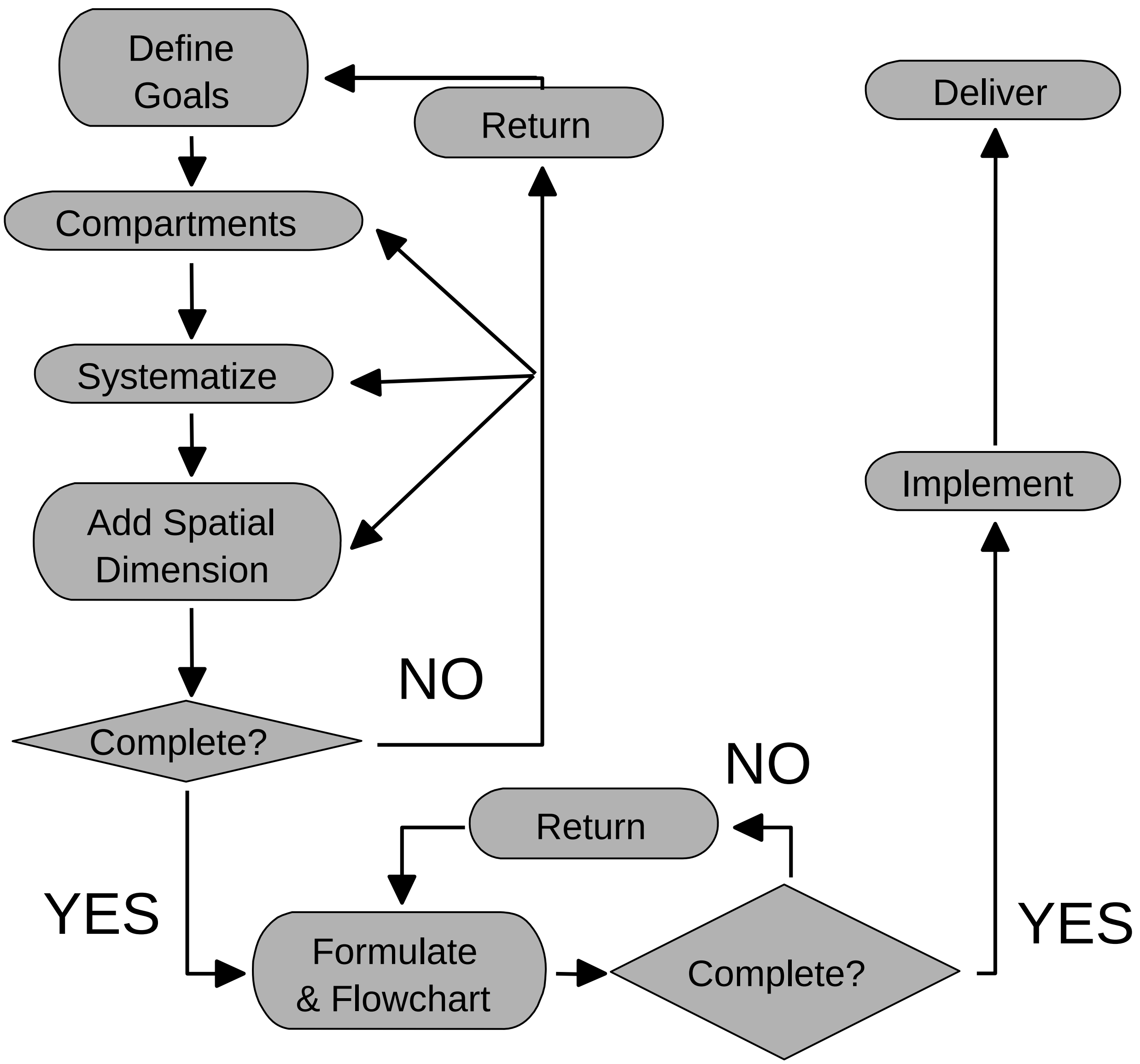
Building Spatial Models II

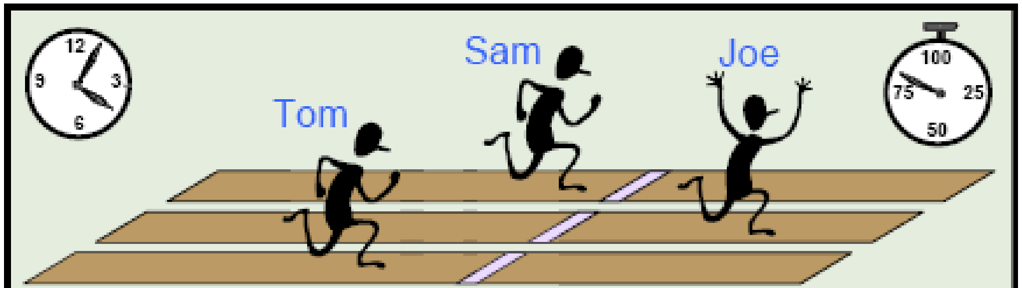
The General Modeling Process



Data Types and Math in Modeling

Type	Examples	Legal math
Nominal	ID, Landuse code, Phone number	=
Ordinal	Importance, Order of completion	<, =, >
Interval	Time of day, Temperature, pH level	<, =, >, +, -
Ratio	Age, Distance, Weight, Counts	<, =, >, +, -, *, /

- **Valid** mathematical operations depends on the **data type**



Runner	#72 – Tom	#43 – Sam	#10 – Joe	Nominal
Finished	3rd	2nd	1st	Ordinal
Time-of-day	4:05:09	4:05:07	4:05:03	Interval
Elapsed time	81 sec	79 sec	75 sec	Ratio

Scales of Measurement

- **Attribute data** can be divided into four types

1. The Nominal Scale
2. The Ordinal Scale
3. The Interval Scale
4. The Ratio Scale

↓ As we progress through these scales, the types of data they describe have increasing information content

The Nominal Scale

- **Nominal data** - information that is simply grouped into categories on the basis of qualitative considerations

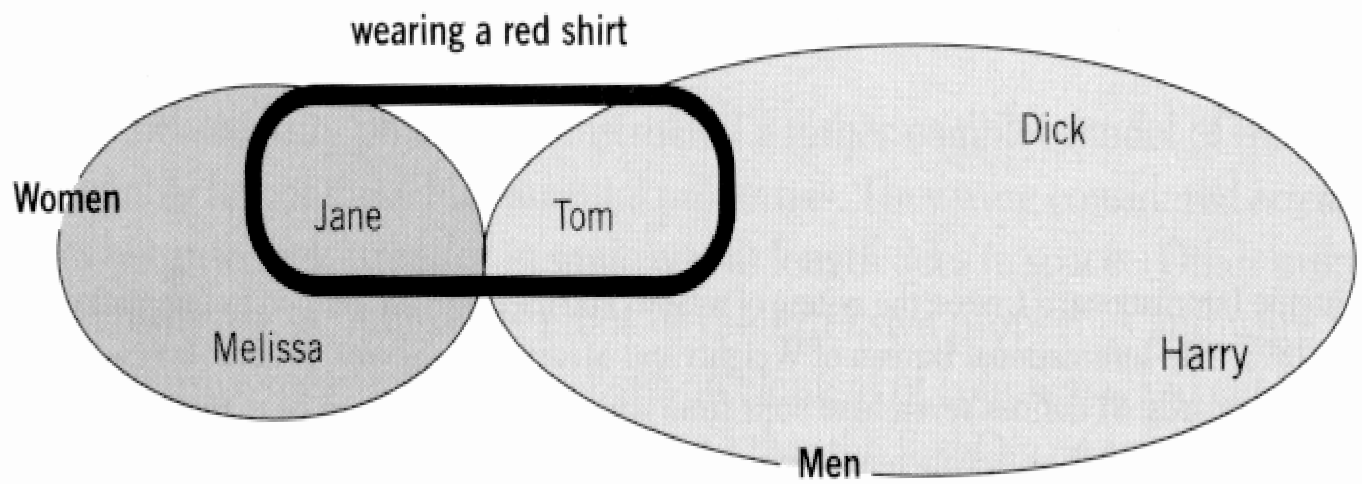


Figure 1-2: Nominal measures are not on scales at all. They create categories that can be treated as sets.

The Ordinal Scale

- **Ordinal data** - grouped by rank on the basis of some quantitative measure

Ordinal

Order of arrival of contestants	*Women's race*	*Men's race*
First	Jane	Tom
Second	Melissa	Dick
Third	Leila	Harry

Figure 1-3: Strictly ordinal scales can arise from a total ordering, but ordinal scales may also arise from partial orderings.

The Interval Scale

- **Interval data** - information that can be arranged using a standard scale along which operations of addition and subtraction have meaning

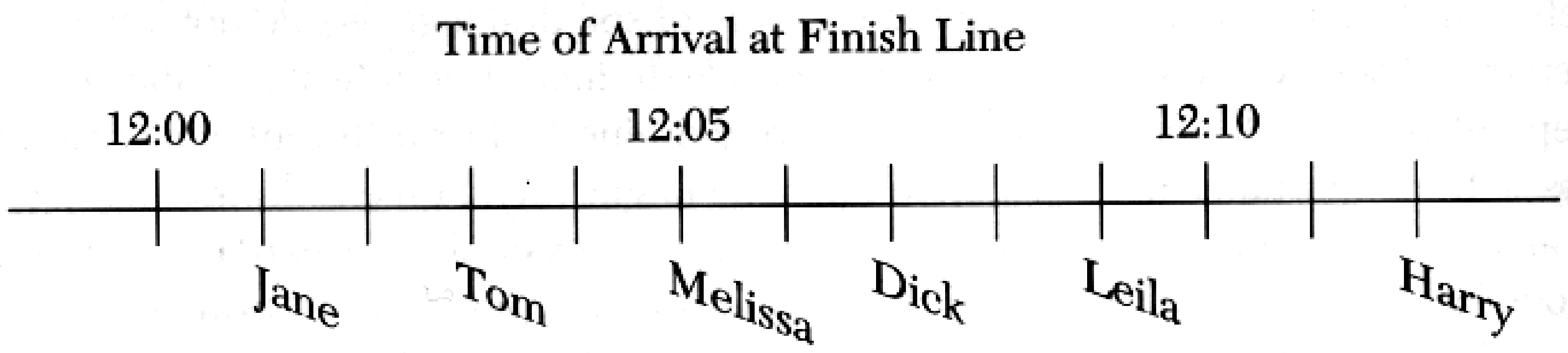


Figure 1-4: Interval scales mobilize a number line, but the origin and the unit are arbitrary.

The Ratio Scale

- **Ratio data** - other type of continuous data that can be arranged along a scale but, in addition, the scale begins at a non-arbitrary zero point
  – **Multiplication and division** can be employed with ratio data to consider proportions and magnitudes

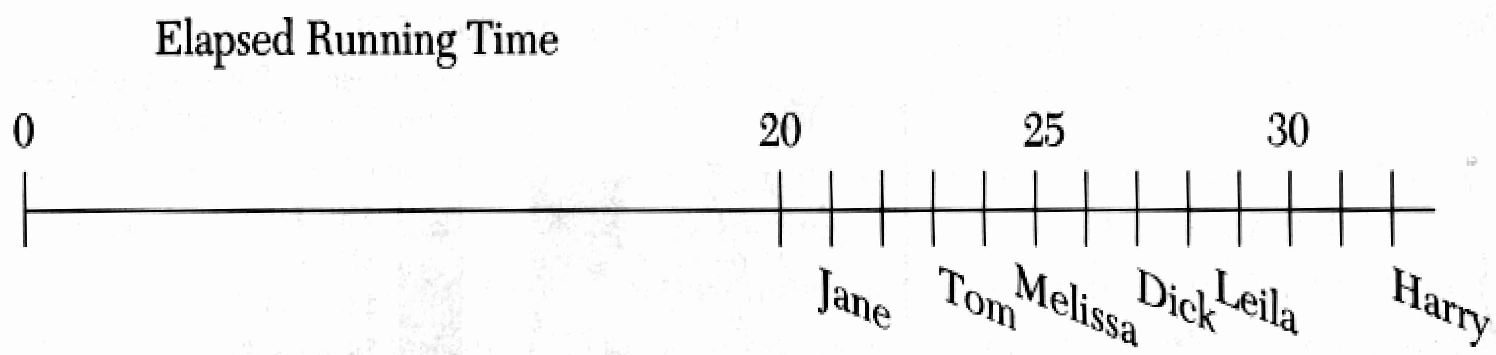
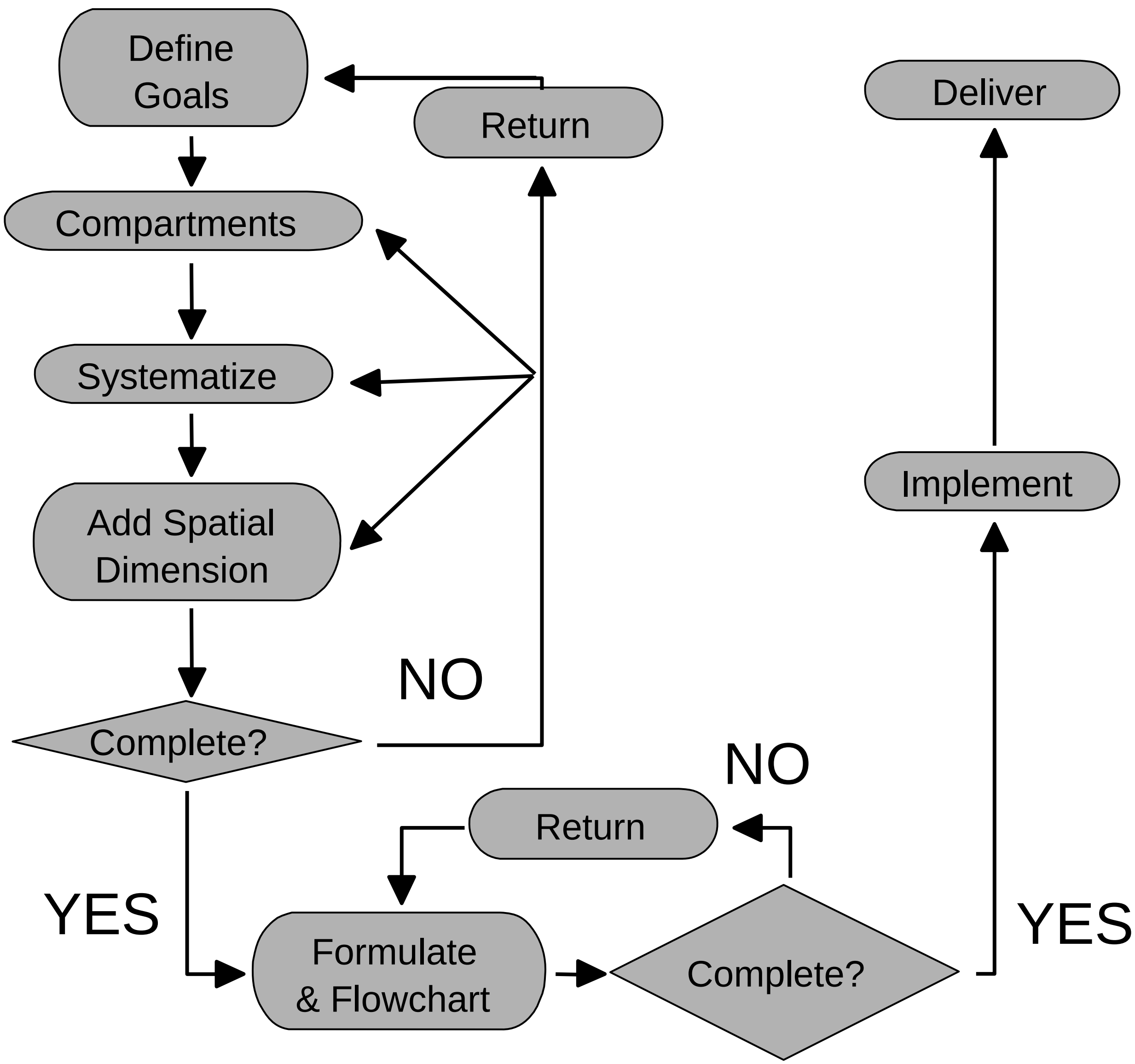


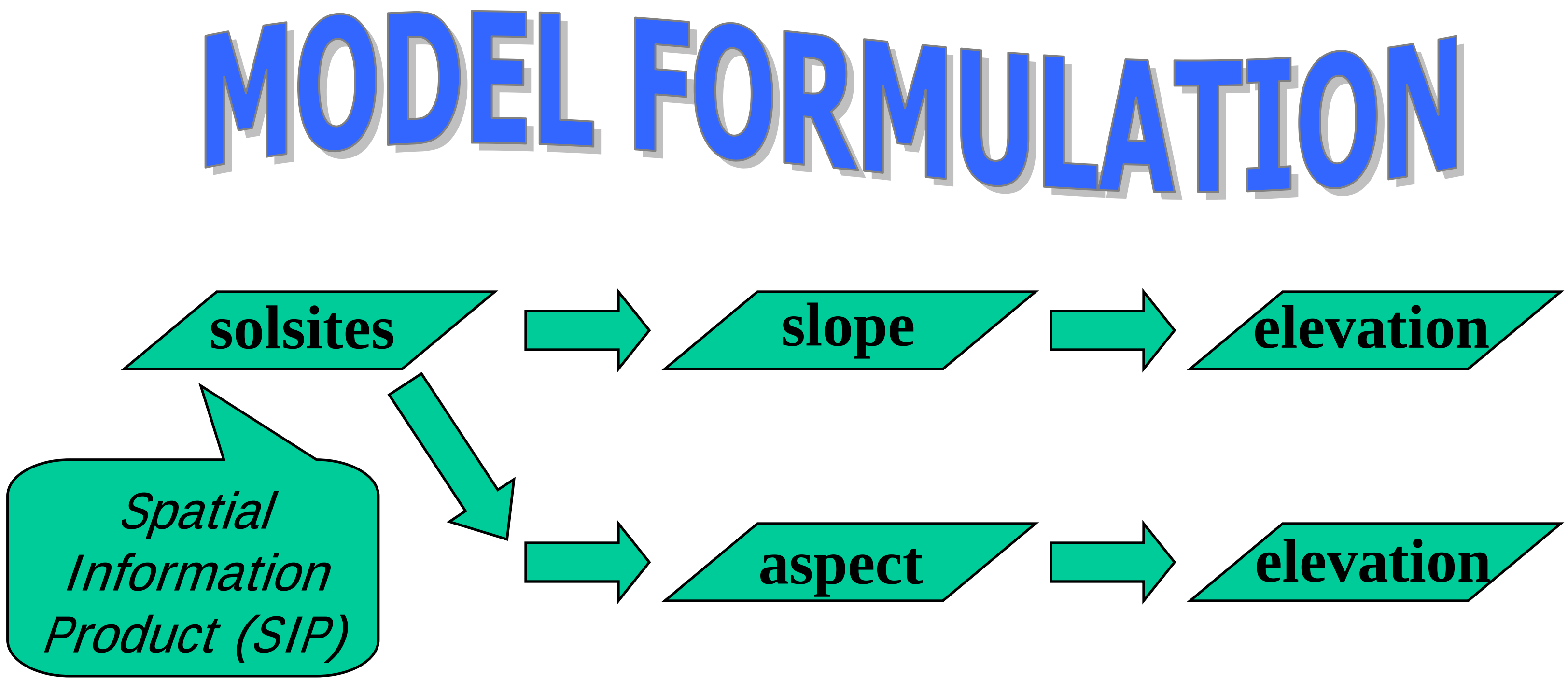
Figure 1-5: Ratio scales, the classical ideal for physical measurement, have a true origin and an arbitrary unit of measure.

The General Modeling Process



Flowcharting

- Helps to **identify and isolate** model elements
  - Example: solar facility sitting

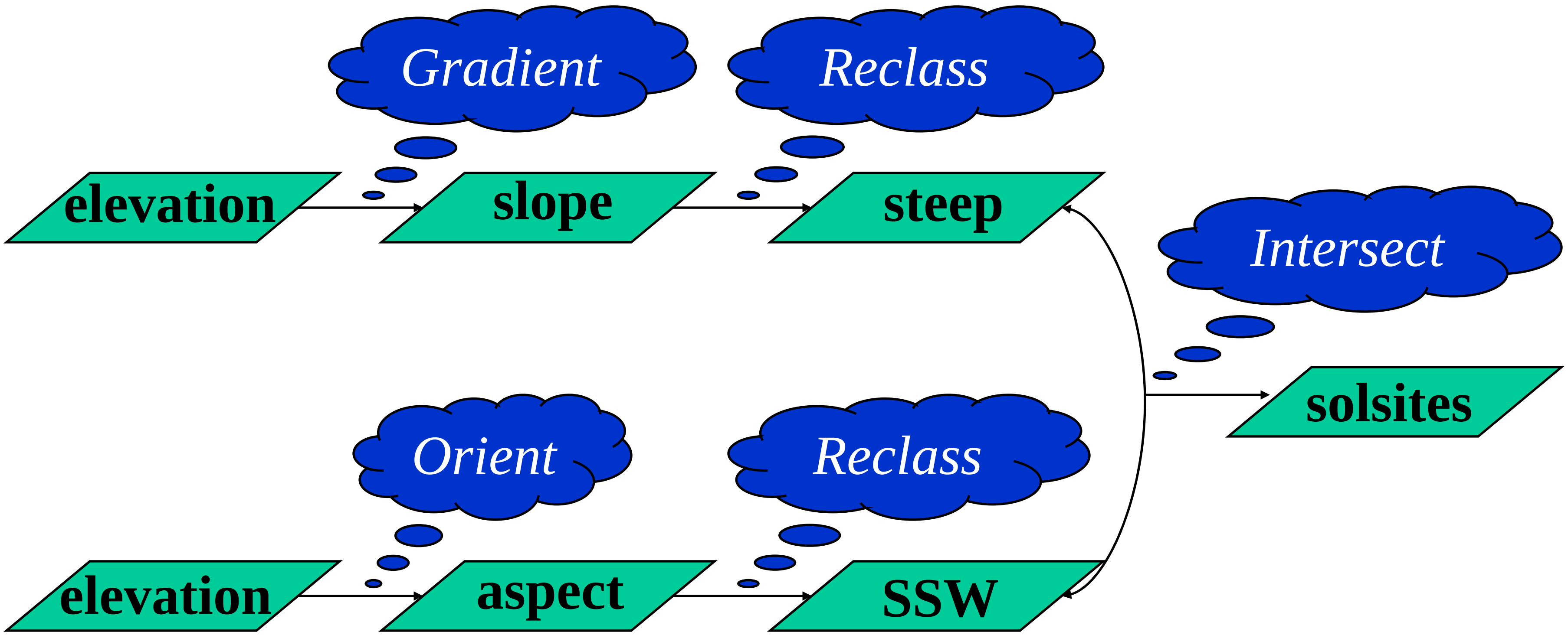


Descriptive Models

- **First steps** in modeling are **descriptive**
  - Describe what “is” or …
  - Describe what “could be”
- **Later** shifts to what **“should be”**
  - More **prescriptive** intent
- Most require a **synthetic approach**
  - Synthesis of data:
    - To **expose significant facts** using data, or…
    - To **express a meaning** a user may attribute to a set of facts
      - Importance of **certain characteristics** over others
- Requires **formulation and implementation**

Flowcharting (in Reverse)

MODEL IMPLEMENTATION

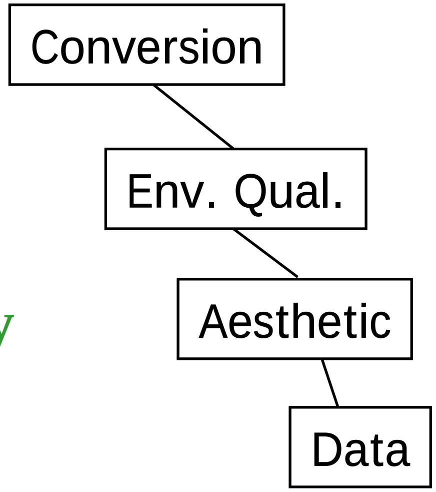


Formulating Descriptive Models

- Start with what data are **needed to answer the question** … **NOT** with what data are available
    - e.g. suppose we are in charge of building a new highway:
        - what factors **should we consider? VS.**
        - We have elevation, water, vegetation data **can we consider**? We might limit ourselves to physical and cost considerations in our routing … what about people?
    - Soon apparent **some factors recur**
        - e.g. construction cost, user appeal, environmental impact and maintenance all arise with respect to vegetation, hydrographic or demographic considerations
        - List criteria **without regard for where they come from**
        - List may be **exhaustive** – probably will show gaps

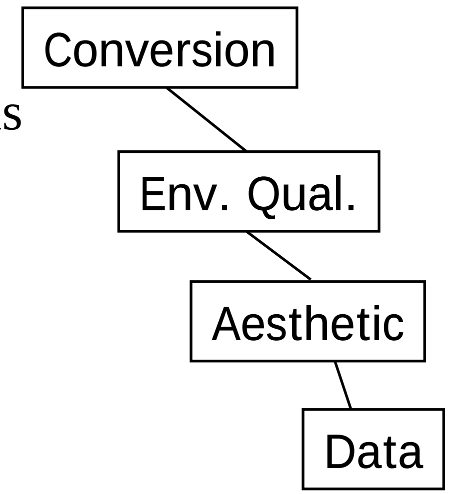
Organizing Data Hierarchically

- **One method** of compiling a list of criteria
  - Some items **general**, others **specific**
  - e.g. aesthetic quality may have several **sub-issues**
    - e.g. aural or visual qualities
    - Or it may be part of a larger category called environmental quality
  - This process **improves clarity and consistency**
    - Helps **identify missing elements**
    - May require **inductive** and/or **deductive** logic
    - Affords a **better understanding** of the model under consideration

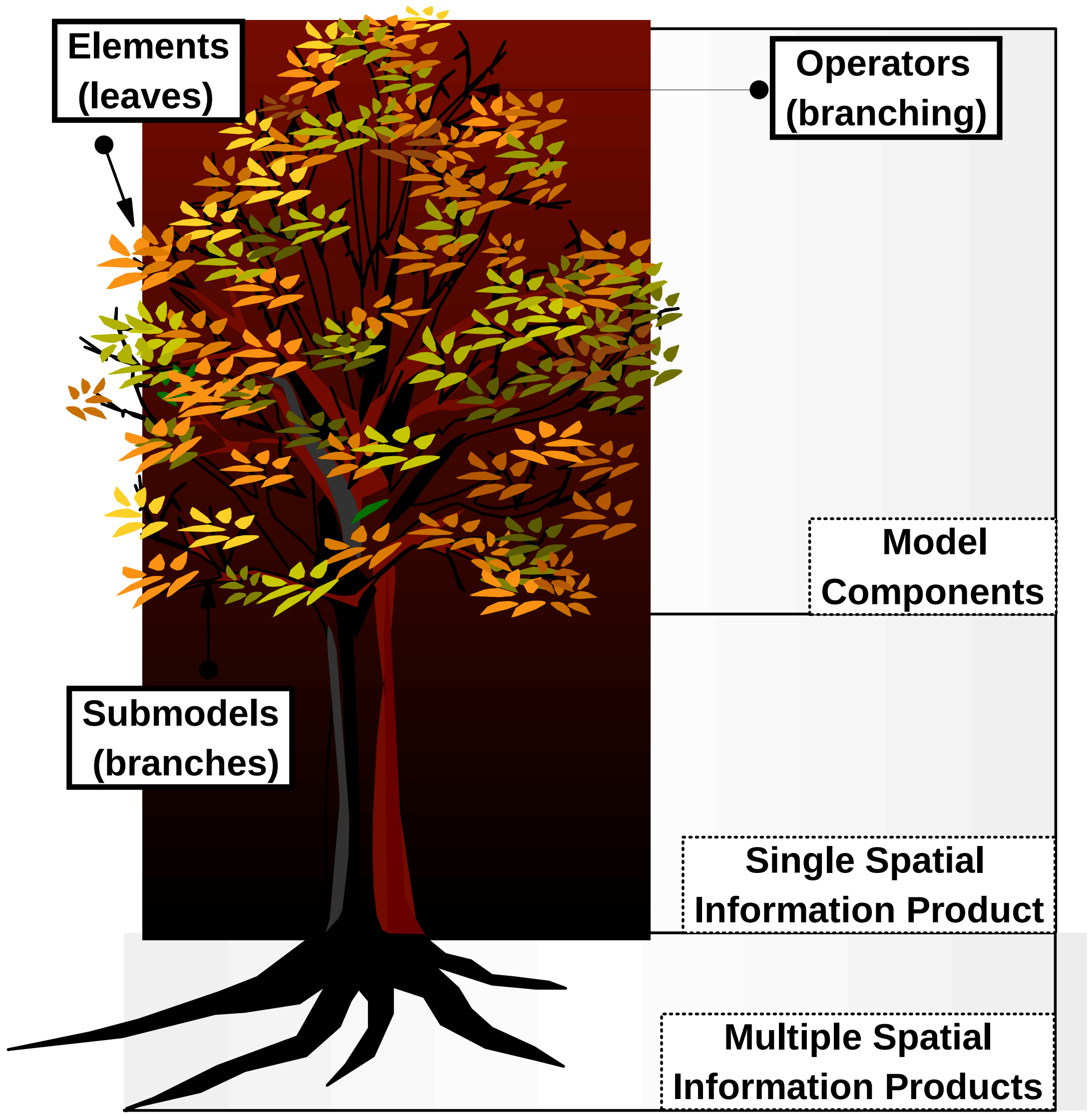


Organizing Data Hierarchically

- Best to **start with a hierarchy** in mind rather than reorganizing an existing list
  - May also use **deductive and inductive** logic
  - Deductive **generally** more useful
    - **Start with intended outcome** then figure out what is required to get there
    - At each hierarchical level, **components should be**:
      - **Inclusive (exhaustive?)** – account for all aspects
      - **Mutually exclusive** – encompass a single set of concerns to reduce redundancy
      - **Meaningful** – components act or are acted on as a whole
  - There is often a struggle between **conceptual elegance** and **utility**



Dana Tomlin's Hierarchical Model



Implementing Descriptive Models

- Flowchart **terminates** in leaves
  - Acquirable data:
    - **Roots** (anticipated [spatial information] products)
    - Products dictate **organization and content** of GIS database
    - Certain components can be **generated** from others
    - Some **similar components** might be **combined**
    - Once the leaves have been expressed as **elements** the model can be **implemented** as a **procedure**
      - **Pruning** the tree by **generating intermediate** themes / GRIDs / coverages

Implementing Descriptive Models

- Each theme / GRID / coverage should **map** to one of the hierarchical components
  - This is **not always easy** to manage ... how do you assign specific numbers to…
    - Habitat quality or historical significance?
      - Specifically, how to obtain numbers assigned to specific locations
  - One method:
    - Use **observable characteristics that act as indicators** of such qualities: e.g. dominant vegetation as a surrogate of wildlife habitat quality, or number of homes sold as an indicator of neighborhood instability

Implementing Descriptive Models

- **Separating** the subjective and synthetic from the analytical portions of the model is **important** because…
  - **Subjective assertions** embodied in the formulation must be **explicit in the implementation**
    - Synthetic & subjective parts
      - Concern **WHAT** is **important**
    - Analytical parts
      - Concern **HOW** they are important and **HOW IMPORTANT** they are
    - Separating the two parts **separates the what from the how**
      - Requires themes as common media
      - Also requires standardization in how they are combined

Record Keeping

- In the past, **intermediate maps** needed to be saved to allow us to re-run a model
- **No longer necessary** with newer software
  - Model Builder Flowcharts
  - ERDAS Spatial Modeler Software
  - Sequences of modeling commands maintained
    - AML (Arc Macro Language) , SML (spatial modeling language)

Documentation: Metadata

- **Metadata**: Data about data – extends the data dictionary
- Federal Government Digital Data Committee
  - Created **Spatial Data Transfer Standards** (SDTS)
  - **More** than just for data set creation and transfer
  - **Impacts the modeling process** itself
    - Particularly with regard to **model interoperability**
  - More **tools** becoming available all the time
    - **Search the web** for particular spatial data by metadata

Raster Analysis and Functions II

- **Global Functions – Hydrologic**
  - **Watershed -** A geomorphically distinct **landscape unit** defined by topographic boundaries, or drainage 'divides' that acts as a spatially discrete hydrological system

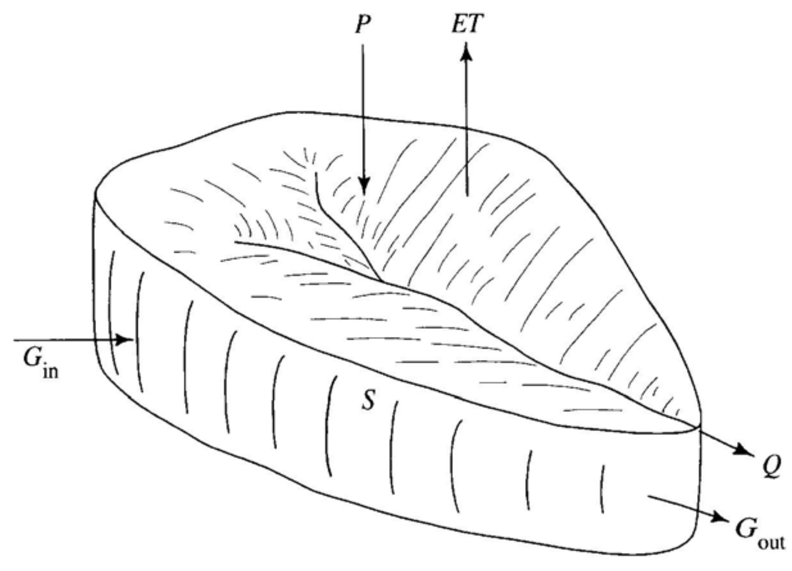
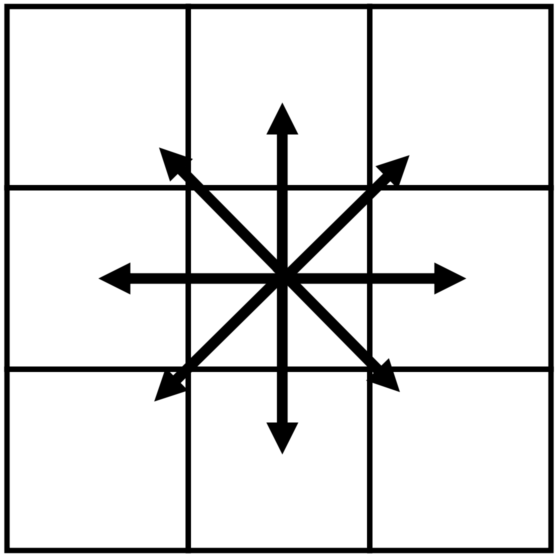


FIGURE 2-3
Schematic diagram of a watershed, showing the components of the regional water balance: P = precipitation, ET = evapotranspiration, Q = stream outflow, G_{in} = ground-water inflow, G_{out} = ground-water outflow.

D8 Analysis Sequence

- Assume we now have a raster DEM and we want to use it **find a watershed and drainage network** through D8 analysis
- We can follow this **sequence of analysis** steps, each of which involves a neighborhood analysis operation:
  - Fill Sinks
  - Slope
  - Aspect
  - Flow Direction
  - Flow Accumulation
  - StreamLink & StreamOrder
  - Watershed

D8 Analysis



Flow Direction

- Flow Direction evaluates the **direction of steepest decent** for each cell in the grid by **comparing a cell with its eight neighbors** in the following fashion:
  - drop = change in z value / distance * 100
  - Note that diagonal neighbors are 1.414214 times as far away as 4-connected orthogonal neighbors
- ArcGIS **encodes** the resulting direction of steepest decent in the grid using **the following scheme**:

32	64	128
16	X	1
8	4	2

- For example:

From ArcView 3.2 Help

78	72	69	71	58	49
74	67	56	49	46	50
69	53	44	37	38	48
64	58	55	22	31	24
68	61	47	21	16	19
74	53	34	12	11	12

elevGrid

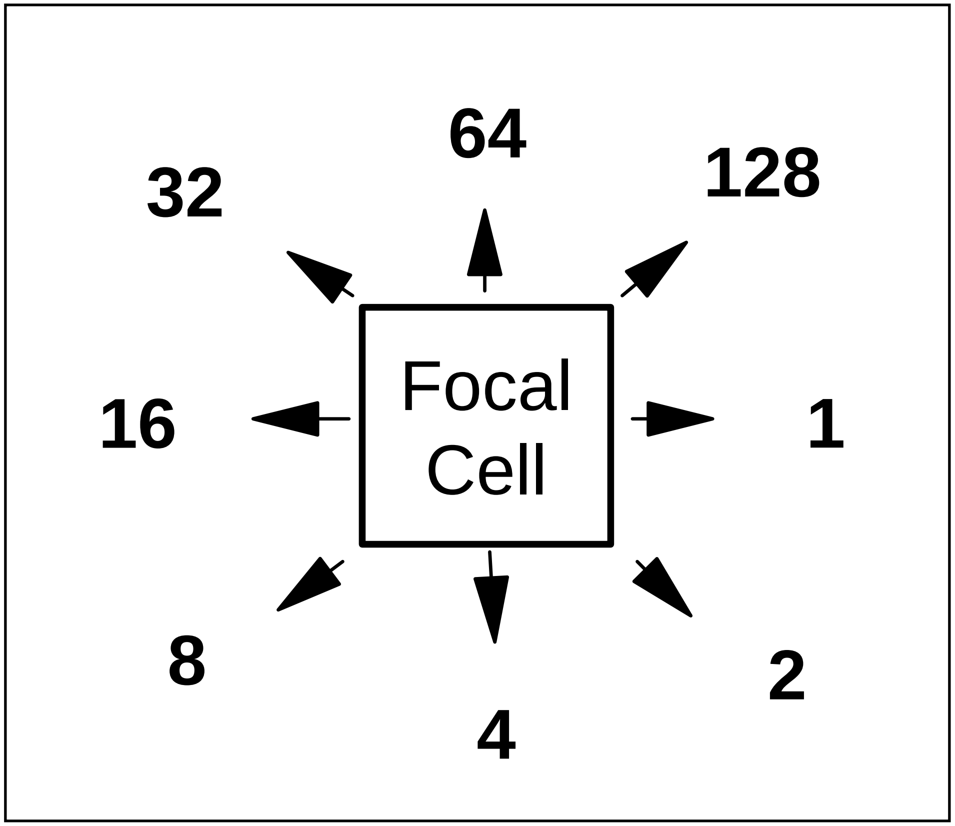
=

2	2	2	4	4	8
2	2	2	4	4	8
1	1	2	4	8	4
128	128	1	2	4	8
2	2	1	4	4	4
1	1	1	1	4	16

flowGrid

Representing Flow Direction

- ESRI's GRID **flow direction encoding** method:
  - If the **change in z** value **is the same** both to the right (1) and down (4) the flow direction is 1+4 for a final value of 5

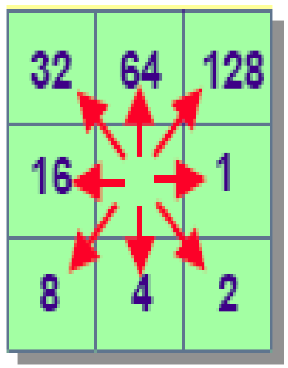


Flow direction

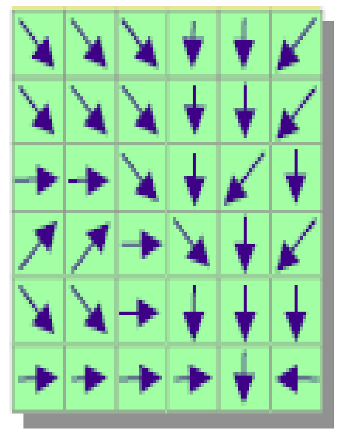
- Created from an elevation surface
  - Determines the direction water flows through a cell

78	72	72	71	58	49
74	67	56	49	46	50
69	53	44	37	38	48
64	58	55	22	31	24
68	61	47	21	16	19
74	53	34	12	11	12

Elevation



Flow direction map



2	2	2	4	4	8
2	2	2	4	4	8
1	1	2	4	8	4
128	128	1	2	4	8
2	2	1	4	4	4
1	1	1	1	4	16

Flow direction

- Required by all other hydrologic analysis

Flow Accumulation

- Flow accumulation find the **number of cells that drain to any cell** in the grid, taking the flow direction grid as input:
  - Output cells with a **high flow accumulation** are areas of concentrated flow and may be used to **identify stream channels**.
  - Output cells with a **flow accumulation of 0** are local topographic highs and may be used to **identify ridges**.
- For example:

From ArcView 3.2 Help

2	2	2	4	4	8
2	2	2	4	4	8
1	1	2	4	8	4
128	128	1	2	4	8
2	2	1	4	4	4
1	1	1	1	4	16

flowGrid

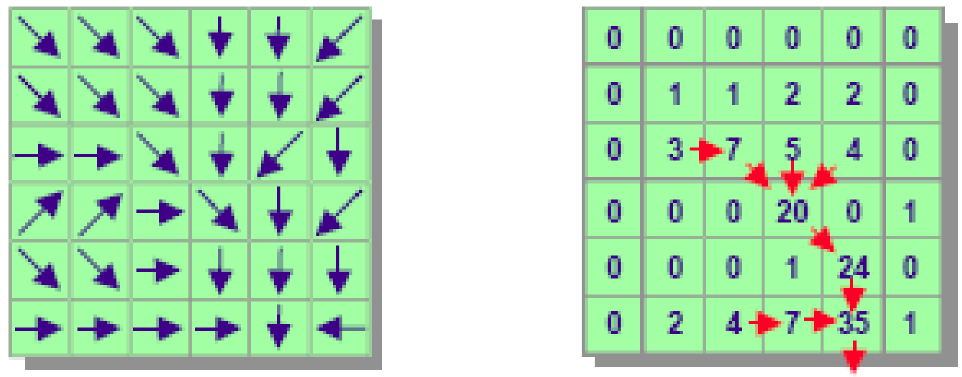
=

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	2

accumGrid

Flow accumulation

- Accumulates uphill water into each cell
  - Based on flow direction

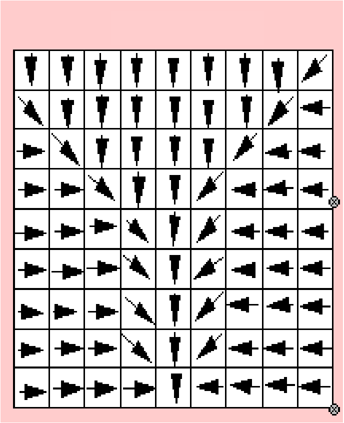


Flow direction   Flow accumulation

- Provide weight layer to model actual rainfall
- High accumulations = streams

Flow Accumulation Types

- **Non-weighted**
  - Simple condition
- **Weighted**
  - Might include precipitation input



a) non-weighted

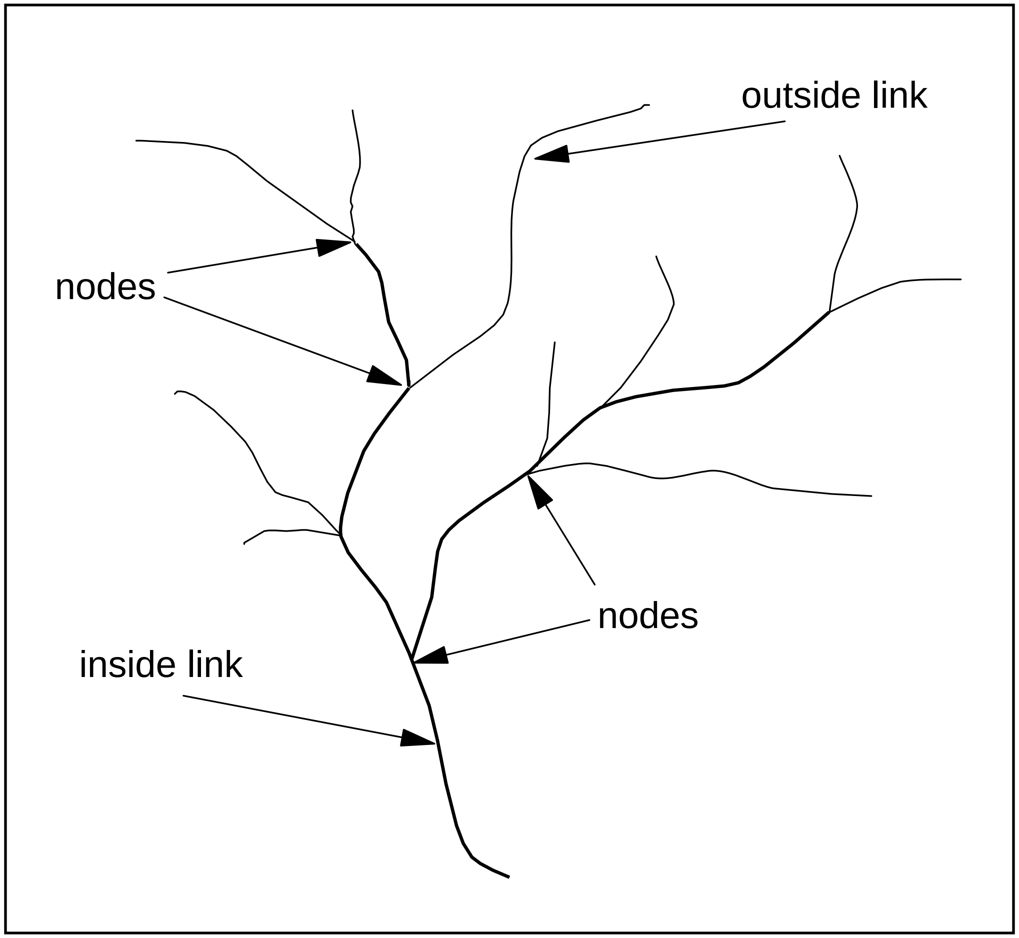
0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	3	0
0	3	1	1	1	1	5	1	0
0	1	5	1	1	7	1	1	0
0	1	1	7	9	1	1	1	0
0	1	1	1	16	1	1	1	0
0	1	1	1	18	1	1	1	0
0	1	1	1	20	1	1	1	0
0	1	1	1	24	1	1	1	0

b) weighted

0	0	0	0	0	0	0	0	0
0	2	2	2	2	2	2	6	0
0	6	2	2	2	2	10	2	0
0	2	10	2	2	14	2	2	0
0	2	2	14	18	2	2	2	0
0	2	2	2	32	2	2	2	0
0	2	2	2	36	2	2	2	0
0	2	2	2	40	2	2	2	0
0	2	2	2	48	2	2	2	0

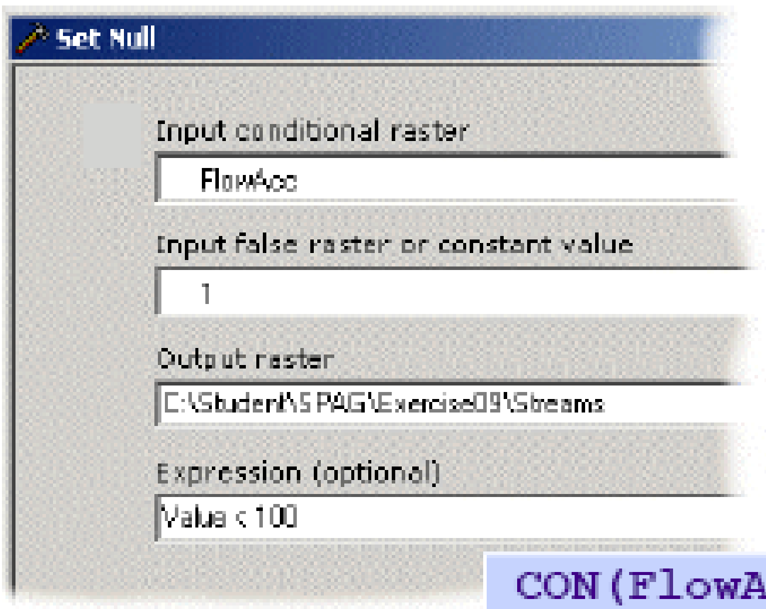
Representing Streams in GIS

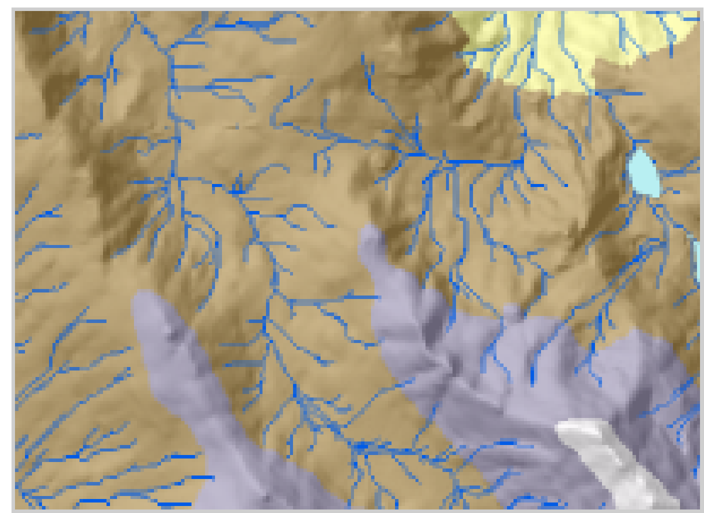
- A **stream network** is just a particular flavor of **network**:
  - **Links** (branches) intersect at **nodes**
  - Outside (or terminal) links have **no tributaries**



Creating streams

- Extract high-value cells from flow accumulation
  - Can use a conditional tool: Con or Set Null
- Choose a threshold value (e.g., accumulation > 100)
  - A low threshold creates more streams

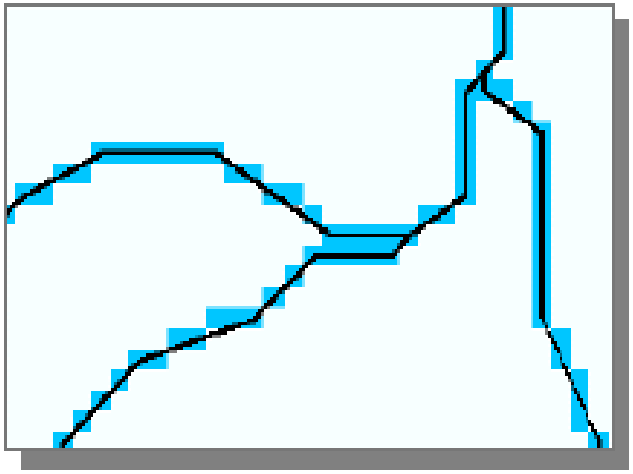




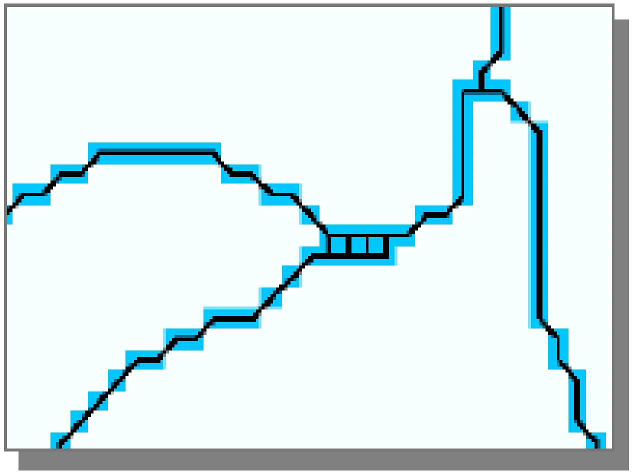
```
CON(FlowAcc GT 100, 1)
SETNULL(FlowAcc LT 100 , 1)
```

Raster streams to features

- Two raster-to-vector conversion functions:



STREAMSHAPE

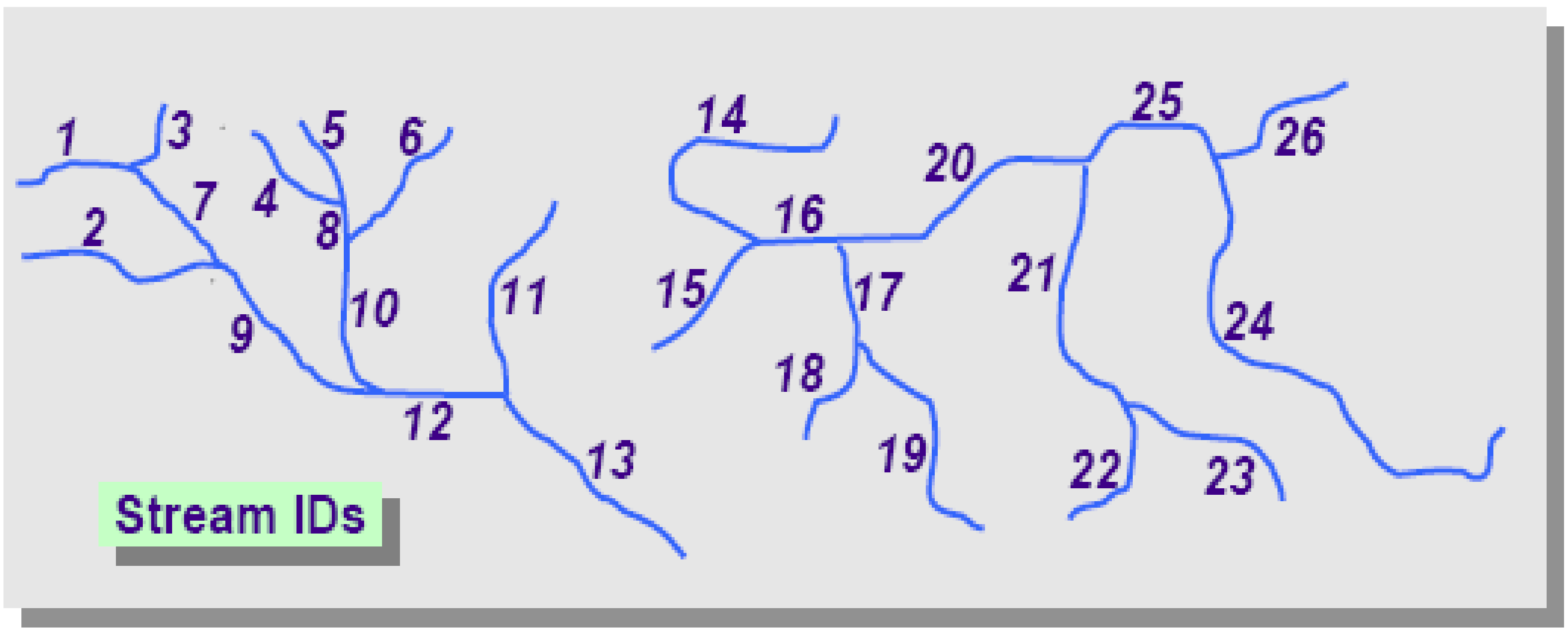


GRIDLINESHAPE

- *STREAMSHAPE* is designed for stream networks
  - Handles parallel stream segments properly
  - Requires stream network and flow direction inputs

Stream link IDs

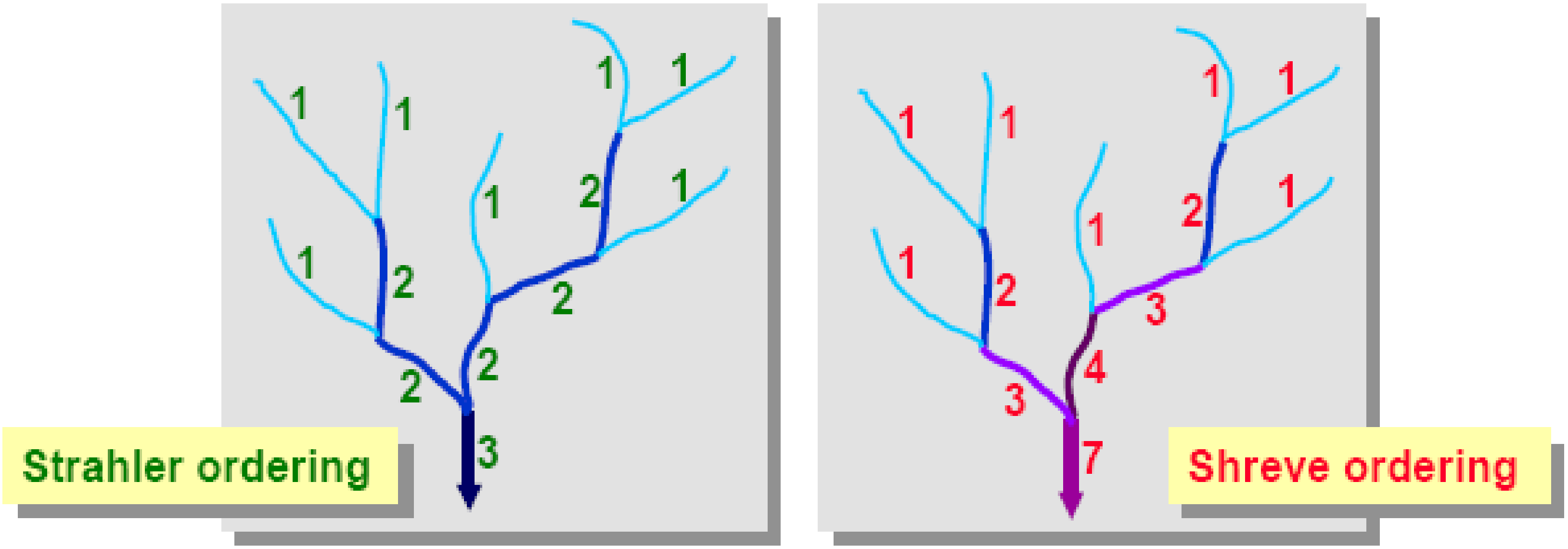
- Assigns unique ID to each *stream segment*
  - May use to link attributes



- Requires stream network and flow direction inputs

Stream order

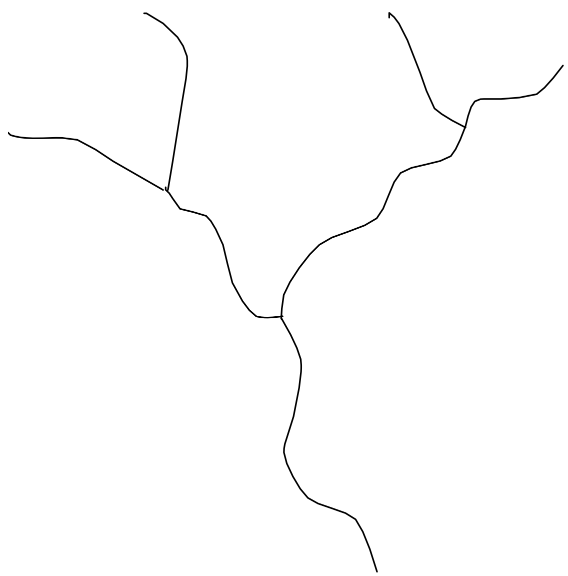
- A relative ranking of streams
  - Based on the number of tributaries
  - Two methods: Shreve or Strahler



- Requires stream network and flow direction inputs

Bifurcation: Strahler Method

- Assigns order of **1 for all external links**
  - Two links of **equal order must meet** to produce a higher order link



		1				1			
		1				1	1	1	
1	1	1					2		
		2	2			2	2		
			2		2	2			
			2	3	2				
				3					
				3					
				3	3				
					3				

☐ = No Data

Bifurcation: Shreve Method

- Note how **highest magnitude is 7** in this example
  - **Number of outside links** that link to the trunk stream
  - Magnitudes are **additive** (1 link + 3 link = 4 link)

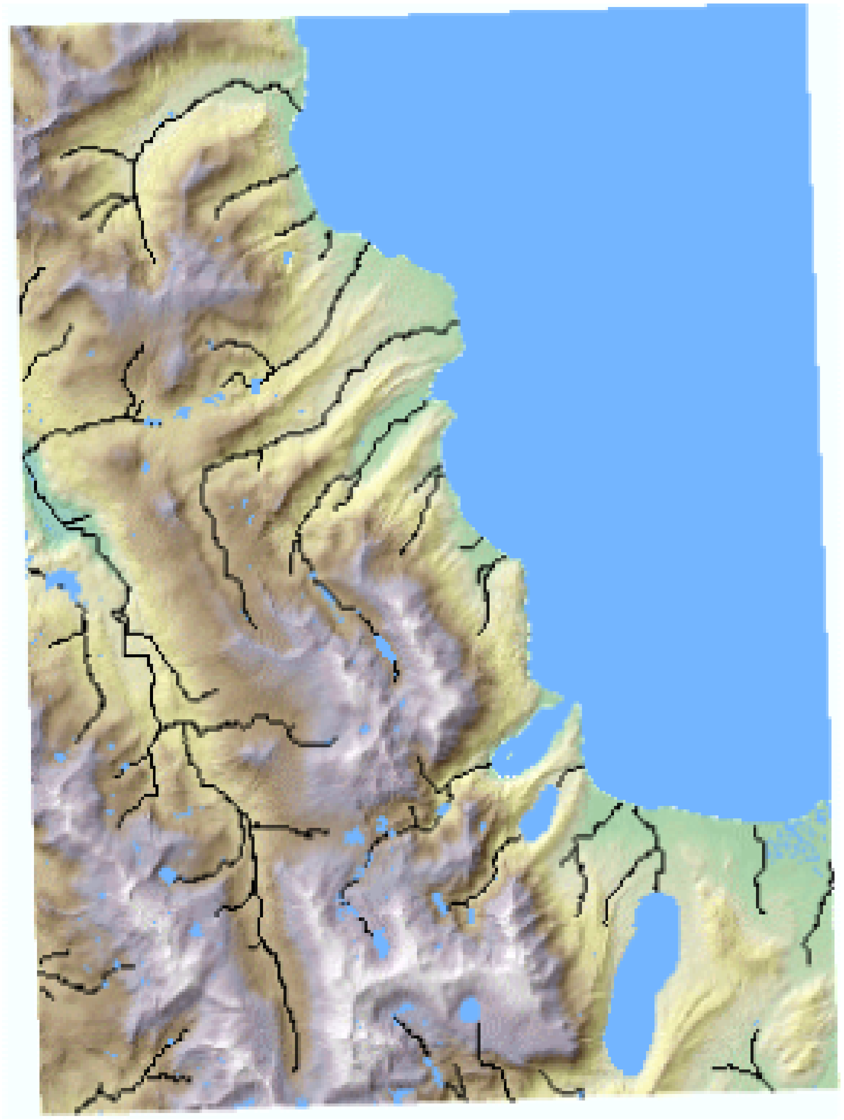
	1			1	1				1				1
	1	1		1					1			1	1
		1	2	1					1	1	2	1	
			2	2						2	2		
1	1			2	2				2	2		1	1
	1	1	1		2			2	2			1	
			1	1	3			2			2	1	
					3	5	2	2			2	1	1
						5				2	2		
						5	5	2	2	2			
							7						
							7	7					
								7					
								7					

☐ = No Data

- Surface hydrology tools
  - Creating hydrologically correct topographic surfaces
  - Defining flow direction over the surface
  - Defining stream networks
    - Analyzing streams
  - Defining drainage basins
  - Determining flow length and tracing rain drops

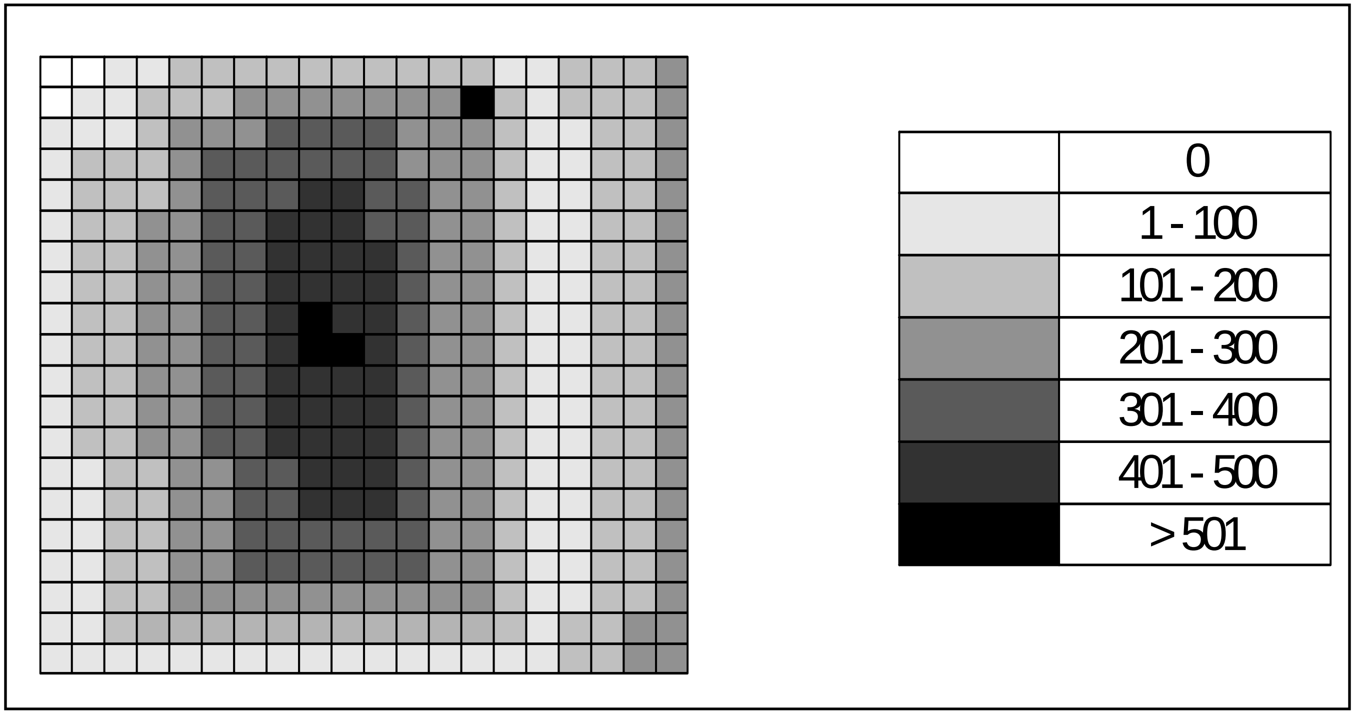
Topographic surfaces

- Use published data (e.g., DEM, DTED)
- Create your own (interpolation)
- Both have errors (sinks)
- Prepare before use (fill sinks)



DEM Problems: Peaks

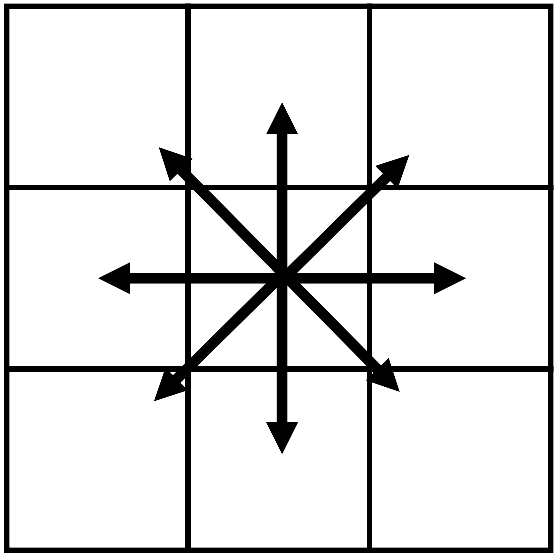
- **Unexplained** high values (spikes)
  – Don't relate to actual surface features



D8 Analysis Sequence

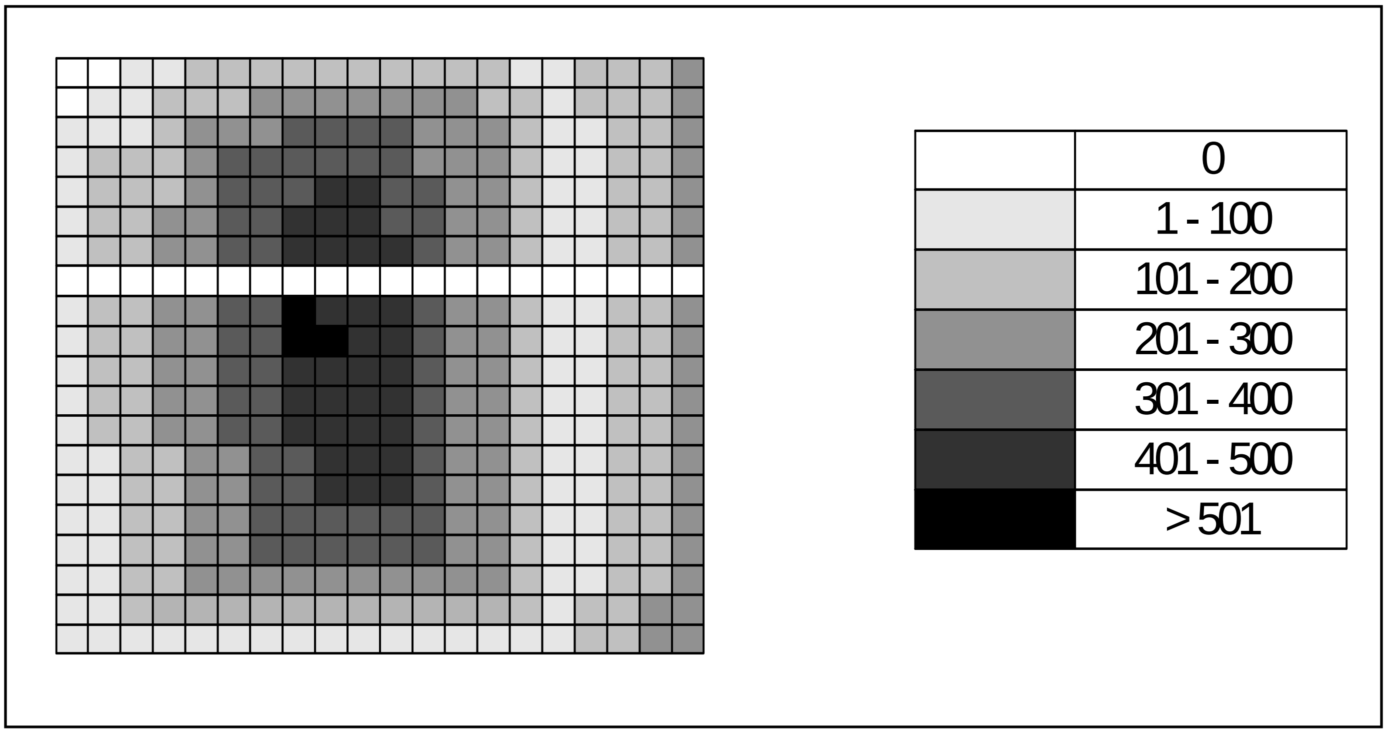
- Assume we now have a raster DEM and we want to use it **find a watershed and drainage network** through D8 analysis
- We can follow this **sequence of analysis** steps, each of which involves a neighborhood analysis operation:
  - Fill Sinks
  - Slope
  - Aspect
  - Flow Direction
  - Flow Accumulation
  - StreamLink & StreamOrder
  - Watershed

D8 Analysis



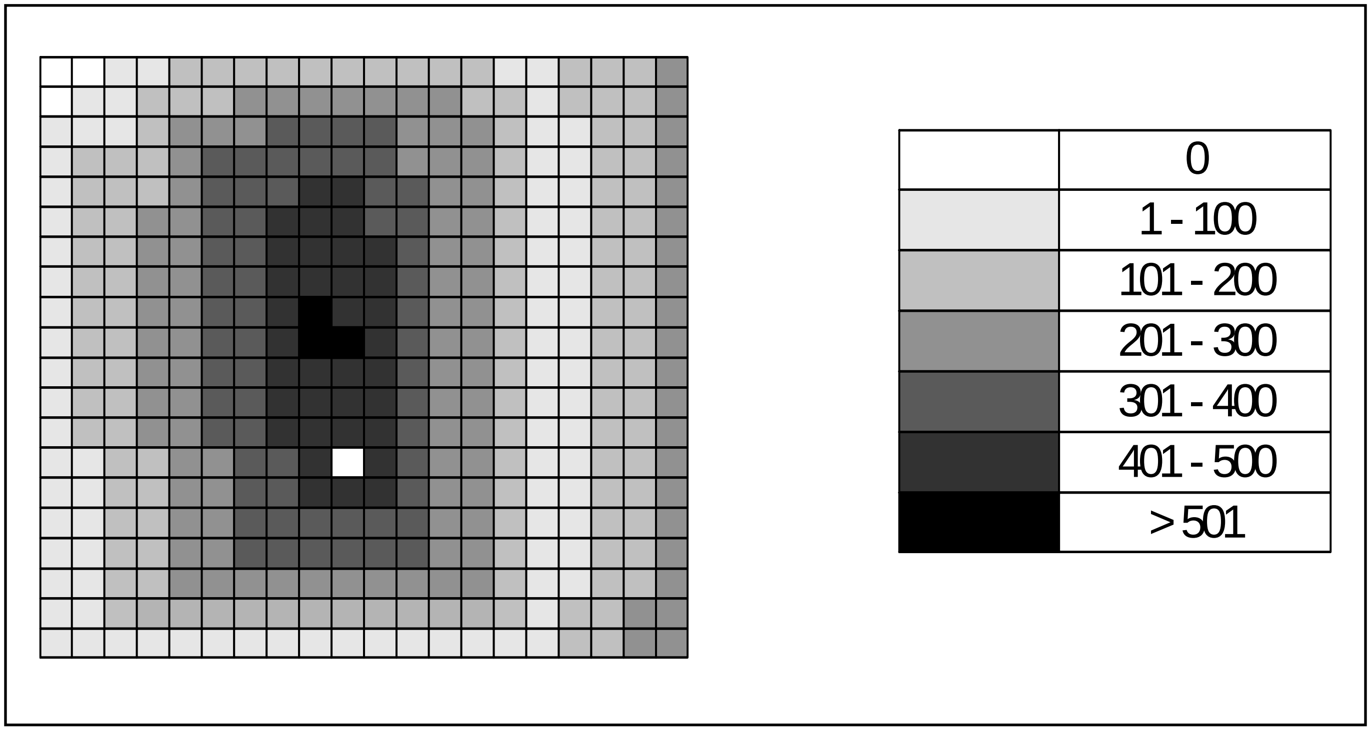
DEM Problems: Striping

- **Sudden shift** in elevation
  - Offset on input device



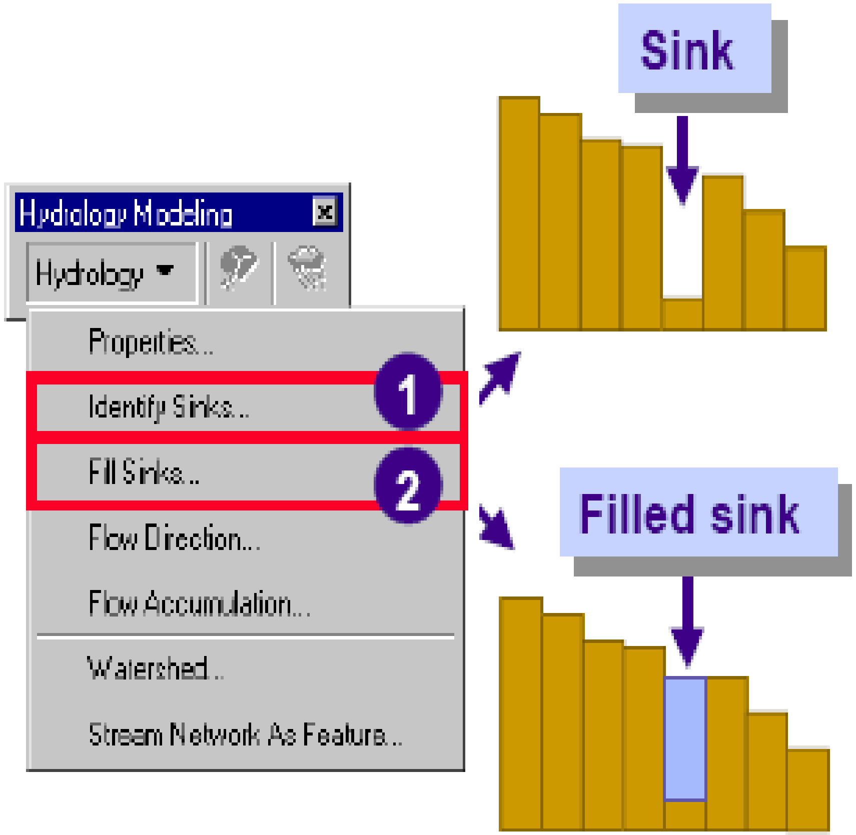
DEM Problems: Pits

- **Unexplained** low values (pits, sinks, etc.)
  - Don’t relate to surface conditions

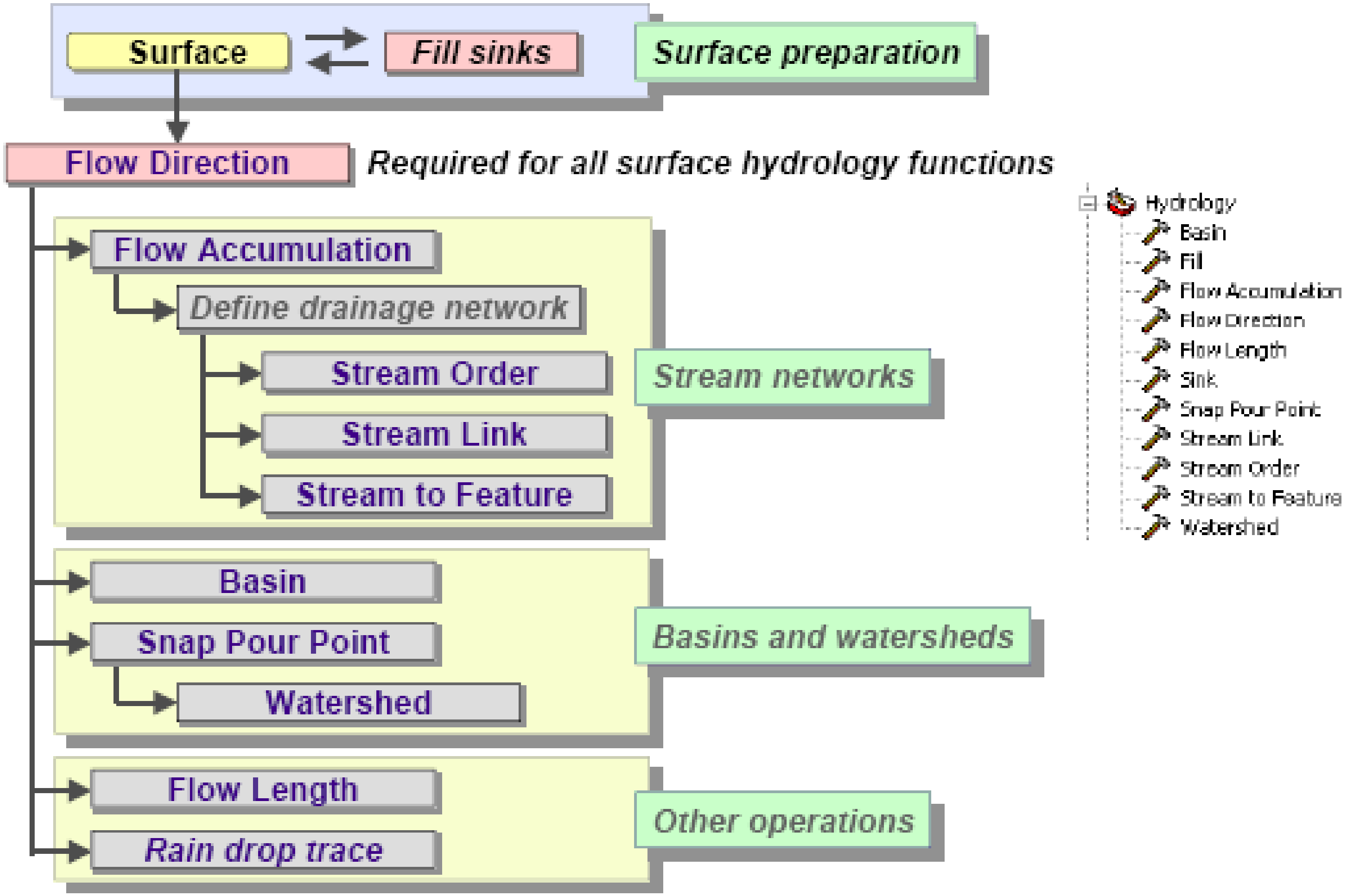


Identifying and filling sinks

- Sinks
  - Surrounded by higher cells
  - Stop surface flow
  - Usually data errors
- Identifying sinks
  - Need flow direction
- Filling sinks
  - Iterative process: May create new sinks
  - Hydrology toolbar is easiest

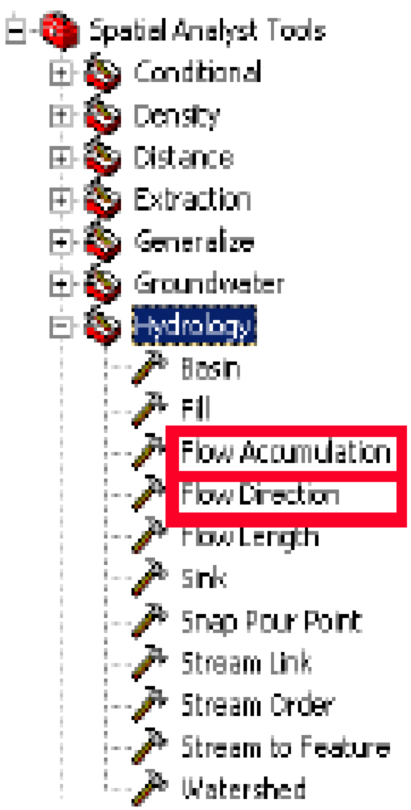


Roadmap: Surface hydrology

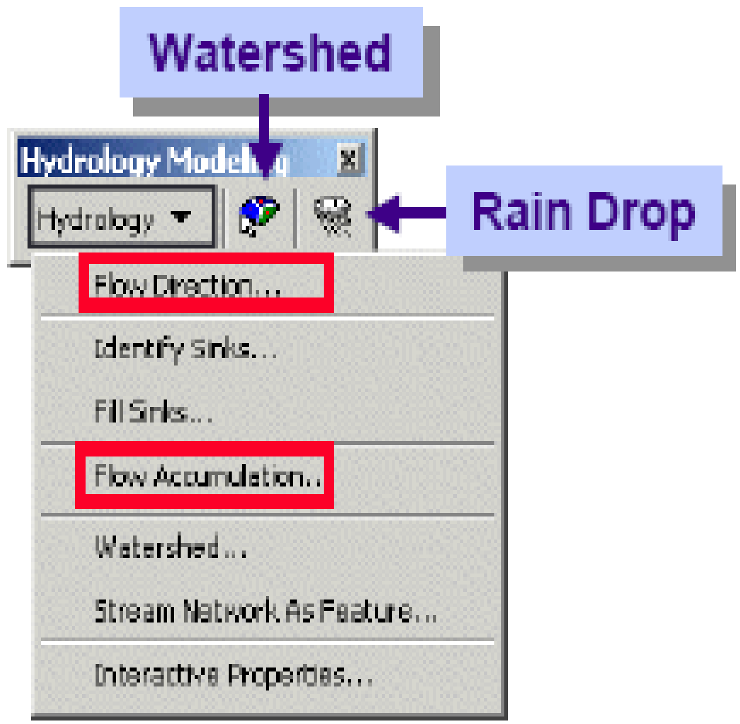


Surface hydrology tools

ArcToolbox Hydrology tools



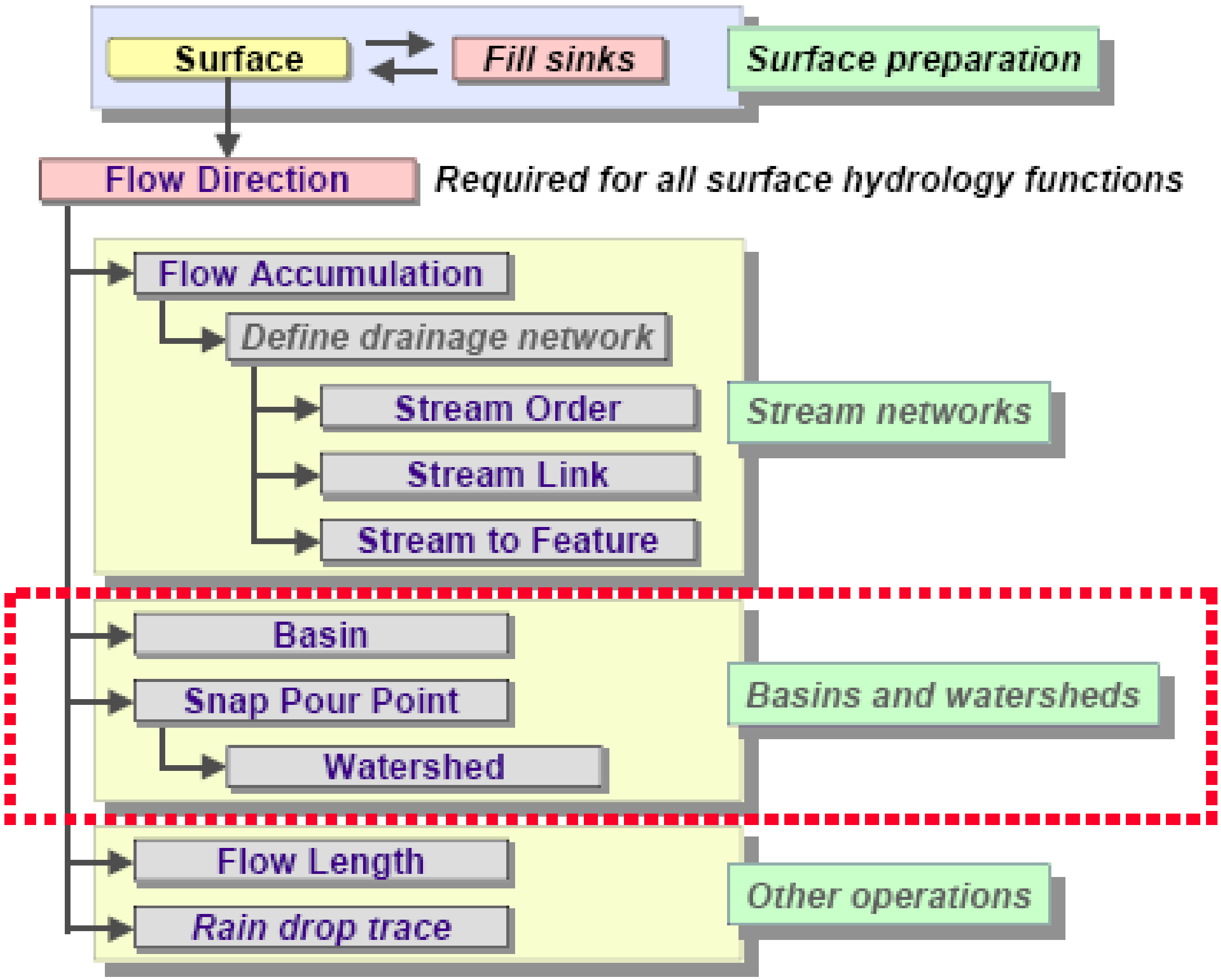
Hydrology Modeling toolbar in ArcMap



Map Algebra functions

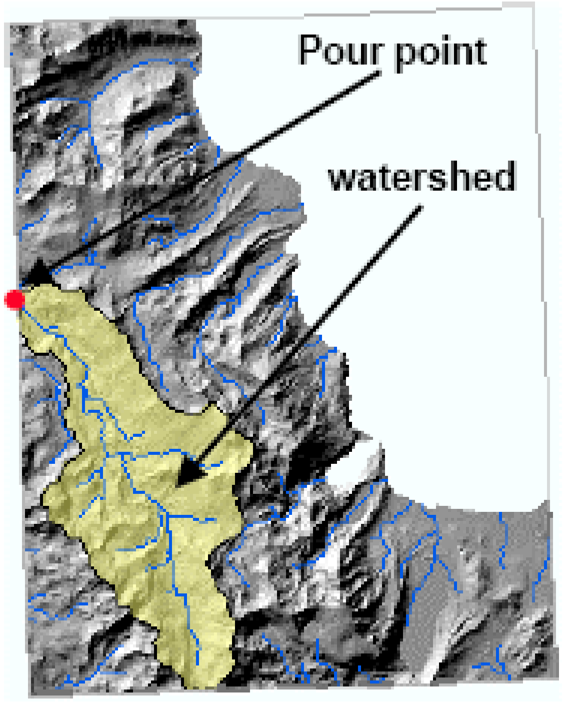
FlowDirection
FlowAccumulation
and so on

Roadmap: Basins and watersheds

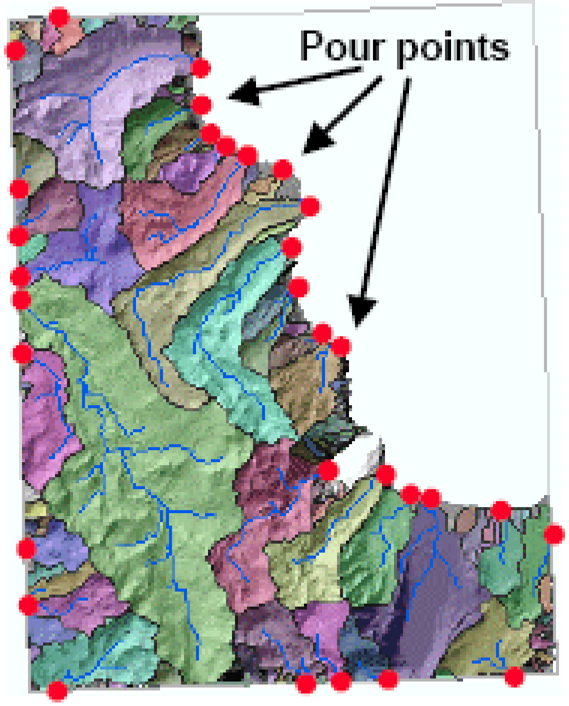


Defining basins and watersheds

- Find all up-slope cells from a pour point
  - WATERSHED: You provide pour points in a source layer
  - BASIN: Finds its own pour points



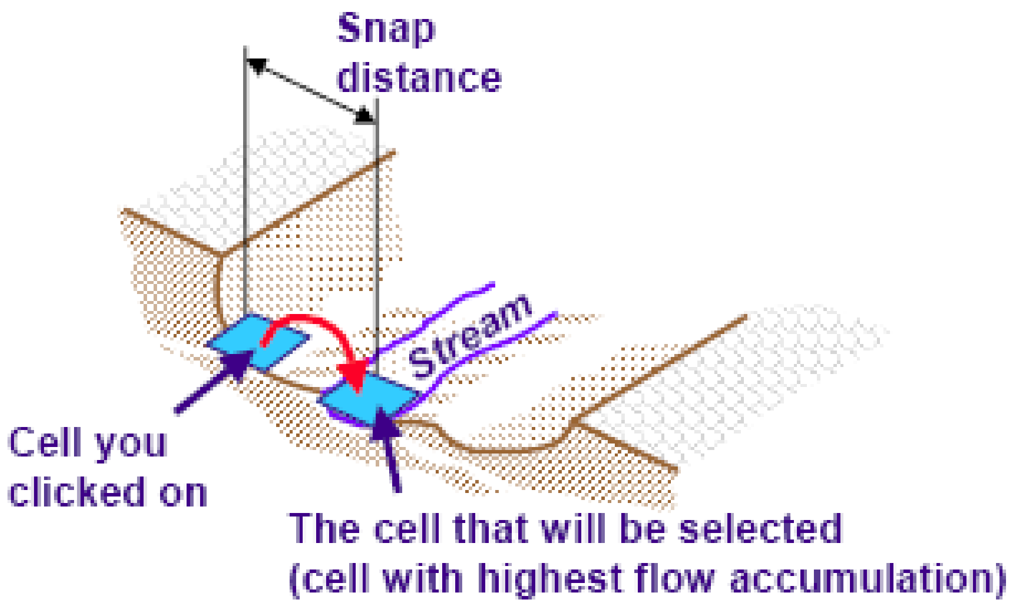
WATERSHED



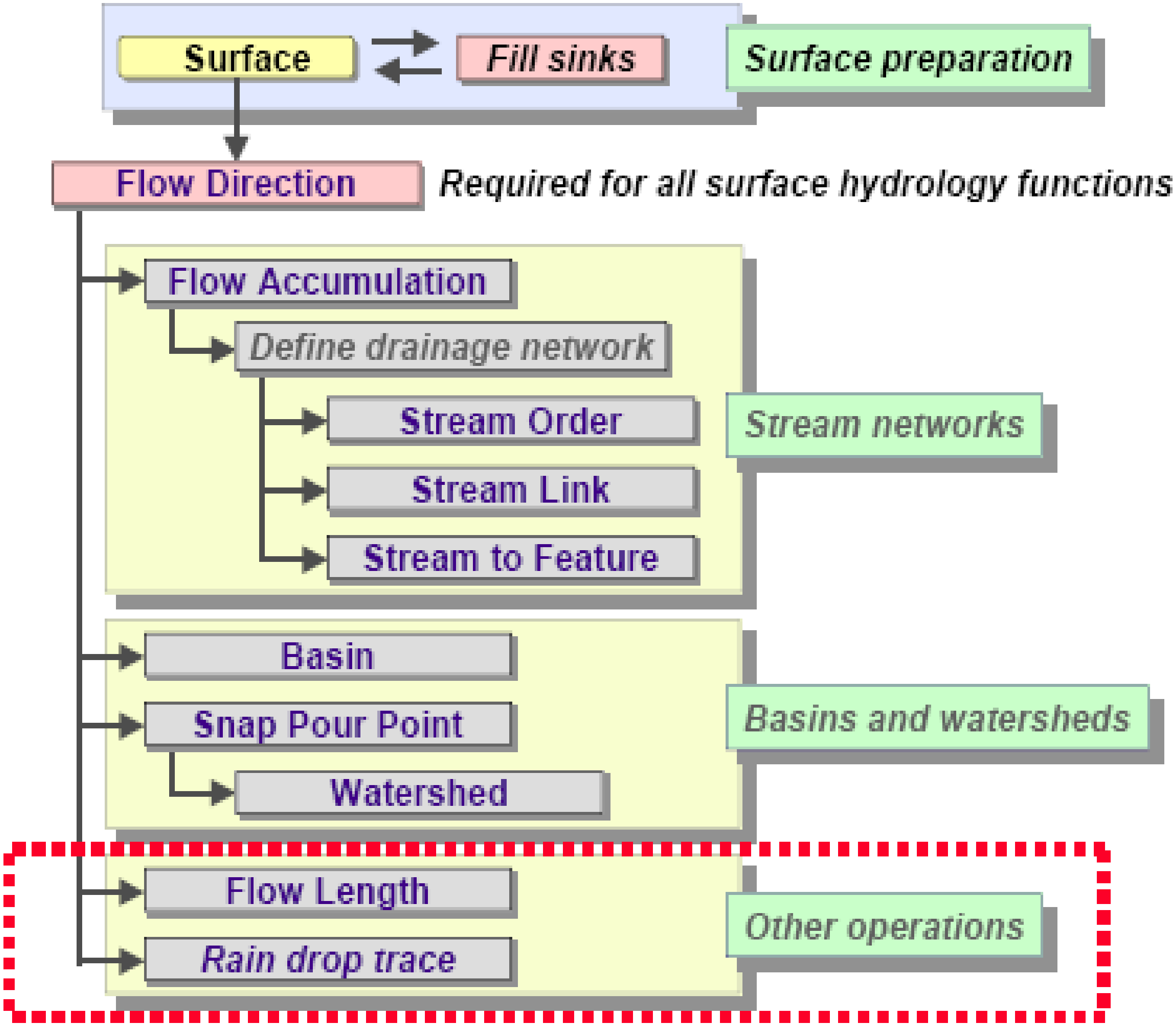
BASIN

Snap Pour Point tool

- Snaps pour point cell to high accumulation neighbor
  - Like the nearest stream cell
- Use before running the Watershed tool

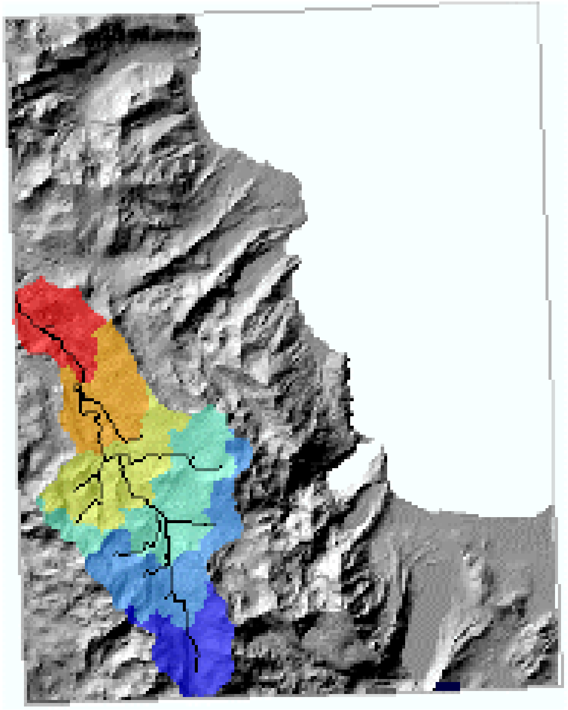


Roadmap: Other operations



Flow length

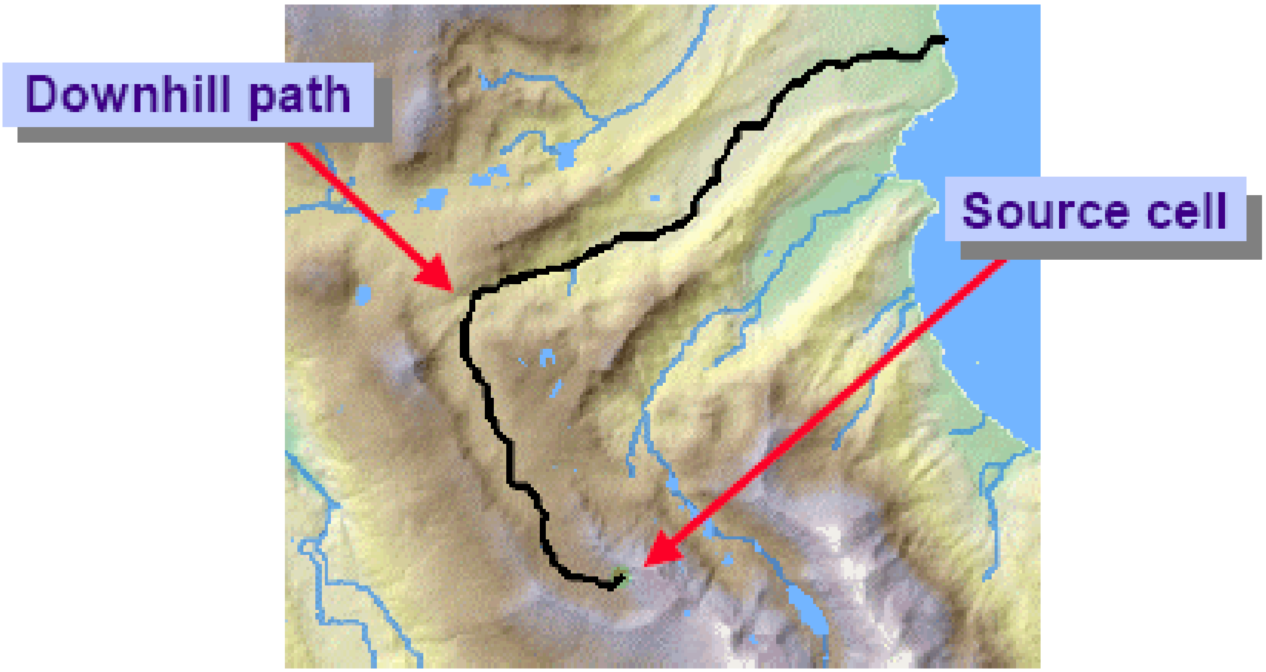
- Calculates flow length for each cell
  - Upstream: To top of basin
  - Downstream: To pour point
- May have a weight layer
- Use to compute a basin's time of concentration



Downstream flow length

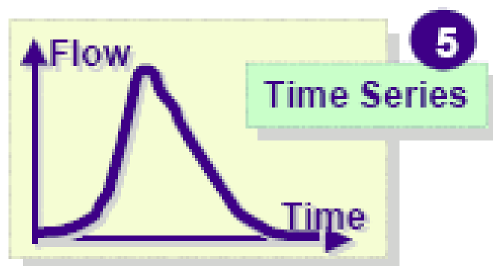
Tracing raindrops

- Trace the downhill path of a raindrop
  - *COSTPATH:* Returns a grid
  - *Hydrology* toolbar: Returns a graphic
- Use: Find where a contaminate spill will go



ArcHydro Data Model

- A geospatial and temporal data model
- Supports hydrologic simulation models
  - It is not itself a simulation model
  - Can construct a simulation model attached to ArcHydro using DLL
  - Can customize the behavior of ArcHydro
- Hydrologic information systems
- Five key areas



Exercise 8: Hydrology Tools

- EXERCISE 8A: SURFACE HYDROLOGY TOOLS
- EXERCISE 8B: GROUNDWATER HYDROLOGY TOOLS

Next Topic:

Model Design and Evaluation