Lecture 5

Spatial Analysis &
Raster Calculations

GIS in Water Resources
Spring 2015

Spatial Analysis Using Grids

Learning Objectives

• The concepts of spatial fields as a way to represent geographical information
• Raster and vector representations of spatial fields
• Perform raster calculations using spatial analyst
• Raster calculation concepts and their use in hydrology
• Calculate slope on a raster using
  – ESRI polynomial surface method
  – Eight direction pour point model
  – $[D\infty \text{ method}]$
Readings – at http://help.arcgis.com

- Elements of geographic information starting from “Overview of geographic information elements” http://help.arcgis.com/en/arcgisdesktop/10.0/help/00v2/00v2000000300000.htm to “Example: Representing surfaces”

- Rasters and images starting from “What is raster data” http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/009t00000200000.htm to end of “Raster dataset attribute tables”
Two fundamental ways of representing geography are **discrete objects** and **fields**.

The **discrete object view** represents the real world as objects with well defined boundaries in empty space.

\[(x_1, y_1)\]

**Points**

**Lines**

**Polygons**

The **field view** represents the real world as a finite number of variables, each one defined at each possible position.

**Continuous surface**

**Raster and Vector Data**

**Raster** data are described by a cell grid, one value per cell

**Vector**

**Polygons**

**Points**

**Lines**

**Zone of cells**
Raster and Vector are two methods of representing geographic data in GIS

- Both represent different ways to **encode** and **generalize** geographic phenomena
- Both can be used to code **both** fields and discrete objects
- In practice a strong association between **raster and fields** and **vector and discrete objects**

Numerical representation of a spatial surface (**field**)
Triangulated Irregular Networks, TINs

No point in a set of points $P$ lies within a circumcircle of any of the created triangles.

$\Rightarrow$ Delauney Triangulation

“Flipping” Algorithm

This is NOT Delauney           This one IS Delauney

Relation to Voronoi Tessellation

Mark center of each circumcircle.

Then connect each center with those surrounding.

If center of CC is inside triangle, then lines connecting centers are perpendicular to the common edge of two neighboring triangles.

Does not always work out! Has implications for numerical models needing orthogonal grid.
Six approximate representations of a field used in GIS

(A) Regularly spaced sample points
(B) Irregularly spaced sample points
(C) Rectangular Cells
(D) Irregularly shaped polygons
(E) Triangulated Irregular Network (TIN)
(F) Polylines/Contours


A grid defines geographic space as a matrix of identically-sized square cells. Each cell holds a numeric value that measures a geographic attribute (like elevation) for that unit of space.
The grid data structure

- Grid size is defined by **extent, spacing and no data value** information
  - Number of rows, number of column
  - Cell sizes (X and Y)
  - Top, left, bottom and right coordinates
- Grid values
  - Real (floating decimal point)
  - Integer (may have associated attribute table)

**Definition of a Grid**

Number of rows

(X,Y)

Cell size

NODATA cell

Number of Columns
Points as Cells

Line as a Sequence of Cells
Polygon as a Zone of Cells

NODATA Cells
Cell Networks

Grid Zones
Floating Point Grids

Continuous data surfaces using floating point or decimal numbers

Value attribute table for categorical (integer) grid data

Attributes of grid zones
Raster Sampling

(1) values are averages for cells

(2) values are samples at cell corners

(3) values are samples at the grid nodes


Raster Generalization

Largest share rule

Central point rule
Raster Calculator

Cell by cell evaluation of mathematical functions

Example
Precipitation
- Losses (Evaporation, Infiltration)
= Runoff

Runoff generation processes

Infiltration excess overland flow aka Horton overland flow
\[ P - q_e - f = q_o \]

Partial area infiltration excess overland flow
\[ P - q_e - f = q_o \]

Saturation excess overland flow
\[ P - q_e - q_s = q_i \]
Runoff generation at a point depends on

- Rainfall intensity or amount
- Antecedent conditions
- Soils and vegetation
- Depth to water table (topography)
- Time scale of interest

These vary spatially which suggests a spatial geographic approach to runoff estimation

Cell based discharge mapping flow accumulation of generated runoff

- Radar Precipitation grid
- Soil and land use grid
- Runoff grid from raster calculator operations implementing runoff generation formula’s
- Accumulation of runoff within watersheds
Raster calculation – some subtleties

Resampling or interpolation (and reprojection) of inputs to target extent, cell size, and projection within region defined by analysis mask

Analysis mask

Analysis cell size

Analysis extent

Spatial Snowmelt Raster Calculation Example

The grids below depict initial snow depth and average temperature over a day for an area.

One way to calculate decrease in snow depth due to melt is to use a temperature index model that uses the formula

\[ D_{\text{new}} = D_{\text{old}} - m \cdot T \]

Here \( D_{\text{old}} \) and \( D_{\text{new}} \) give the snow depth at the beginning and end of a time step, \( T \) gives the temperature and \( m \) is a melt factor. Assume melt factor \( m = 0.5 \text{ cm/}^{\circ}\text{C/day}. \)

Calculate the snow depth at the end of the day.
New depth calculation using Raster Calculator

“snow100” - 0.5 * “temp150”

Example and Pixel Inspector
The Result

• Outputs are on 150 m grid.

• How were values obtained?

Nearest Neighbor Resampling with Cellsize Maximum of Inputs

- 40 - 0.5 * 4 = 38
- 55 - 0.5 * 6 = 52
- 42 - 0.5 * 2 = 41
- 41 - 0.5 * 4 = 39
Scale issues in interpretation of measurements and modeling results

The scale triplet

a) Extent  b) Spacing  c) Support

Fig. 2.3.1 The scale triplet of measurement scale and modelling scale: (a) extent; (b) spacing; (c) support in either space or time.


Fig. 2.3.6 The effect of sampling for measurement scales not commensurate with the process scale. (a) Spacings larger than the process scale cause aliasing in the data; (b) Extents smaller than the process scale cause a trend in the data; (c) Supports larger than the process scale cause excessive smoothing in the data. The process scale is the period.

Use Environment Settings to control the scale of the output

Raster Calculator “Evaluation” of “temp150”

Nearest neighbor to the E and S has been resampled to obtain a 100 m temperature grid.
Calculation with cell size set to 100 m grid

“snow100” - 0.5 * “temp150”

• Outputs are on 100 m grid as desired.

• How were these values obtained?

100 m cell size raster calculation

Nearest neighbor values resampled to 100 m grid used in raster calculation
What did we learn?

• Raster calculator automatically uses nearest neighbor resampling
• The scale (extent and cell size) can be set under options

• What if we want to use some other form of interpolation?

  From Point
  Natural Neighbor, IDW, Kriging,
  Spline, …

  From Raster
  Project Raster (Nearest, Bilinear, Cubic)

Interpolation

Estimate values between known values.
A set of spatial analyst functions that predict values for a surface from a limited number of sample points creating a continuous raster.

Apparent improvement in resolution may not be justified
Interpolation methods

- Nearest neighbor
- Inverse distance weight
- Bilinear interpolation
- Kriging (best linear unbiased estimator)
- Spline

Nearest Neighbor “Thiessen” Polygon Interpolation

Spline Interpolation
Interpolation Comparison

Figure 2.4. Comparison of interpolation methods for a one-dimensional example: (a) Thiessen method; (b) inverse distance squared; (c) example of overfitting using a sixth-order polynomial; (d) thin plate splines with different tension parameters; (e) kriging (zero nugget, large range); (f) kriging (solid line: large nugget, large range, dashed line: zero nugget, very short range).


Further Reading


Chapter 2. Spatial Observations and Interpolation

Full text online at:
Spatial Surfaces used in Hydrology

Elevation Surface — the ground surface elevation at each point

3-D detail of the Tongue river at the WY/Mont border from LIDAR.

Roberto Gutierrez
University of Texas at Austin
Topographic Slope

- Defined or represented by one of the following
  - Surface derivative $\nabla z$ (dz/dx, dz/dy)
  - Vector with x and y components (S_x, S_y)
  - Vector with magnitude (slope) and direction (aspect) (S, $\alpha$)

ArcGIS “Slope” tool

$$\frac{dz}{dx} = \frac{(a + 2d + g) - (c + 2f + i)}{8 \times x\_mesh\_spacing}$$

$$\frac{dz}{dy} = \frac{(g + 2h + i) - (a + 2b + c)}{8 \times y\_mesh\_spacing}$$

$$\text{rise} = \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}$$

$$\text{deg} = \text{atan} \left( \frac{\text{rise}}{\text{run}} \right)$$
ArcGIS Aspect – the steepest downslope direction

\[
\frac{dz}{dy} = \tan \left( \frac{dz/dx}{dz/dy} \right)
\]

\[
\begin{align*}
\text{Slope} &= \sqrt{(0.229)^2 + (0.329)^2} \\
&= 0.401 \\
\text{atan}(0.401) &= 21.8^\circ \\
\text{Aspect} &= \tan \left( \frac{0.229}{-0.329} \right) = -34.8^\circ + 180^\circ \\
&= 145.2^\circ
\end{align*}
\]

**Example**

\[
\frac{dz}{dx} = \frac{(a + 2d + g) - (c + 2f + i)}{8 \times x_{\text{mesh spacing}}}
\]

\[
= \frac{(80 + 2 \times 69 + 60) - (63 + 2 \times 56 + 48)}{8 \times 30}
\]

\[
= 0.229
\]

\[
\frac{dz}{dy} = \frac{(g + 2h + i) - (a + 2b + c)}{8 \times y_{\text{mesh spacing}}}
\]

\[
= \frac{(60 + 2 \times 52 + 48) - (80 + 2 \times 74 + 63)}{8 \times 30}
\]

\[
= -0.329
\]
Hydrologic Slope (Flow Direction Tool) - Direction of Steepest Descent

\[
\text{Slope: } \frac{67 - 48}{30\sqrt{2}} = 0.45 \quad \frac{67 - 52}{30} = 0.50
\]

Eight Direction Pour Point Model

ESRI Direction encoding
Limitation due to 8 grid directions.

The $D\infty$ Algorithm

Proportion flowing to neighboring grid cell 4 is $\alpha_1/(\alpha_1+\alpha_2)$

Proportion flowing to neighboring grid cell 3 is $\alpha_2/(\alpha_1+\alpha_2)$

Steepest direction downslope

The $D_\infty$ Algorithm

If $\alpha_1$ does not fit within the triangle the angle is chosen along the steepest edge or diagonal resulting in a slope and direction equivalent to $D_8$

$$\alpha_1 = \tan^{-1} \left( \frac{e_1 - e_2}{e_0 - e_1} \right)$$

$$S = \sqrt{\left(\frac{e_1 - e_2}{\Delta}\right)^2 + \left(\frac{e_0 - e_1}{\Delta}\right)^2}$$

$D_\infty$ Example

$$\alpha_i = \tan^{-1} \left( \frac{e_7 - e_8}{e_0 - e_7} \right) = \tan^{-1} \left( \frac{52 - 48}{67 - 52} \right) = 14.9^\circ$$

$$S = \sqrt{\left(\frac{52 - 48}{30}\right)^2 + \left(\frac{67 - 52}{30}\right)^2} = 0.517$$
Key Spatial Analysis Concepts

- Contours and Hillshade to visualize topography

Zonal Average of Raster over Subwatershed
Subwatershed Precipitation by Thiessen Polygons

- Thiessen Polygons
- Feature to Raster (Precip field)
- Zonal Statistics (Mean)
- Join
- Export to DBF (Excel)

Subwatershed Precipitation by Interpolation

- Kriging (on Precip field)
- Zonal Statistics (Mean)
- Join
- Export
Runoff Coefficients

- Interpolated precip for each subwatershed
- Convert to volume, P
- Sum over upstream subwatersheds
- Runoff volume, Q
- Ratio of Q/P

Summary Concepts

- **Grid (raster) data structures** represent surfaces as an array of grid cells
- **Raster calculation** involves algebraic like operations on grids
- **Interpolation and Generalization** is an inherent part of the raster data representation
Summary Concepts (2)

- The elevation surface represented by a grid digital elevation model is used to derive surfaces representing other hydrologic variables of interest such as
  - Slope
  - Drainage area (more details in later classes)
  - Watersheds and channel networks (more details in later classes)

Summary Concepts (3)

- The eight direction pour point model approximates the surface flow using eight discrete grid directions.
- The $D_\infty$ vector surface flow model approximates the surface flow as a flow vector from each grid cell apportioned between down slope grid cells.