Synopsis of Spatial Analysis Using Grids
GIS in Water Resources

David Tarboton, September 2019

Learning objectives

By the end of this class you should be able to:

- describe some ways continuous surfaces or spatial fields are represented in GIS
- list and describe the data elements that comprise a grid data structure
- describe how integer, categorical and real valued data fields are represented using grids
- use map algebra to perform raster calculations on grids and explain the scale issues involved in raster calculations where grids may not have the same cell size
- define interpolation and describe aspects to be considered in interpolating a spatial field
- use interpolation to determine watershed average precipitation (After Ex 3)
- calculate the slope of a topographic surface represented by a grid using: (i) ArcGIS method based on finite differences, (ii) D8 model in direction of steepest single flow direction, and (iii) D∞ model in direction of steepest outward slope on grid centered triangular facets.

Reading

- ArcMap help: What is raster data? http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-raster-data.htm
- ArcMap help: Fundamentals of raster data from Cell size of raster data to Raster dataset attribute tables http://desktop .arcgis.com/en/arcmap/latest/manage-data/raster-and-images/cell-size-of-raster-data.htm
- ArcGIS Pro An overview of the Spatial Analyst Toolbox https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/an-overview-of-the-spatial-analyst-toolbox.htm. Read this to get a general sense of the tools available. Tools we will use now or later include, from the Surface toolset: Slope, Aspect, Contour; from the Zonal toolset: Zonal Statistics; from Map Algebra: Raster Calculator; from Interpolation: Spline and Kriging; and many tools in the Hydrology toolset.
- How Slope Works https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-slope-works.htm
- How Aspect Works https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-aspect-works.htm

Synopsis

This class will introduce the concept of spatial fields as a form of geographical information used in GIS and the use of raster or grid data structures to represent spatial fields. The differences between vector and raster representations of geographic information will be discussed and key concepts involved in the use of raster data representations in hydrology will be introduced.

There are two fundamental ways of representing information in a GIS, as either discrete spatial objects or spatial fields. The class has to date discussed mainly the discrete object or vector representation based on points, lines and polygons. By contrast, a *spatial field* may be thought of as a continuous surface or a layer for which a value is defined at every point in geographic space. Certain things are more amenable to the object representation, while others are more amenable to the field representation. For example, a river or a road is most easily represented as a linear object, whereas the elevation surface or atmospheric pressure may be better represented by a field since there is a value defined everywhere.

There are a number of approaches in GIS to numerically represent spatial fields. A triangulated irregular network (TIN) is one. A TIN is a data structure for representing surfaces as a connected network of triangles. Each triangle node has an x,y coordinate and a z or surface elevation value. TIN representations can be used to estimate the surface value for any other location using interpolation. Contours are another way of representing spatial fields. Contours are lines that represent locations having an equal value, such as surface elevation height. Contours actually provide a way to represent a continuous surface using vector feature classes with interpolation between contours being implied to obtain the value of the surface at intermediate points. The third method used to represent a spatial field is a raster or grid. The terms "raster" and "grid" are used more or less interchangeably in GIS. This class focuses on the grid representation.

A *grid* defines geographic space as a mesh of identically-sized cells that are usually square. Each cell holds a numeric value that measures a geographic attribute (like elevation) for that unit of space. Complete specification of a grid includes the cell size, number of rows and columns, corner coordinates, a specification for how *no data* values are encoded and numerical values at each grid cell. Grid values may be integer representing discrete or categorical information, or real valued (floating decimal point). Integer grids may have *value attribute tables* associated with them to hold information about the categorical quantities they represent. For example land cover or land use classes are encoded using integer grids with value attribute table that contains information such as the name for the corresponding class. By contrast, the national elevation dataset digital elevation model (DEM) is a real valued grid that has no value attribute table. Grids may be used to represent points as discrete grid cells, lines as sequences of grid cells and polygons as zones of grid cells. Since grids have a fixed cell size they are unable to represent information at a scale smaller than the cell (pixel) size. The cell must be small enough to capture the required detail but large enough so computer storage and analysis can be performed efficiently. Grid representation of data involves rules for resolving the spatial variability of the phenomenon into a single value for each cell in the grid.

ArcGIS provides the capability for *map algebra*, or raster calculations that allow you to perform cell by cell evaluation of algebraic expression on entire grids in one operation. This is a way to calculate new information using grids. For example given grids of precipitation (P) and evaporation loss (L), a runoff grid could be calculated using the expression (P-L). The power of raster calculations is that much more complex algebraic expressions can also be evaluated. A complication with raster calculations is that the cell sizes of the rasters used in the calculations may be different. By default ArcGIS then interpolates all rasters in the calculation to the grid scale of the coarsest raster in the calculation using a nearest neighbor scheme. It is possible to use other interpolation methods and to define the cell size of the output using environment settings in ArcGIS.

Interpolation is the estimation of the values of a spatial field between known values. It is an intrinsic part of raster analysis and occurs whenever rasters with different cell sizes or misaligned grids are used in an analysis. It also occurs when rasters are projected or resampled. ArcGIS supports a number of interpolation methods such as nearest neighbor, bilinear or cubic for raster resampling. The default is nearest neighbor, usually a good choice for integer for categorical rasters, but for real valued spatial fields the smoother bilinear or cubic methods are usually better. Spatial fields may also interpolated from sample points using point to raster interpolation methods, such as inverse distance weight, kriging and splines. Spatial interpolation of point precipitation measurements may be used to calculate area average precipitation over a watershed. The commonly used Thiessen polygon method is a special case based on nearest neighbor concepts.

Digital elevation model (DEM) grids that represent the ground surface elevation at each point enable hydrologic terrain analysis calculations to derive much useful hydrologic information. The starting point in many of these calculations is evaluation of topographic slope, because slope dictates the direction of overland water flow as well as playing a role in processes such as the surface energy balance (radiation exposure), erosion and terrain stability, among others. Mathematically, slope is a vector derivative, or gradient, of elevation represented as a continuous surface. ArcGIS provides a method for evaluation of slope that is based on a finite difference approximation. This depends on the elevation values of all surrounding grid points, including those that are higher than the grid point in question. If we are interested in the flow of water downhill from a grid cell, higher neighboring grid cells are not important, so methods for calculating hydrologic slope for each grid cell have been developed that evaluate slope in only a downward direction towards one or more neighboring grid cells.