imposes constraints that ensure a hydrologically correct digital elevation model that contains a connected drainage structure and correctly represents ridges and streams from input contour data. It uses an iterative finite difference interpolation technique that optimizes the computational efficiency of local interpolation without losing the surface continuity of global interpolation. It was specifically designed to work intelligently with contour inputs.

4. Hydrological modeling

Hydrologic models are simplified, conceptual representations of a part of the hydrologic cycle. They are primarily used for hydrologic prediction and for understanding hydrologic processes. Two major types of hydrologic models can be distinguished:

- ❖ Stochastic Models. These models are black box systems, based on data and using mathematical and statistical concepts to link a certain input (for instance rainfall) to the model output (for instance runoff). Commonly used techniques are regression, transfer functions, neural networks and system identification. These models are known as stochastic hydrology models.

- ❖ Process-Based Models. These models try to represent the physical processes observed in the real world. Typically, such models contain representations of surface runoff, subsurface flow, evapotranspiration, and channel flow, but they can be far more complicated. These models are known as deterministic hydrology models. Deterministic hydrology models can be subdivided into single-event models and continuous simulation models.

In this lecture the focus is on Process-based models that are specific to surface runoff. Hydrologic Modeling is used to: delineate watershed; delineate the path when there is a breach in a dam.

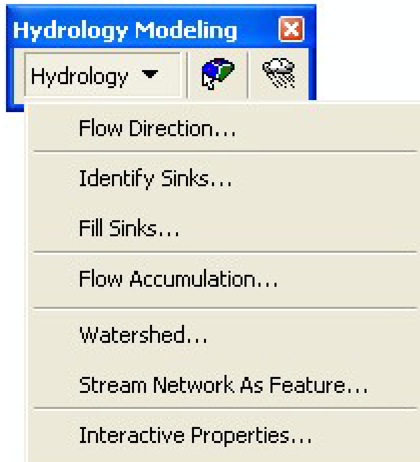
The shape of a surface determines how water will flow across it. The sample hydrologic analysis extension in ArcGIS provides a method to describe the physical characteristics of a surface. Using a digital elevation model as input, it is possible to delineate a drainage system and then quantify the characteristics of that system. The tools in the extension let users determine, for any location in a grid, the upslope area contributing to that point and the downslope path water would follow.

Watersheds and stream networks, created from DEMs using the sample extension, are the primary input to most surface hydrologic models. These models are used for determining the height, timing, and inundation of a flood, as well as locating areas contributing pollutants to a stream, or predicting the effects of altering the landscape.

The tool used is specific to the AcrGIS software and the extension is available for free which is called ArcHydro[2]. It is developed by the Department of Civil Engineering, Texas A&M University. The tools are used for Watershed and Stream Delineation. Hydrology Modeling is a set of tools created by ESRI for the application of the hydrological functions in ArcGIS.

[2] http://resources.esri.com/arcHydro/

The hydrologic sample extension has a set of dialogs to delineate watersheds or define stream networks. In addition to watershed and network analysis, it has two interactive tools, one to model how water if dropped at a specified location will flow through the landscape and the second identifies the contributing watershed for specified locations.



No matter what a user goal is, it starts with an elevation model. From the elevation model users find out what cells flow into which cells (i.e., the flow direction). However, if there are errors in the elevation model or if users are modeling karst geology, there may be some cell locations that are lower than all the surrounding cells. If this is the case, all water traveling into the cell will not travel out. These depressions are called sinks. The sample extension allows users to identify the sinks and gives users the tool to fill them. The result is a depressionless elevation model. Users then determine the flow direction on this depressionless elevation model.

If users are delineating watersheds, they then need to identify pour points, which are locations that they wish to know the contributing watershed for. Usually these locations are mouths of streams or some other hydrologic point of interest. In the sample extension, users can specify their pour points or they can use the stream network as the pour points. In the later case, to create the stream network users must first calculate the flow accumulation for each cell location (see below for more explanation).

If users are defining stream networks, they will not only need to know the direction water flows from cell to cell but also how much water flows through a cell, or how many cells flow into another cell. When enough water flows through a cell, then the location is considered to have a stream passing through it.

When delineating watersheds or defining stream networks, users must proceed through a series of step processes as shown in Figure 8. Some steps are mandatory while others are optional depending on the characteristics of the input data. Flow across a surface will always be in the steepest downslope direction. Once the direction of flow out of each cell is known, it is possible to determine which and how many cells flow into any given cell. This information can be used to define watershed boundaries and stream networks. The following flowchart shows the process of extracting hydrologic information, such as watershed boundaries and stream networks, from a DEM.

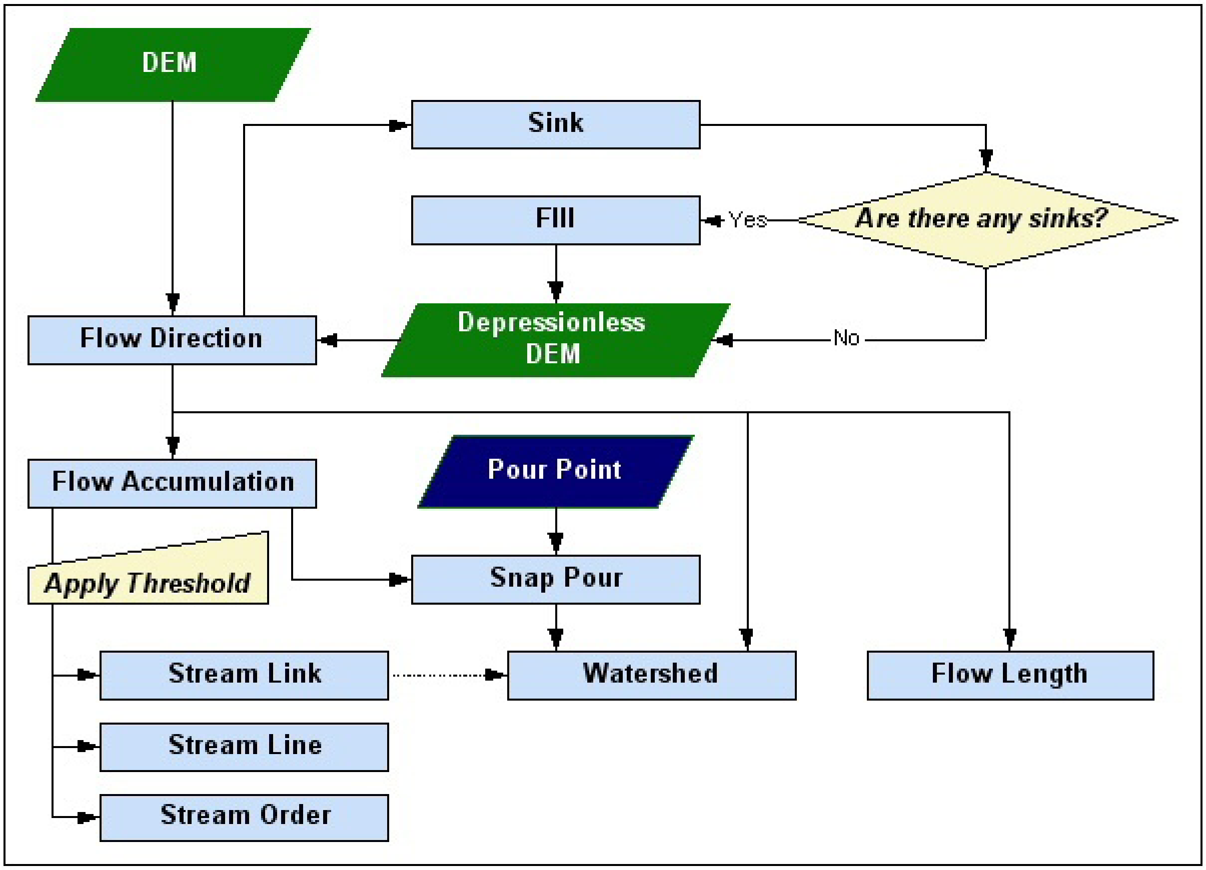


Figure 8: Hydrological Function Processing

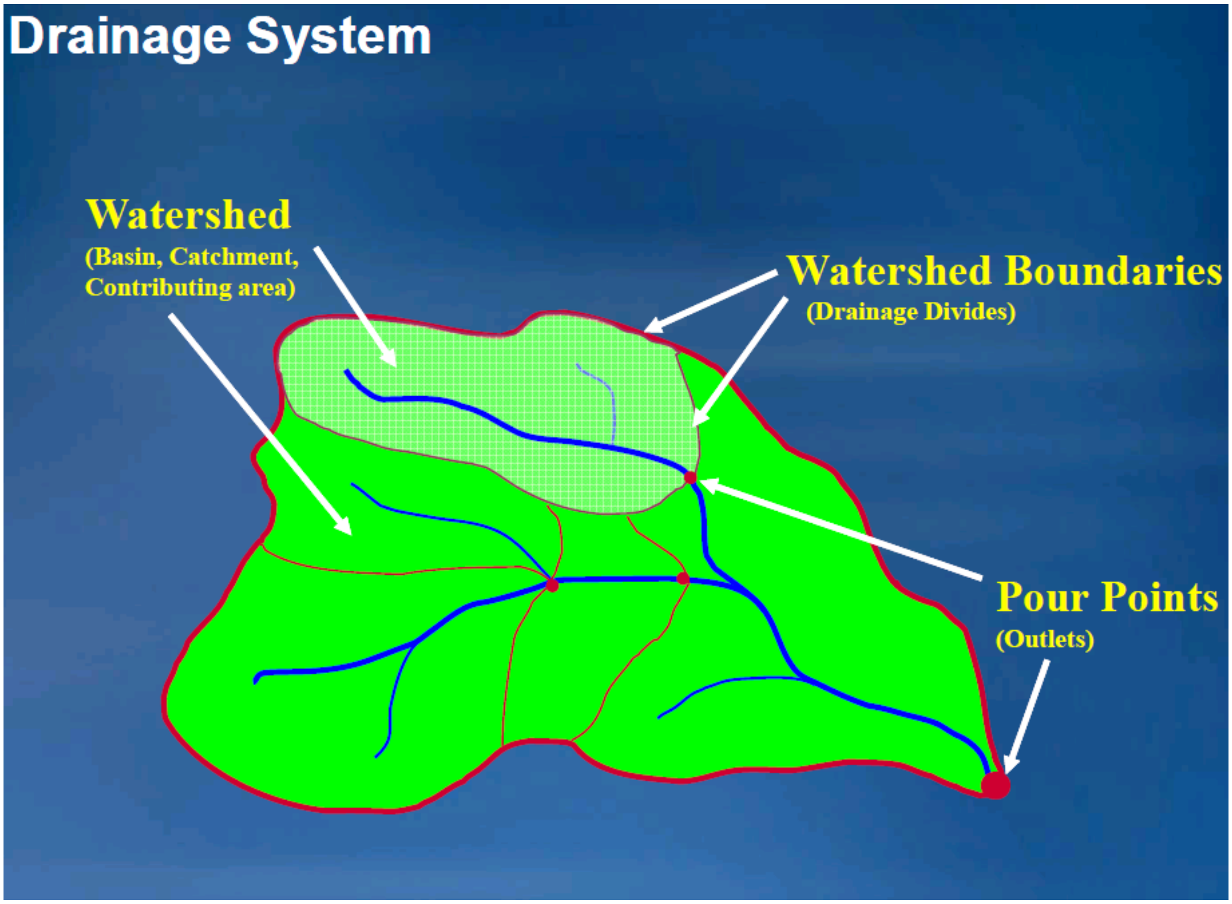


Figure 9: Hydrological characteristic features

The basic features that are important in Hydrological modeling are shown in Figure 9. The characteristics shown are as follows:

Watershed – A watershed is the area of land where all of the water that is under it or drains off of it goes into the same place. Watersheds come in all shapes and sizes. They cross county, state, and national boundaries.

Drainage Basin – A drainage basin is an extent or area of land where surface water from rain and melting snow or ice converges to a single point, usually the exit of the basin, where the waters join another waterbody, such as a river, lake, reservoir, estuary, wetland, sea, or ocean. In closed drainage basins the water converges to a single point inside the basin, known as a sink, which may be a permanent lake, dry lake, or a point where surface water is lost underground.

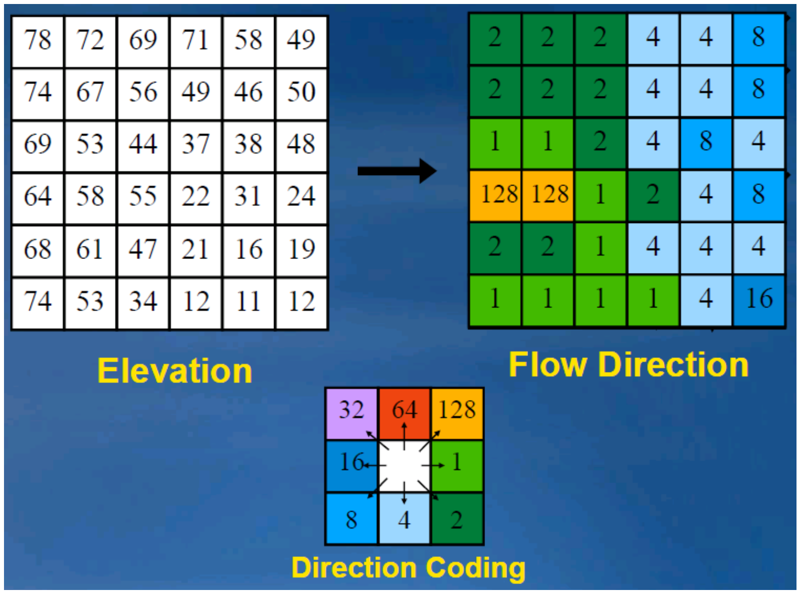
Catchment – Other terms that are used to describe a drainage basin are catchment, catchment area, catchment basin, drainage area, river basin, water basin and watershed.

Drainage divides – A drainage divide, water divide, divide or watershed is the line separating neighboring drainage basins (catchments).

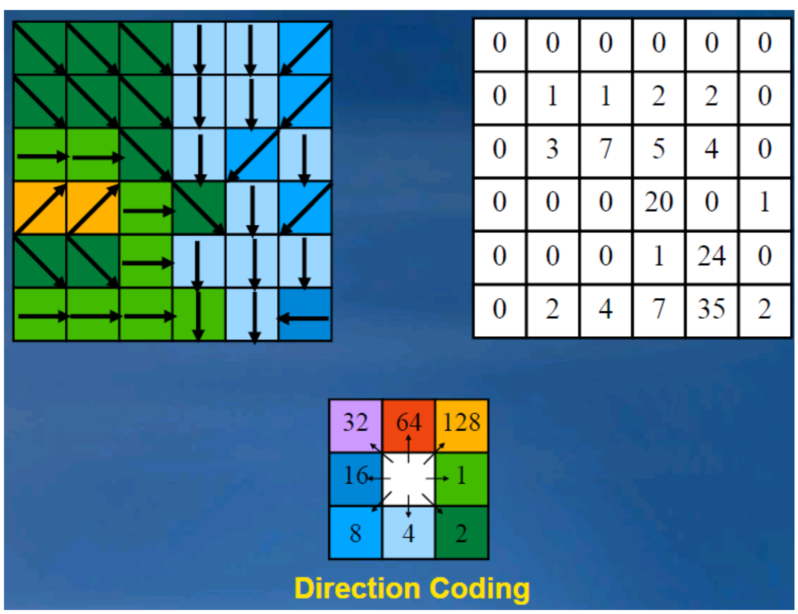
Pour Points – Also called Snap the "pour point" of a watershed to the cell of flow accumulation within a neighborhood.

Flow Length – Calculates the length of the upstream or downstream flow path from each cell.

Flow Direction – a raster data set of flow direction from each cell to its steepest downslope neighbor.



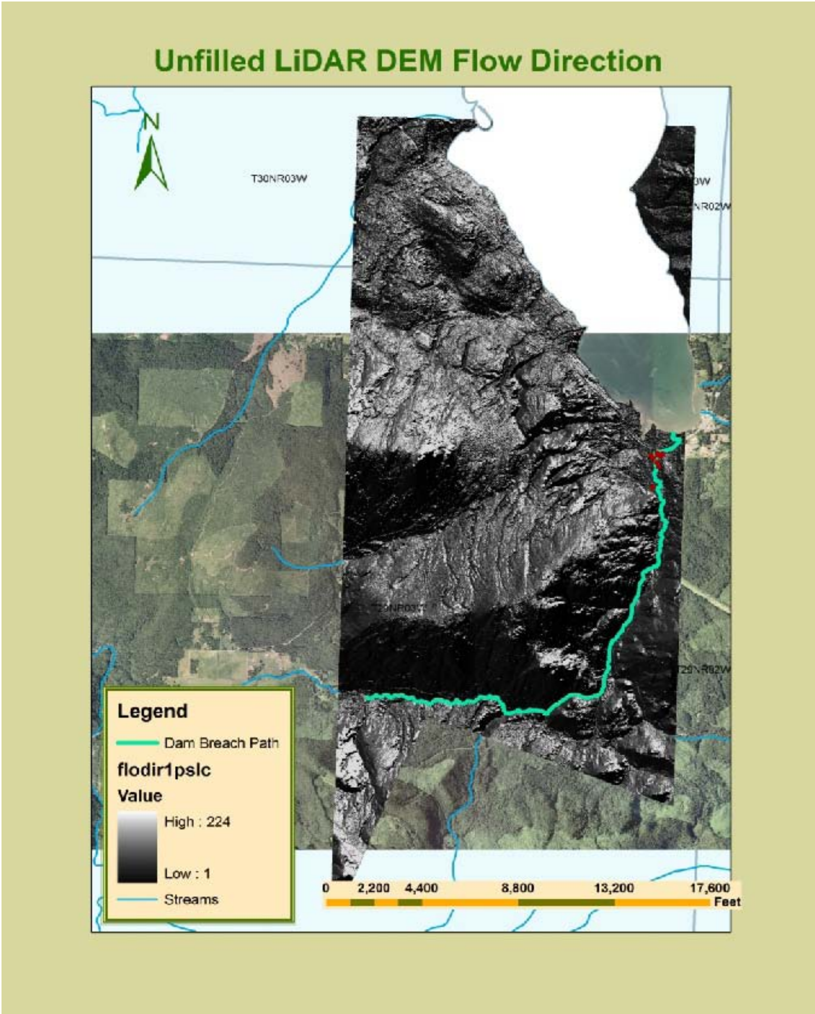
Flow Accumulation – a raster data set of accumulated flow to each cell.



Steps taken to perform Hydrological modeling

Step 1: Create an Initial Flow Direction & Identify the Number of Sinks

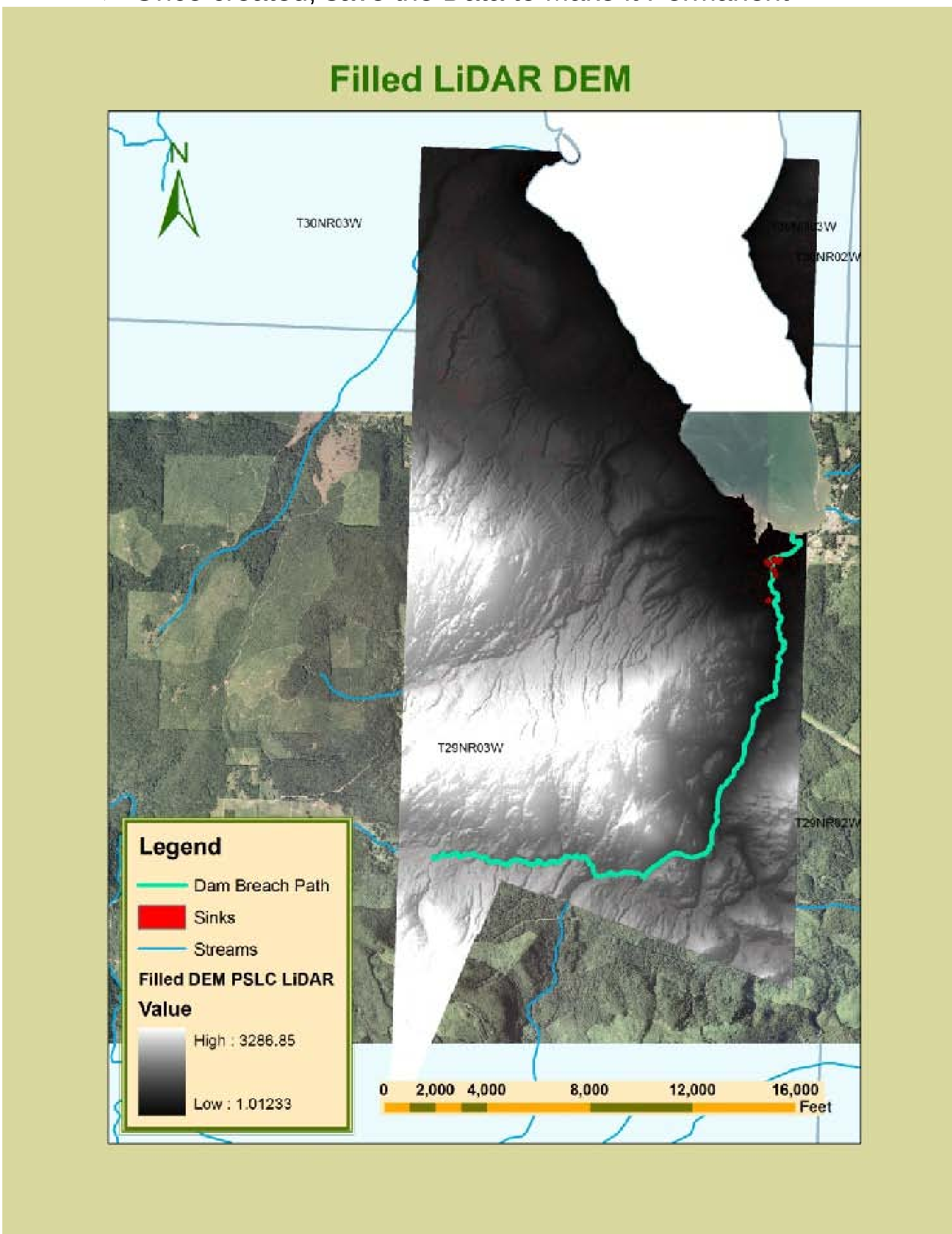
- A DEM free of sinks, a *depressionless DEM*, is required for Hydrologic Modeling.
- Create 1st Flow Direction raster, determining the flow direction of every cell in the grid.
- Determine sinks on the DEM – Sinks on the DEM requires that they be filled using the filling tool in the hydrologic modeling software. Sinks in the DEM prevents further hydrologic computations.
- DEMs with Sinks in excess of 5,000 should be clipped.
- This sets the stage for filling the DEM.



An example of an unfilled DEM

Step 2: Fill Sinks, Creating New Filled DEM

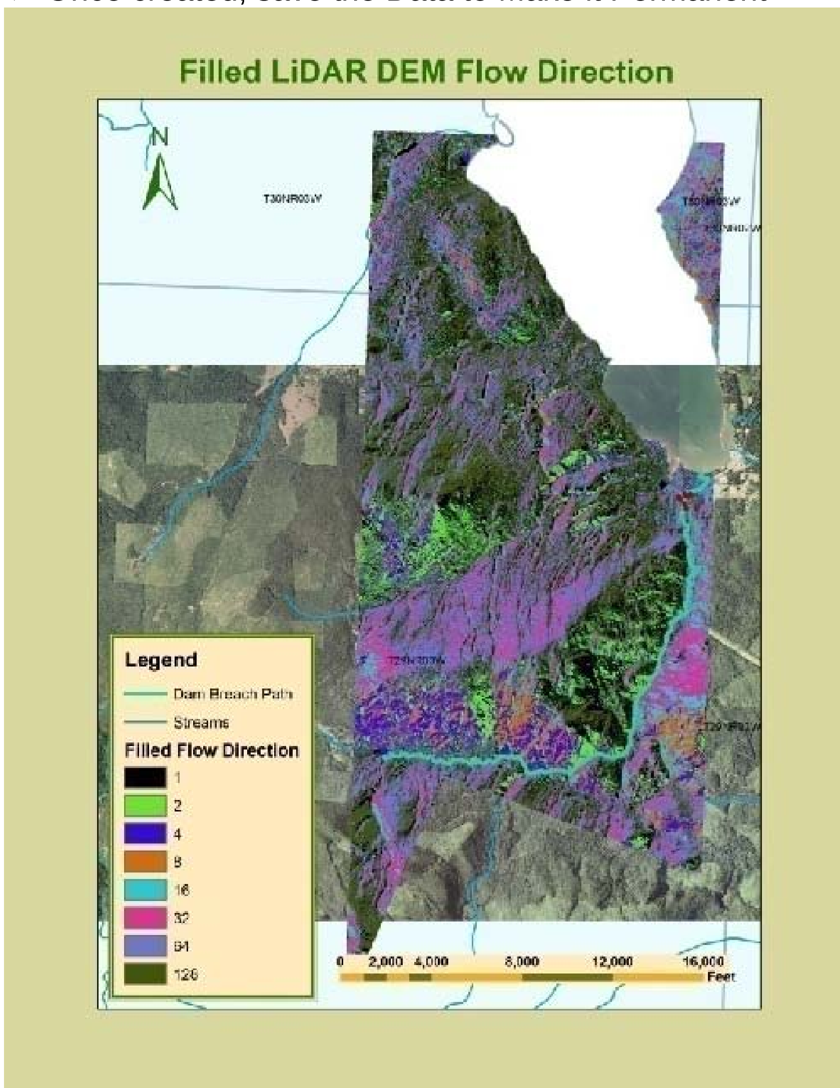
- ❖ Input original LiDAR DEM and set fill limit as necessary
- ❖ Allow output Raster to be Temporary
- ❖ Once created, save the Data to Make it Permanent



An example of a filled DEM

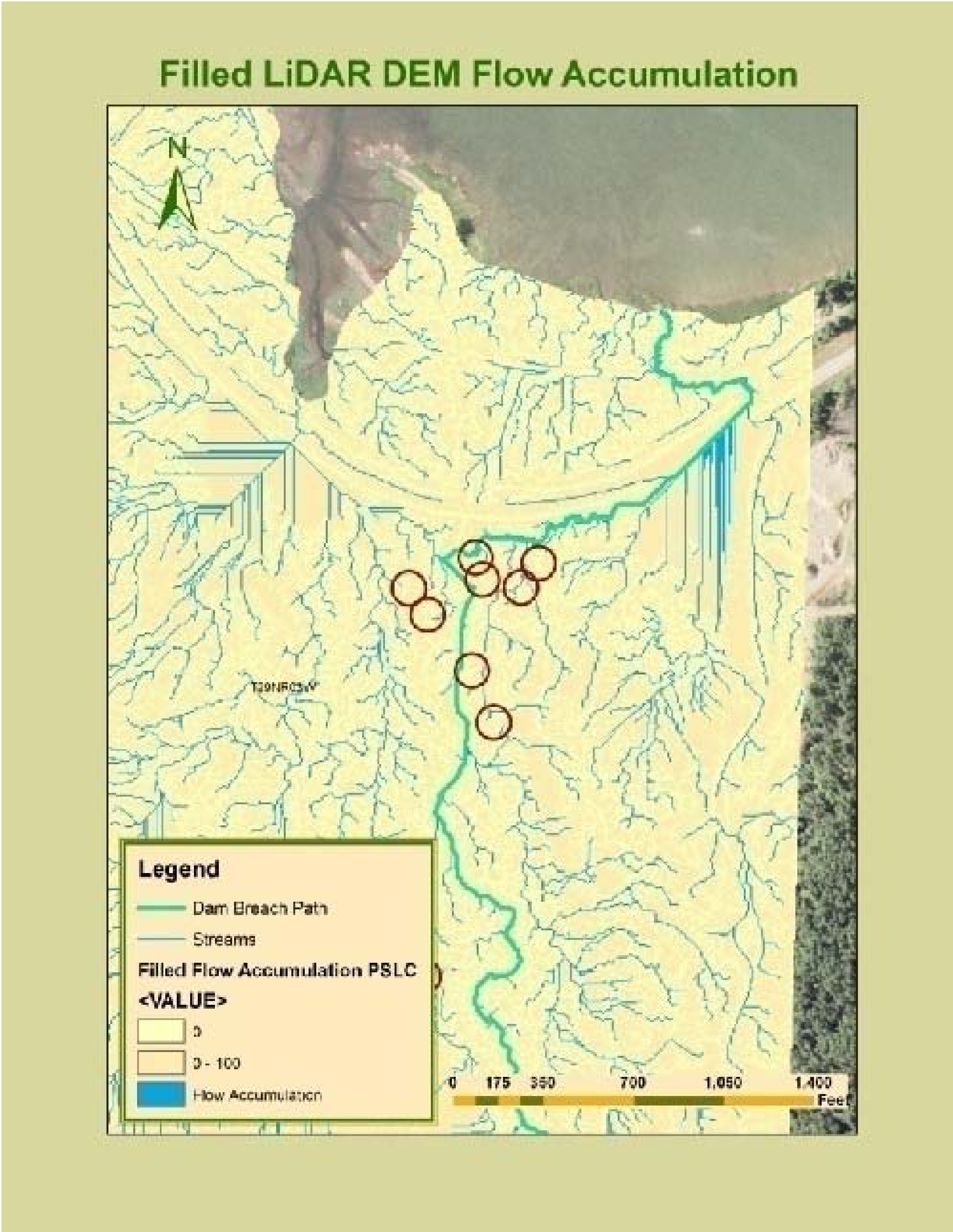
Step 3: Create Filled Flow Direction and Flow Accumulation Rasters

- For Flow Direction use the Filled DEM as Input
- Output is Temporary
- Once created, save the Data to Make it Permanent



An example of a filled DEM with Flow Direction

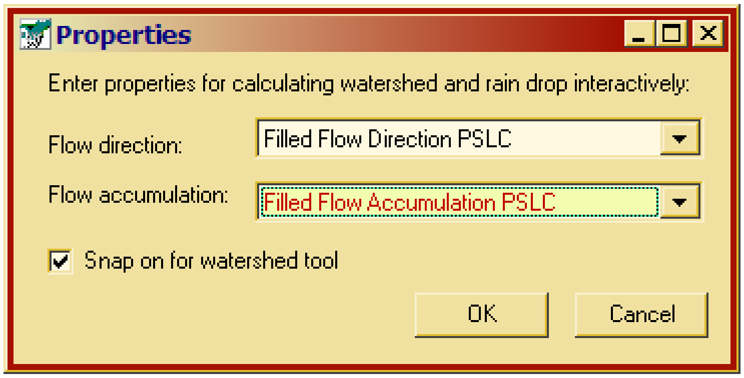
- For **Flow Accumulation** use Filled Flow Direction
- Allow output Raster to be Temporary
- Once created, save the Data to Make it Permanent



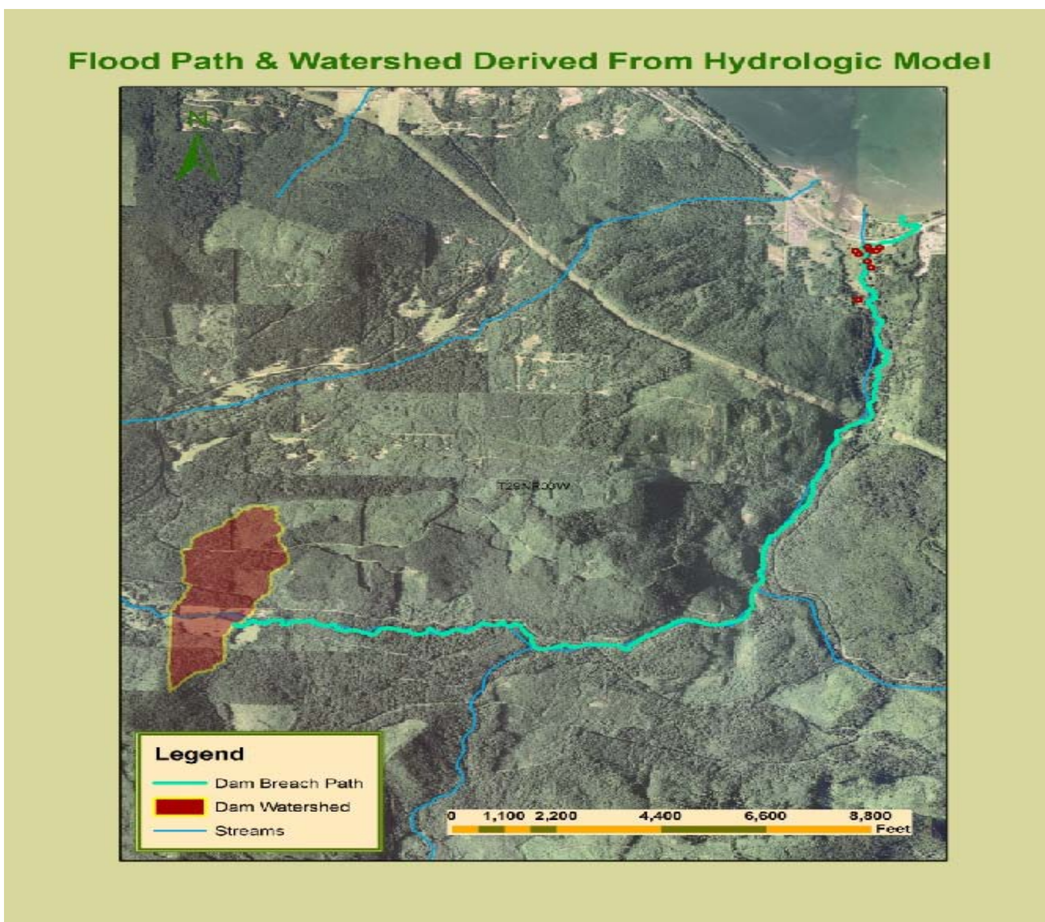
An example of a filled DEM with Flow Accumulation

<u>Step 4 :</u> Create Flood Path & Watershed Extent Using Raindrop and Watershed Delineation

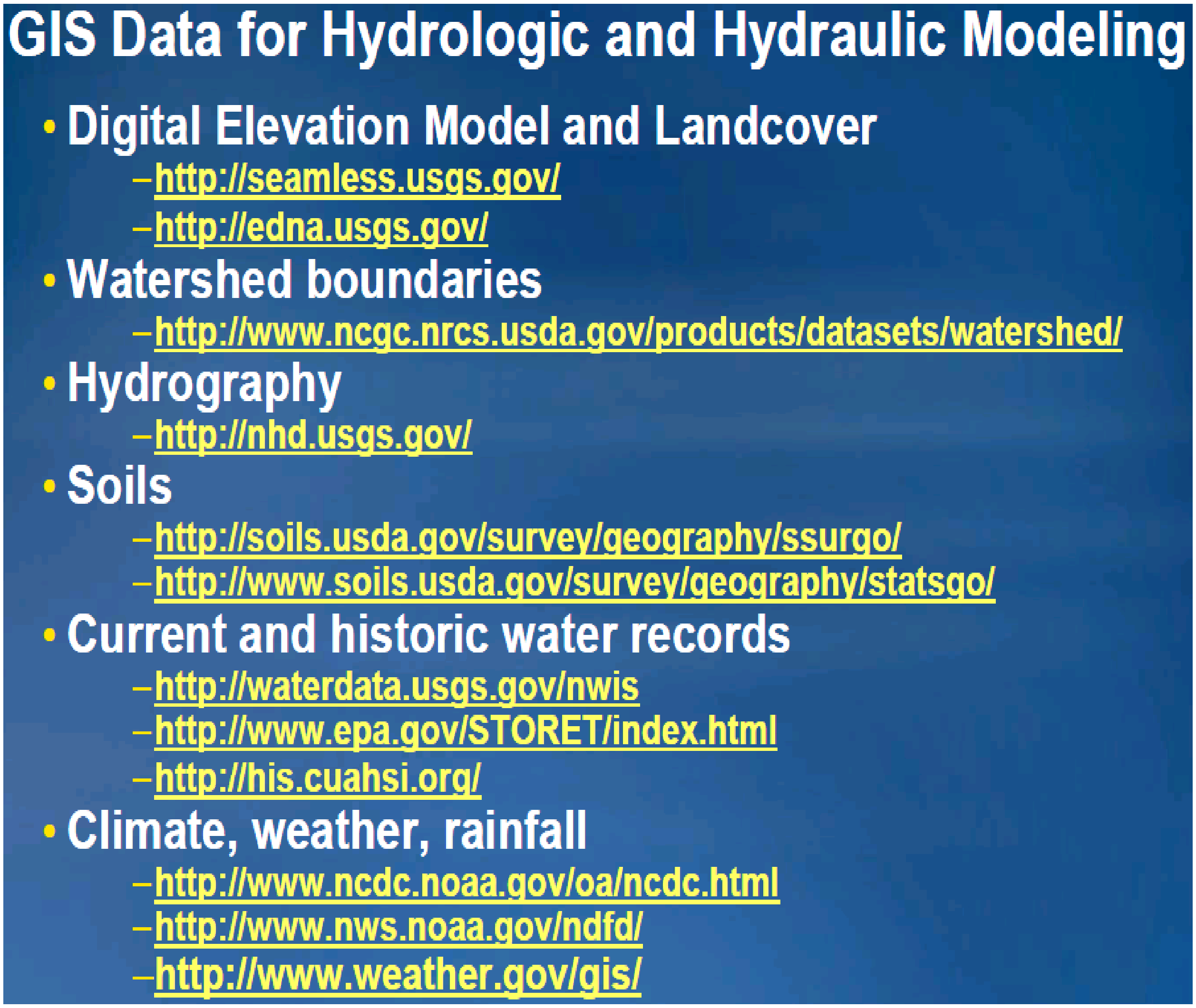
- Enter Filled **Flow Direction** and **Flow Accumulation**



- Opens Watershed & Raindrop Buttons in Toolbar
- Choose **Raindrop** & Select Start Point
- Choose **Watershed** & Select Start Point



An example of a flood path and watershed


GIS Data for Hydrologic and Hydraulic Modeling
• Digital Elevation Model and Landcover
–http://seamless.usgs.gov/
–http://edna.usgs.gov/
• Watershed boundaries
–http://www.ncgc.nrcs.usda.gov/products/datasets/watershed/
• Hydrography
–http://nhd.usgs.gov/
• Soils
–http://soils.usda.gov/survey/geography/ssurgo/
–http://www.soils.usda.gov/survey/geography/statsgo/
• Current and historic water records
–http://waterdata.usgs.gov/nwis
–http://www.epa.gov/STORET/index.html
–http://his.cuahsi.org/
• Climate, weather, rainfall
–http://www.ncdc.noaa.gov/oa/ncdc.html
–http://www.nws.noaa.gov/ndfd/
–http://www.weather.gov/gis/