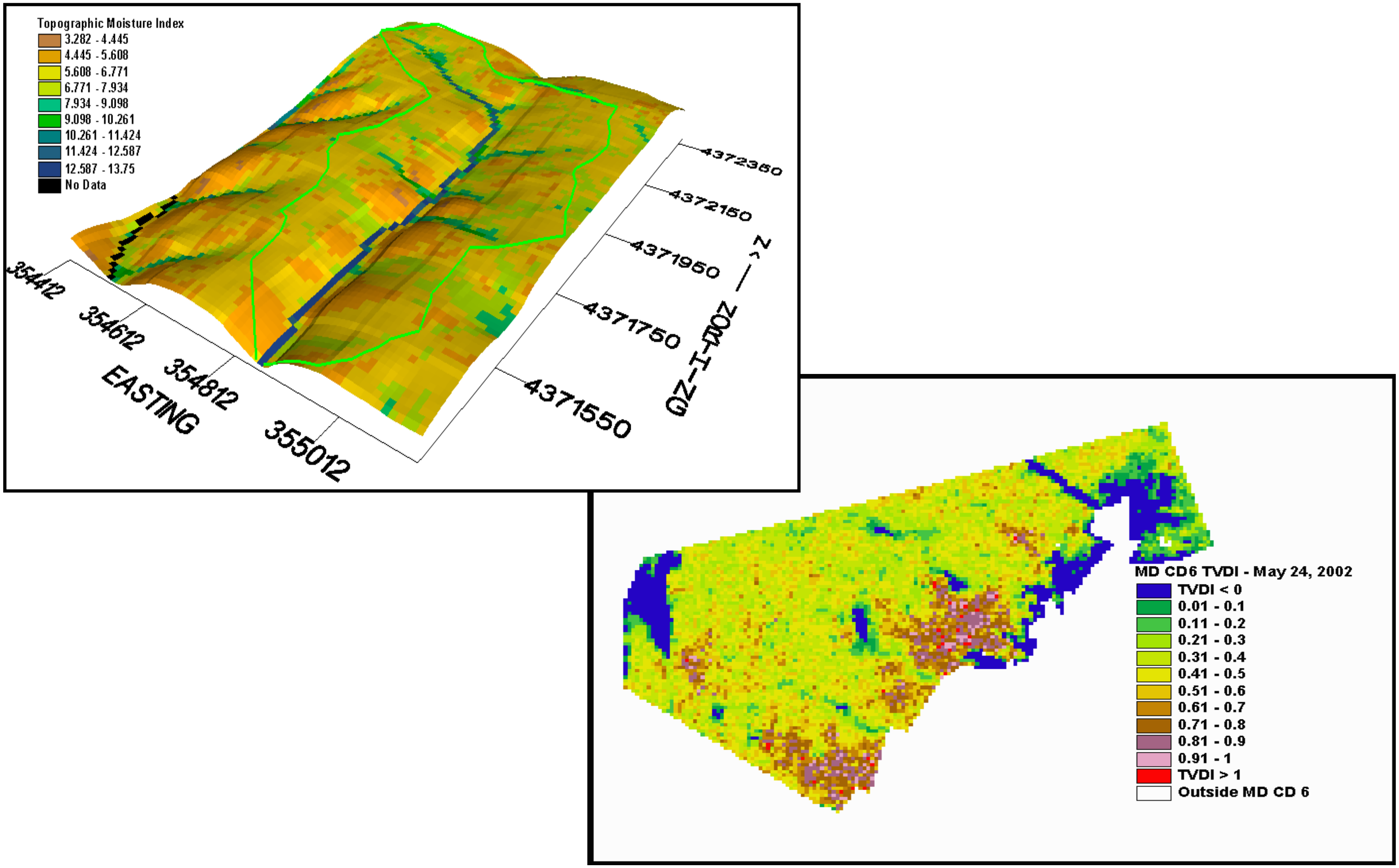
EEOS 383 – GIScience for Water Resources Research



EEOS 383 – GIScience for Water Resources Research

Course Description:

This course will provide students with an introduction to some **key GIScience approaches** for investigating the **movement of water** through terrestrial landscapes. This includes two major themes, that each compose about half of the course: The first focuses upon methods that have been developed to extract and represent the **structure of watersheds** from **digital terrain data**, and this will be followed by approaches that use **passive remote sensing** to describe **landscape hydrologic conditions**.

EEOS 383 – GIScience for Water Resources Research

- We will **work with**:
  - **Digital terrain models** (both isoline and gridded)
  - Datasets produced through the **analysis** of digital terrain models
  - **Remotely sensed data** that:
    - is collected by **passive satellite sensors**
    - includes **shortwave reflectance** data
    - and **longwave emittance** data

EEOS 383 – GIScience for Water Resources Research

- Students will be provided with **hands-on experience**, working with the ArcGIS desktop geographic information system (GIS) and Microsoft Excel. The **goals** are to help students:

1. Understand the **key concepts** that make the methods useful for investigating water resources
2. Become skilled at using the tools to **apply the techniques**
3. Build capability and understanding in the **application of GIScience** techniques to water resources research

EEOS 383 – GIScience for Water Resources Research

- Students will be expected to **read reference materials** and be **prepared to discuss them** each week
  - Class meetings will begin by focusing on the key concepts from the reference materials (both from the literature and from online materials such as slideshows and web pages)
- This will be followed by **weekly lab exercises** designed to augment the readings and to **sequentially build capability and understanding** in the application of GIScience techniques to water resources research
  - These exercises will use a combination of ArcGIS and Microsoft Excel to provide an opportunity for hands-on learning

EEOS 383 – GIScience for Water Resources Research

- **Course Web Page:**
  - http://www.faculty.umb.edu/david.tenenbaum/eeos383

- **Lab Exercises:**
  - Instructions are online, linked from the course web page
  - Data will sit in a course directory on the s:\ drive
  - There be will an exercise each week, and we will begin working on them during the class meetings
  - Due 1 week later at the beginning of the class meeting

- **Lateness Penalty:** **-10%** of total mark per weekday
  - Approach your instructor/TA for extenuating circumstances
  - Exercises more than a week overdue will not be accepted

EEOS 383 – GIScience for Water Resources Research

Course Evaluation:

Lab assignments	**60%**
Mid-term (Mar. 23)	**20%**
Final exam (TBA)	**20%**

EEOS 383 – GIScience for Water Resources Research

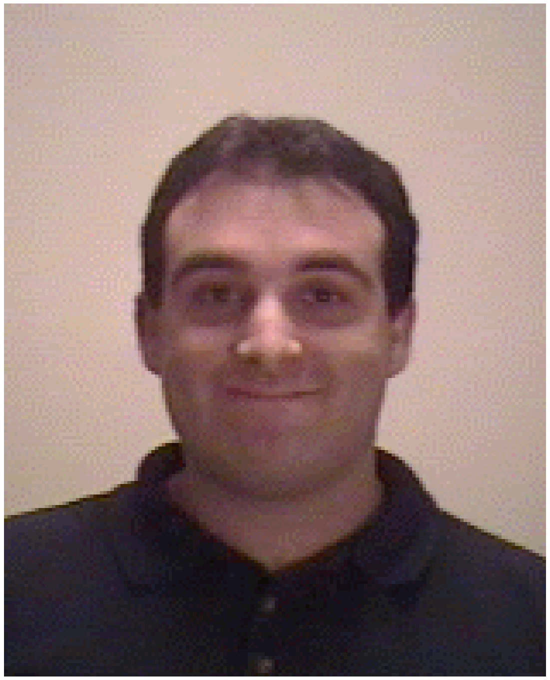
- **Class Meetings**:
  - McCormack M02-0415, moving to S-3-34
  - Tuesdays from 5:30 - 8:00 PM
- **Computer Labs**:
  - S-3-20 when open and not used by a class
  - S-3-34 anytime 24/7 when not used by a class
- **Instructor Office Hours**:
  - S-1-060
  - Mondays and Wednesdays from 1:00 - 2:30 PM

David Tenenbaum

- Hon. B.Sc. at the University of Toronto
  - **Majors**: Physical and Environmental Geography & Environment in Society
- M.Sc. at the University of Toronto
  - **Thesis**: RHESSys-ArcView Integrated Modelling Environment
- Ph.D. at the University of North Carolina at Chapel Hill
  - **Dissertation**: Surface Moisture Patterns in Urbanizing Landscapes
- Canadian Government Lab Visiting Fellow at the Water & Climate Impacts Research Centre
  - **Research**: NAESI - In-Stream Flow Needs

How to Reach Me

David E. Tenenbaum, Ph.D.
Assistant Professor
EEOS, UMass Boston

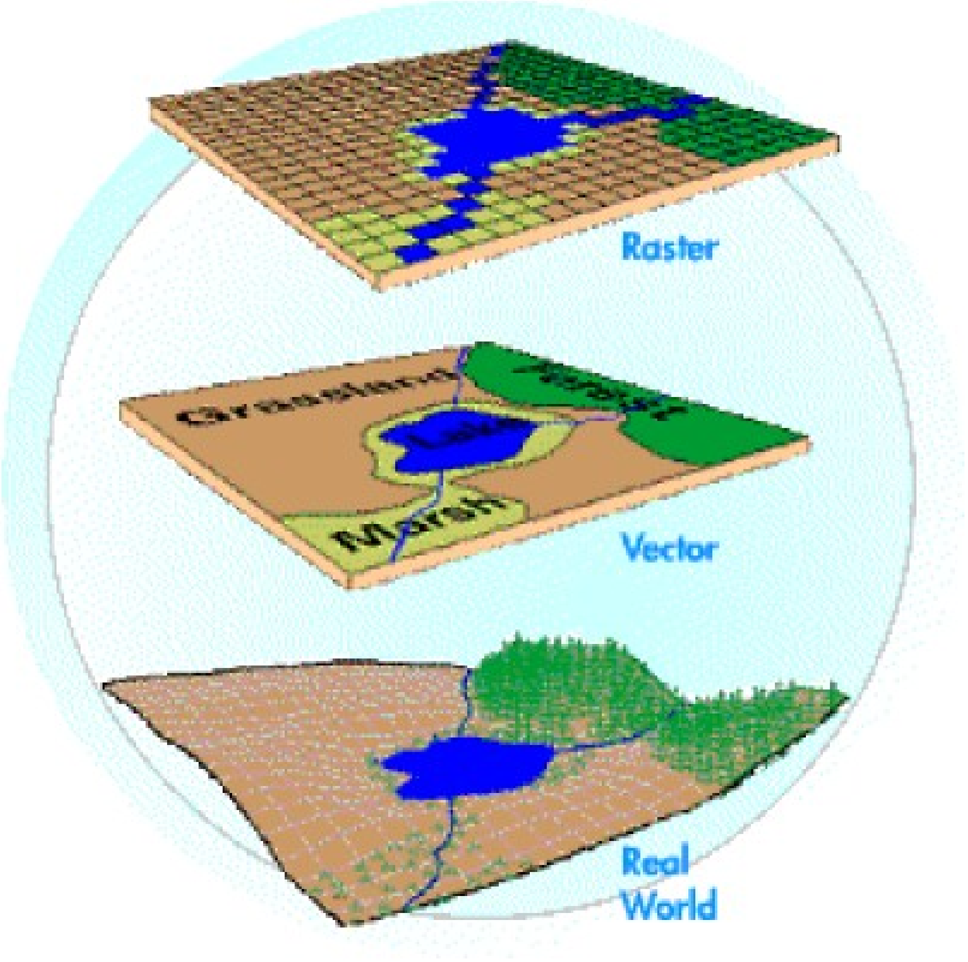


Office: Science Building S-1-60

Email: david.tenenbaum@umb.edu
Tel: 617-472-7396

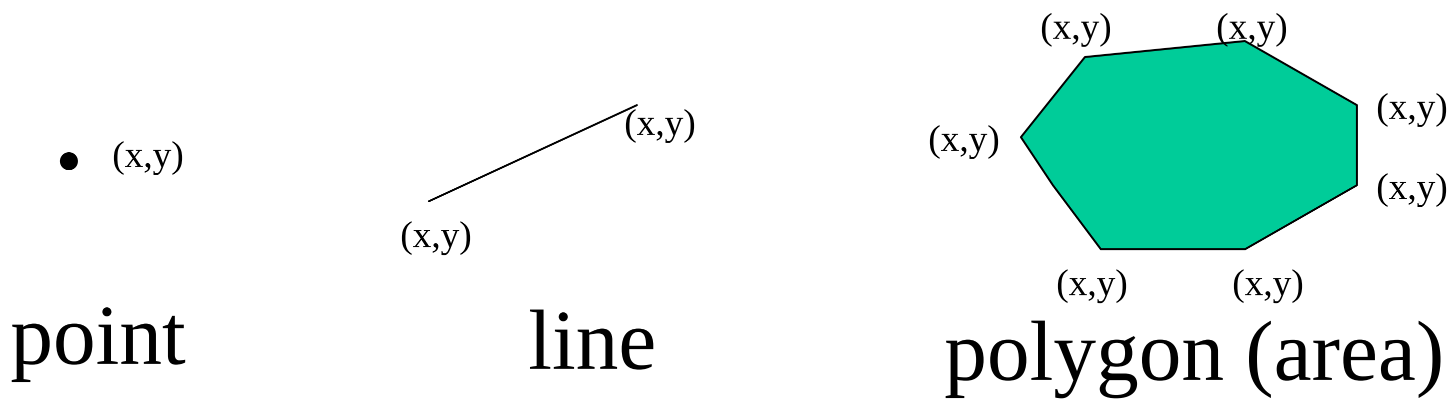
Spatial Data Models

- **Raster**
  uses individual cells in a matrix, or grid, format to represent real world entities



- **Vector**
  uses coordinates to store the shape of spatial data objects

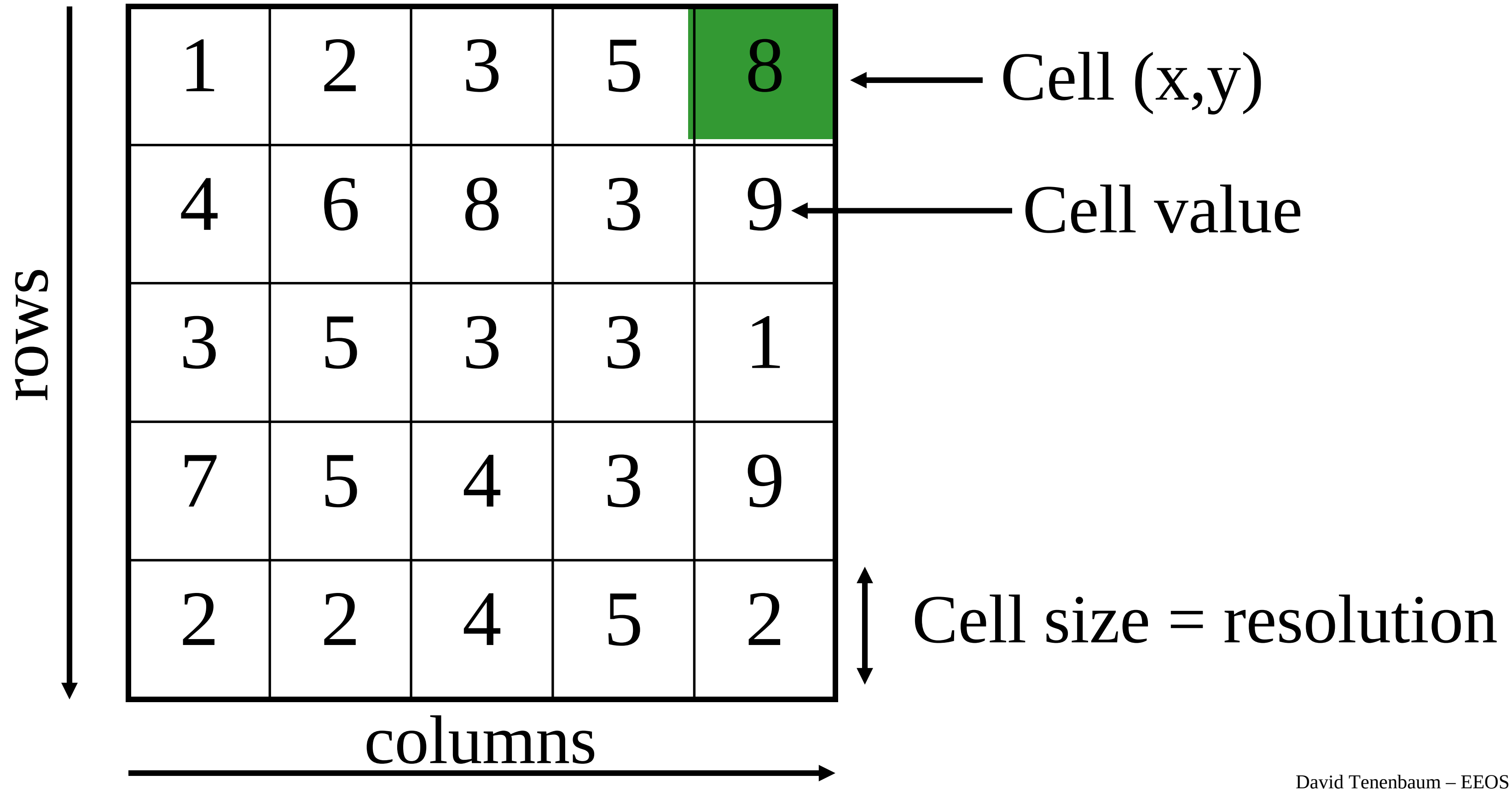
Geographic Features (Vector Model)



- **A point:** specified by a pair of (x,y) coordinates, representing a feature that is too small to have length and area.
- **A line:** formed by joining two points, representing features too narrow to have areas
- **A polygon (area):** formed by a joining multiple points that enclose an area

Raster Data Model

- The raster data model represents the Earth’s surface as an **array** of two-dimensional grid cells, with each cell having an associated value:



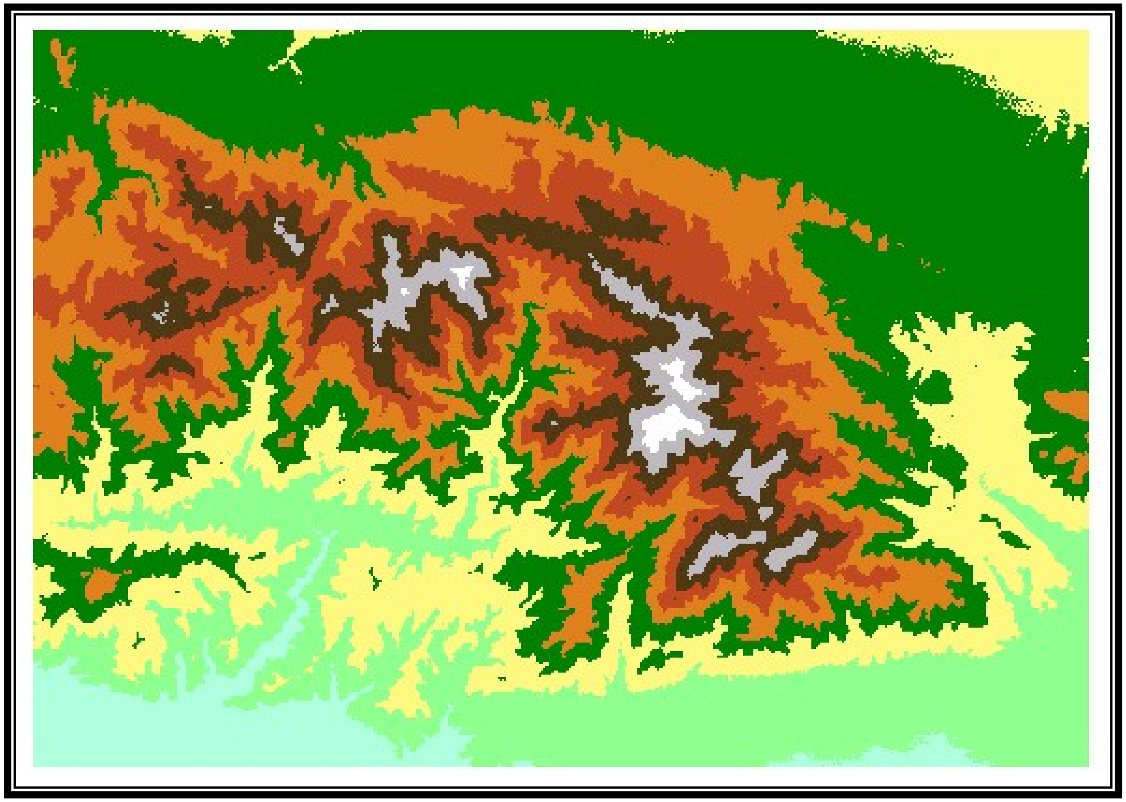
1	2	3	5	8
4	6	8	3	9
3	5	3	3	1
7	5	4	3	9
2	2	4	5	2

Raster Data Model

- Each grid cell in a raster data layer is one **unit** (the minimum amount of information in the raster data model)
- Every cell has a **value**, even if it is a special value to indicate that there is “no data” or that data is “missing” at that location
- The values are numbers, either:
  - **absolute values** **OR**
  - **codes** representing an **attribute**

Cells - Absolute Values

- In this instance, the **value** of the cell is actually the value of the phenomenon of interest, e.g. elevation data (whether floating point or integer):



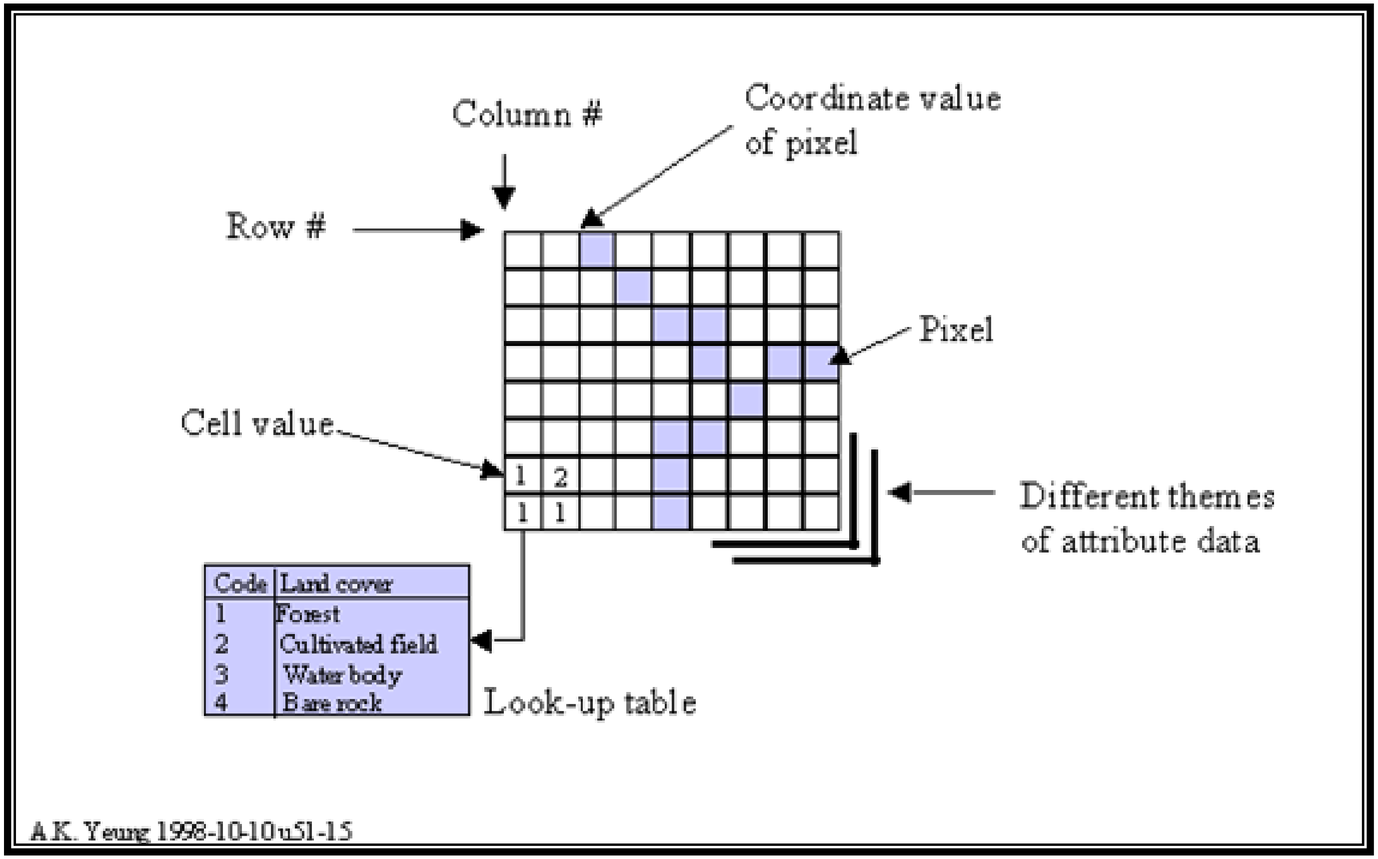
Cells - Coded Values

- Here, the values stored in each cell are used as **substitutes** for some **nominal** or **categorical** data, e.g. land cover classes:

ID	Land Cover Type	Ownership
1	Grass	Smith
…		
8	Water	Smith
13	Sand	Smith

Cells – Coded Values

- The coded values can then **link** to one (or more) attribute tables that associate the cell values with various themes or attributes:

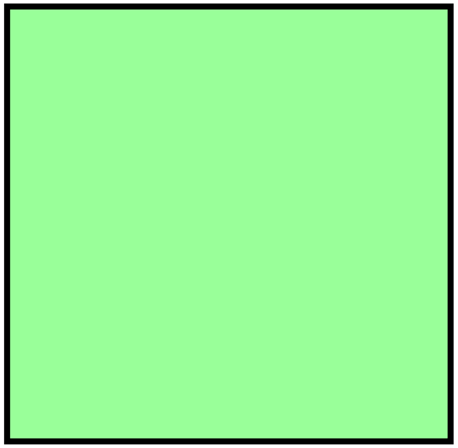


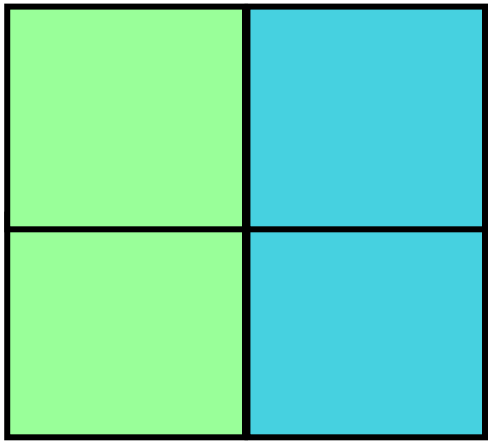
Code	Land cover
1	Forest
2	Cultivated field
3	Water body
4	Bare rock

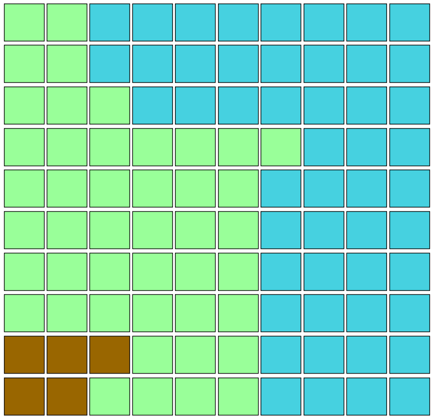
A.K. Yeung 1998-10-10 u51-15

Cell Size & Resolution

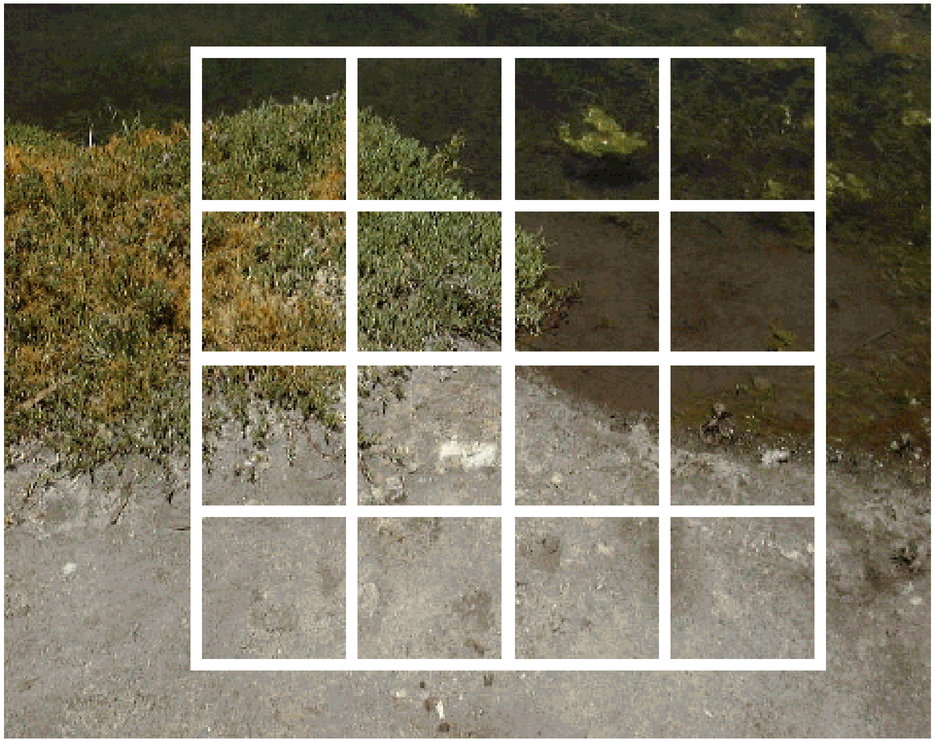
- The **size** of the **cells** in the raster data model determines the **resolution** at which features can be represented
- The selected **resolution** can have an **effect** on how features are represented:

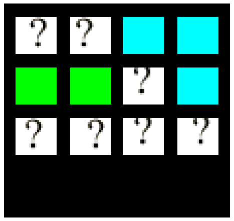

10 m Resolution


5 m Resolution

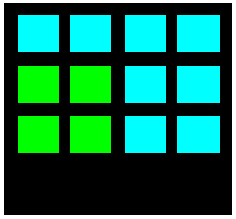

1 m Resolution

Rules for assigning values / The mixed pixel problem

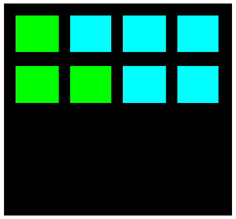




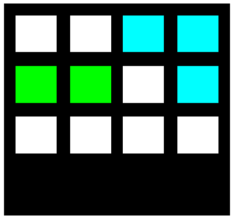
Water/Veg dominates



Winner takes all



Edges separate



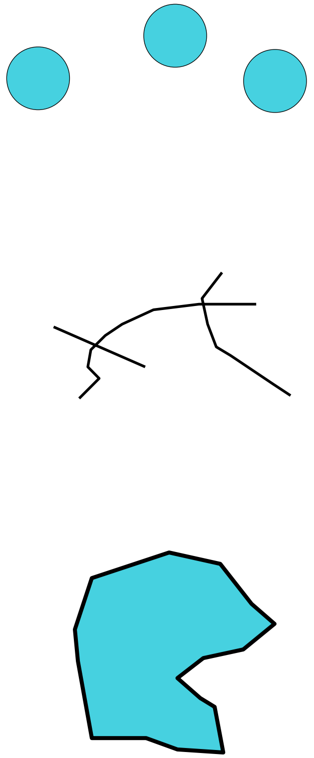
Raster and Vector Spatial Data Models

- **Vector** – implementation of discrete object conceptual model → **Features**
  - Point, line and polygon representations
  - Widely used in cartography, and network analysis
- **Raster** – implementation of field conceptual model → **Continuous**
  - Array of cells used to represent objects
  - Useful as background maps and for spatial analysis

Raster Data Model - Objects

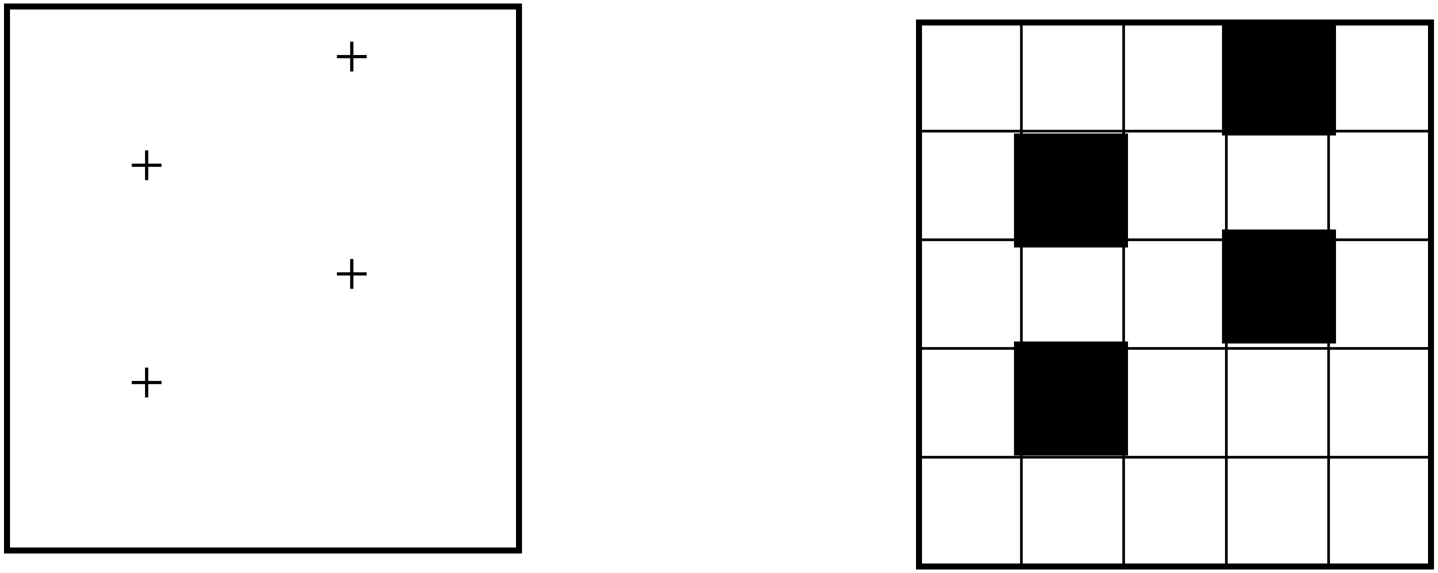
The raster data model still represents spatial objects, but does so differently from the vector model:

<u>Geographic Primitives</u>



- Points
  - –0 dimensional
- Lines
  - –1 dimensional
- Polygons
  - –2 dimensional

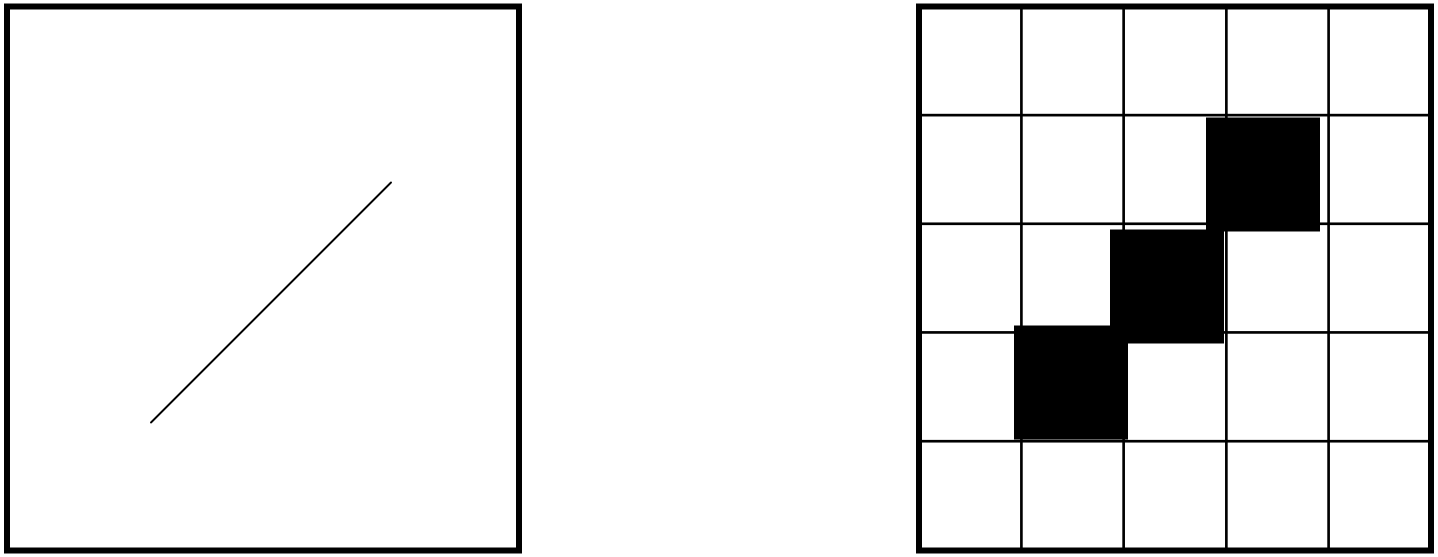
Raster Data Model - Points



1 point = 1 cell

What problem do we have here? How can we solve it?

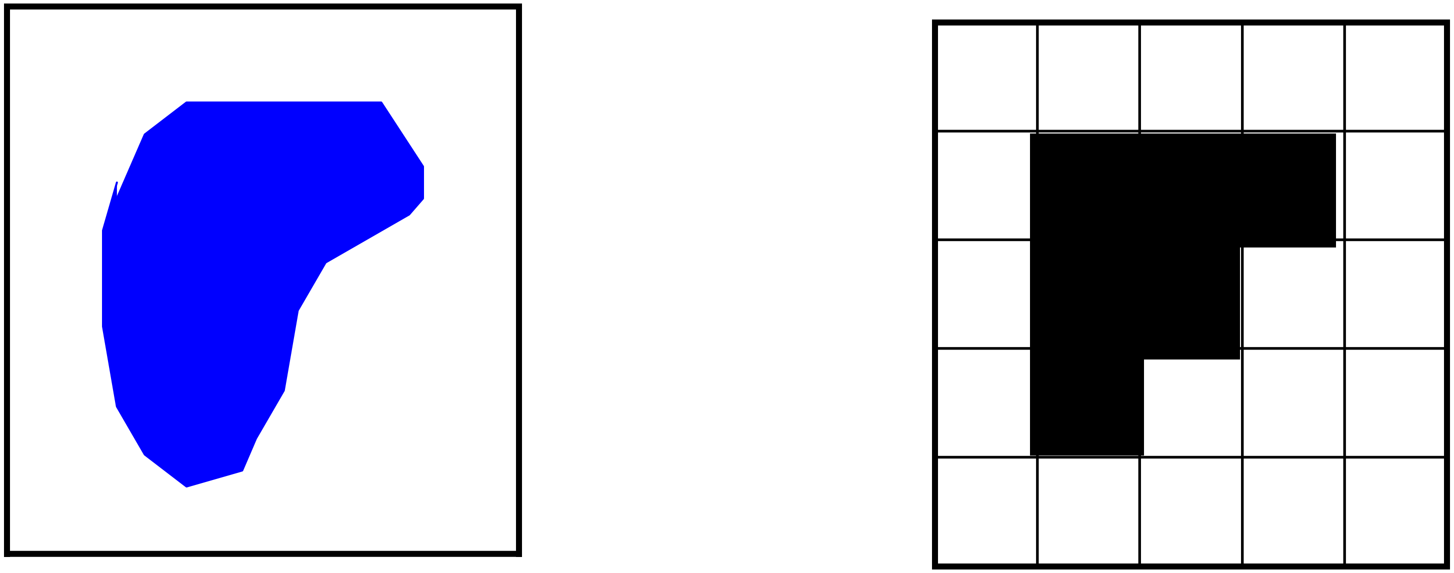
Raster Data Model - Lines



A line = a series of connected cells that portray length

Is there a problem with this representation?

Raster Data Model - Areas

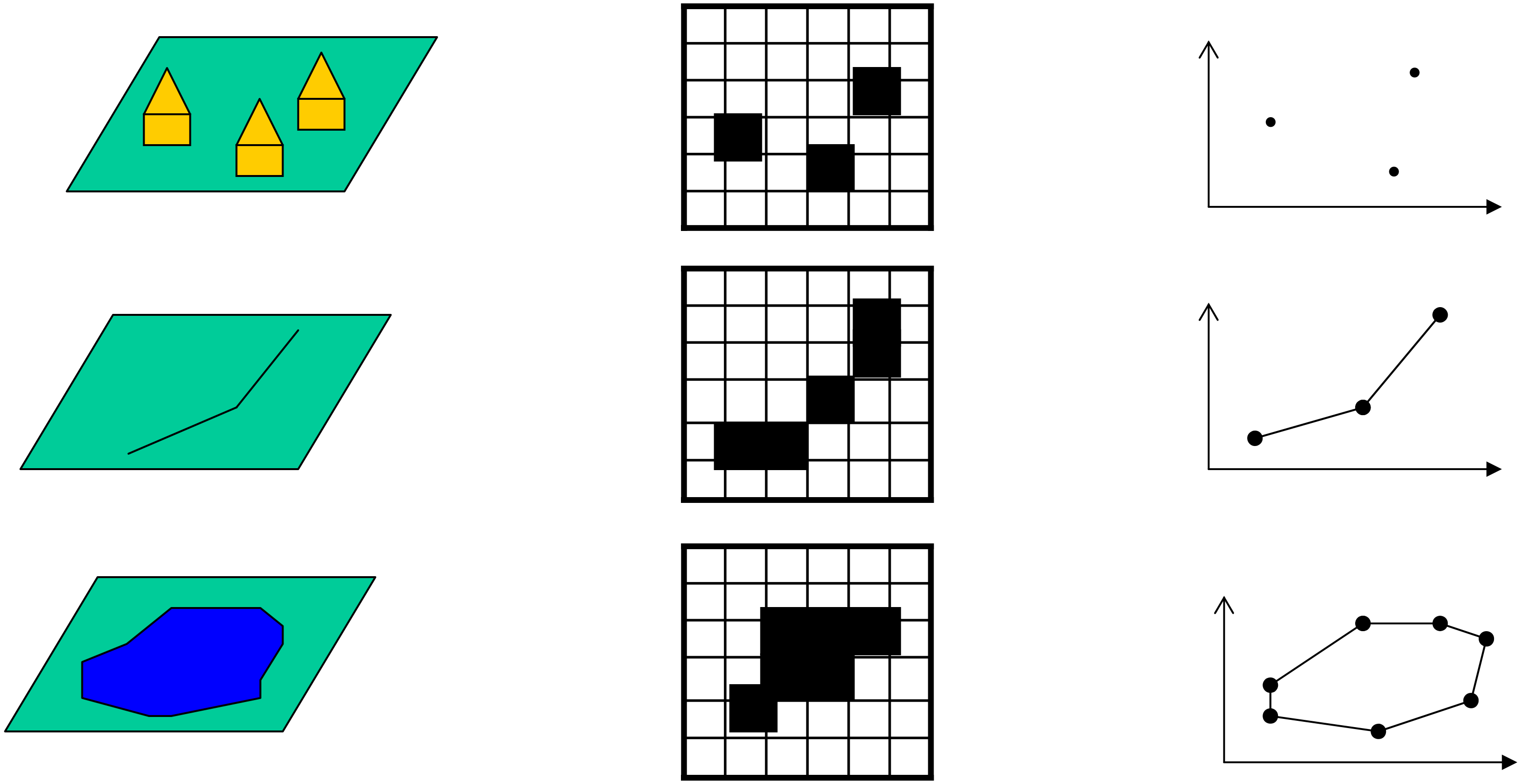


Area = a group of connected cells that portray a shape

What problems could we have with this representation?

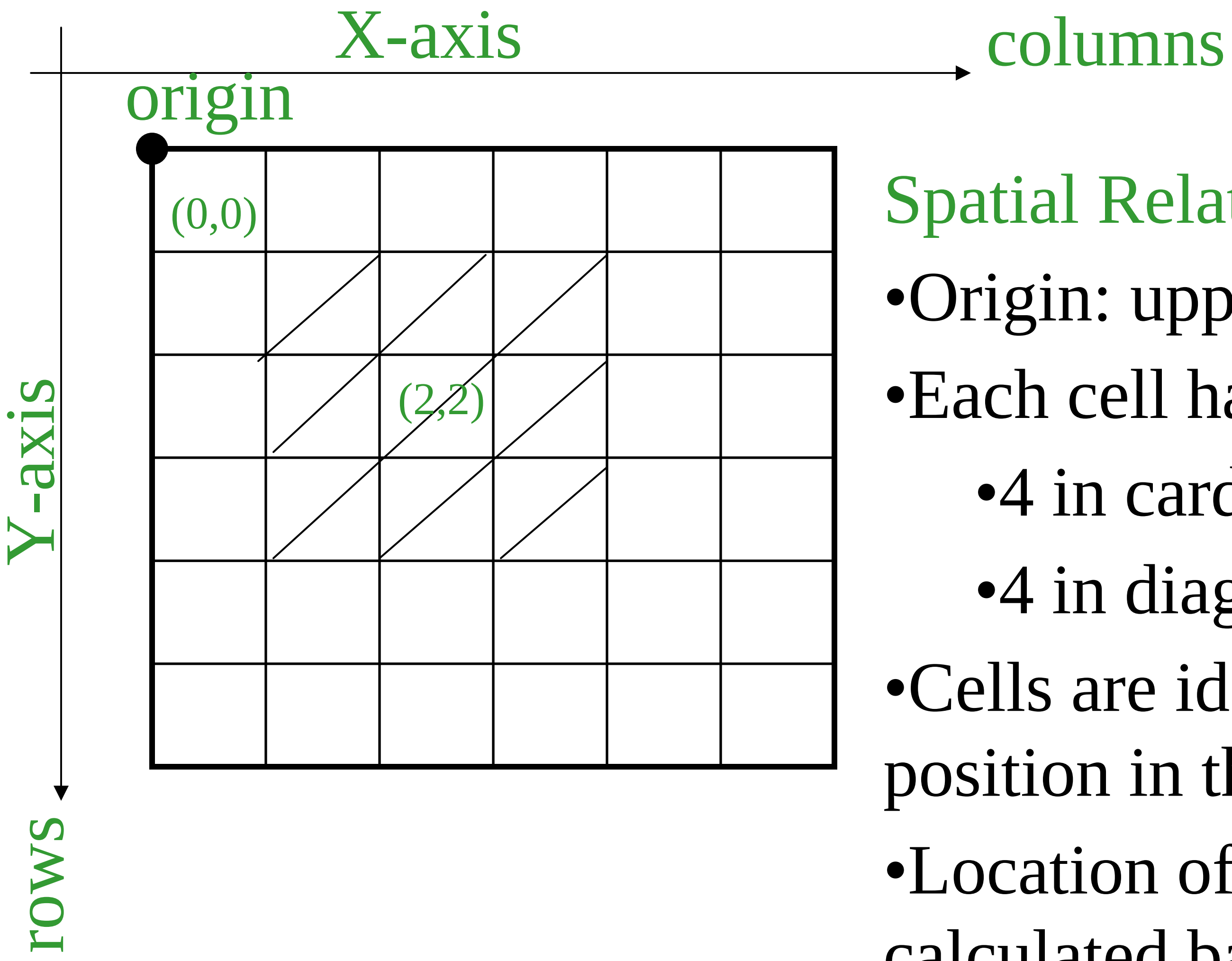
Raster and Vector Data Model Comparison

Real World Features **Raster** **Vector**



"A raster model tells what occurs everywhere, while a vector model tells where every thing occurs"

Raster Data Model - Topology

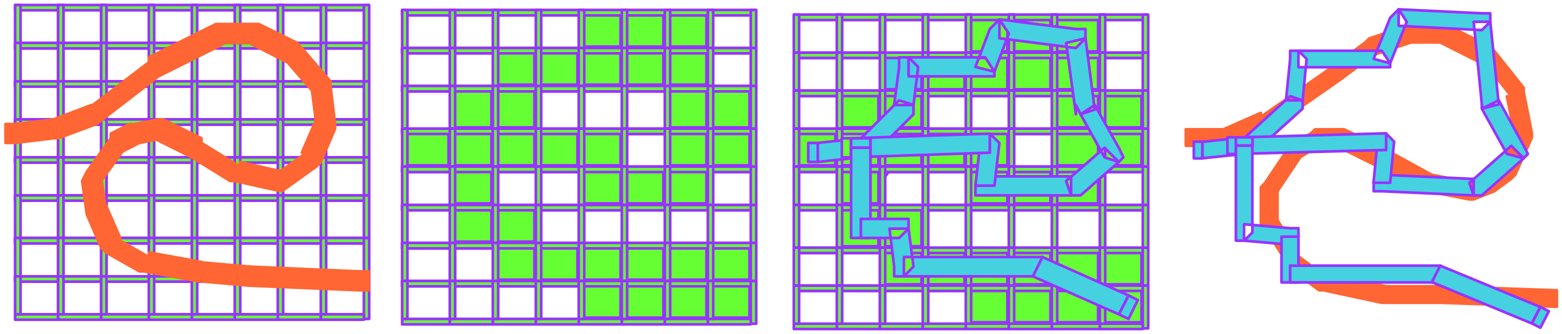


Spatial Relations are implicit

- Origin: upper left corner
- Each cell has 8 neighbors
  - 4 in cardinal directions
  - 4 in diagonal directions
- Cells are identified by their position in the grid
- Location of a cell can be calculated based on its position and the cell size

Vector to Raster Transformations

- Quite often, data in the **vector and raster** models need to be **used together**, and data from one model is **transformed** to be represented in the other model
- Any such transformations can cause **distortions** – consider this line transformed from vector to raster, and then back to vector:

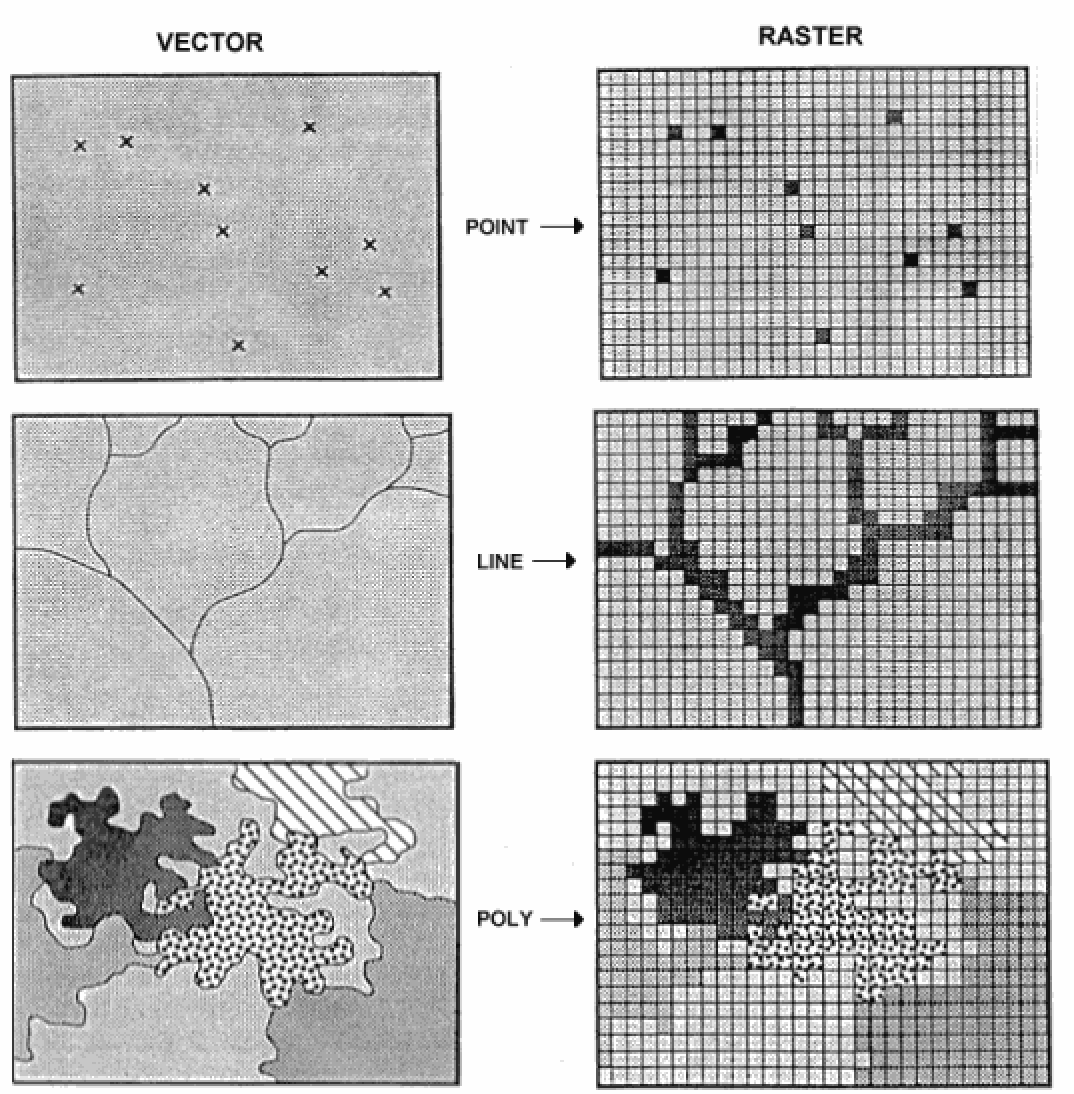


Rasterization –Why?

- There are instances where the raster data model is the **preferred model** and vector data sources are the data that need to be converted:
  - Projects that make extensive use of **satellite remote sensing data** often use the raster model
  - If cell-based **simulation modeling** or analysis is to be used, the raster model is preferred
  - If phenomena that are best represented using **continuous data** are the subject of inquiry, the raster model is ideal

Rasterization - Issues

- The **key issues** to are:
- **cell size**, and its effect on the spatial representation of entities **AND**
- how we choose **cell values** to represent attributes
- This is **simpler** compared to vectorization.



Rasterization - Issues

Selecting a Cell size:

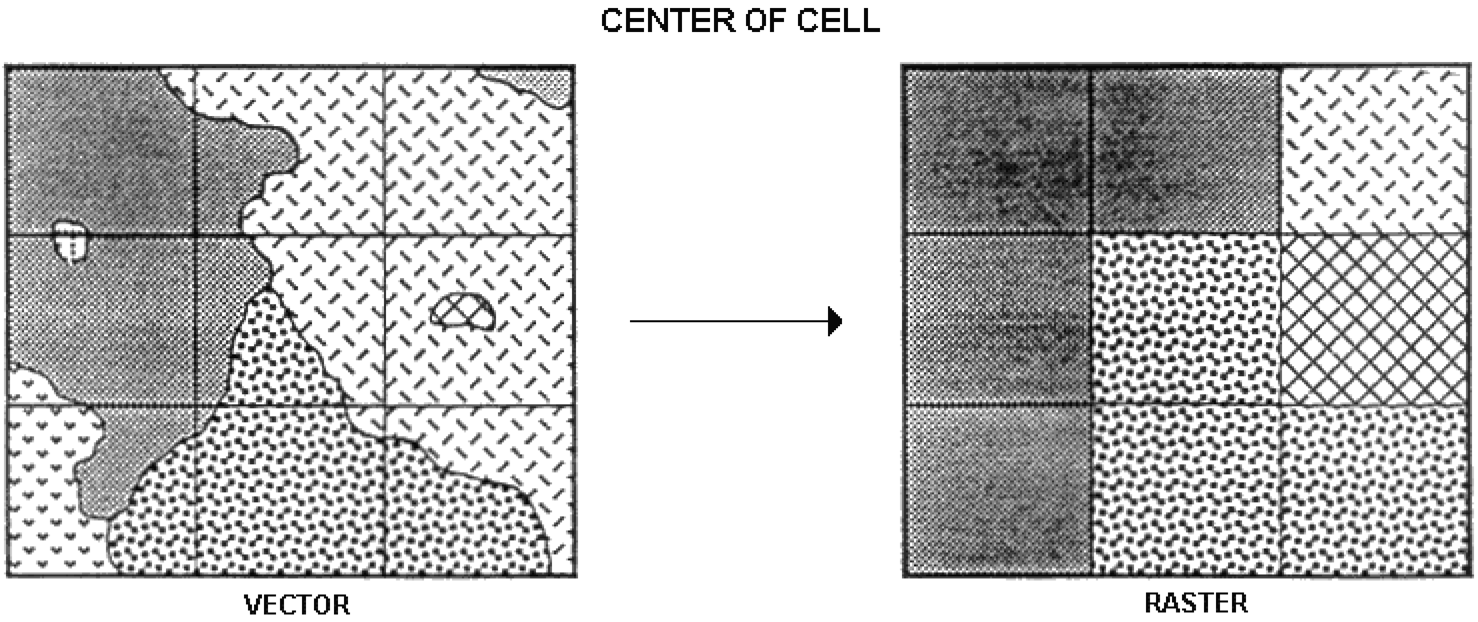
- By selecting a smaller cell size, we get a better representation of the vector data, more cells at a fine spatial resolution
- Using a larger cell size results in less cells, but more information loss at a coarse spatial resolution

Rules for determining cell values:

1. Use the value from the cell center (centroid)
2. Use majority weighting within each cell
3. Calculate weighted values

Rasterization – Cell Values

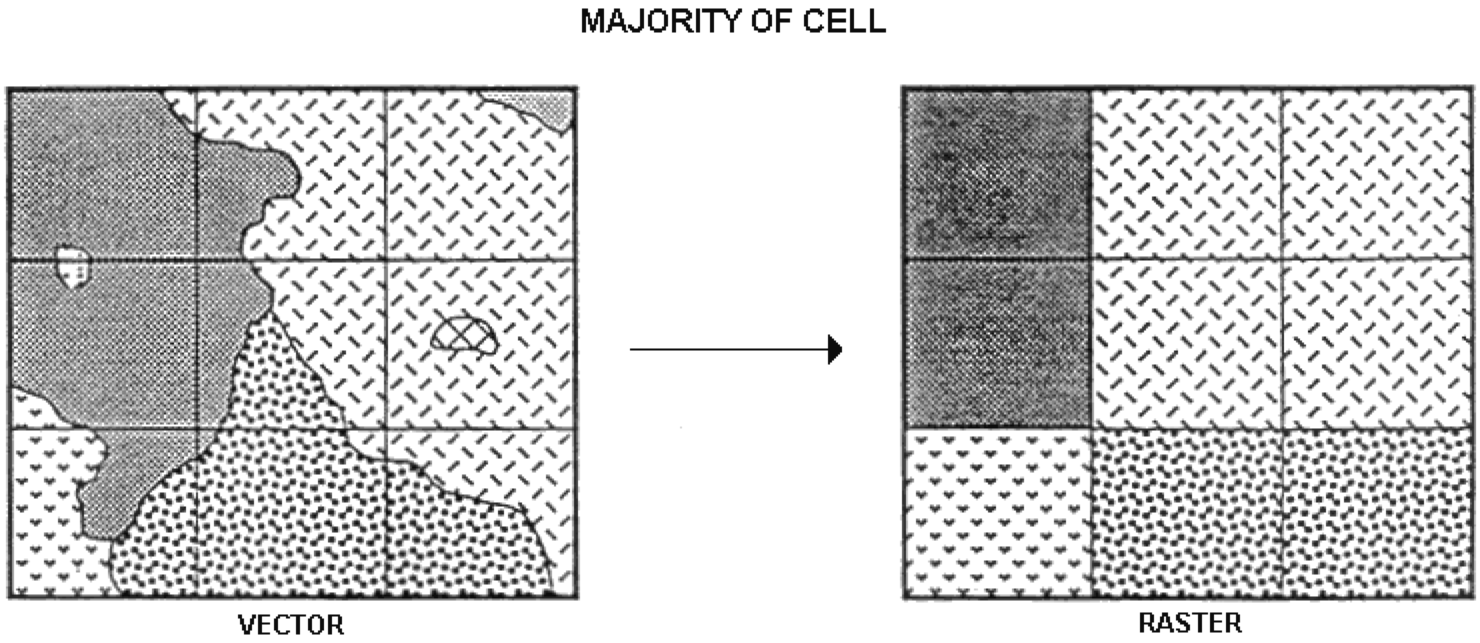
1. Use the value from the cell center (centroid)



- The value at the **centroid** is assigned to the cell
- This is a simple approach, but it can **over-represent** the values from small areas

Rasterization – Cell Values

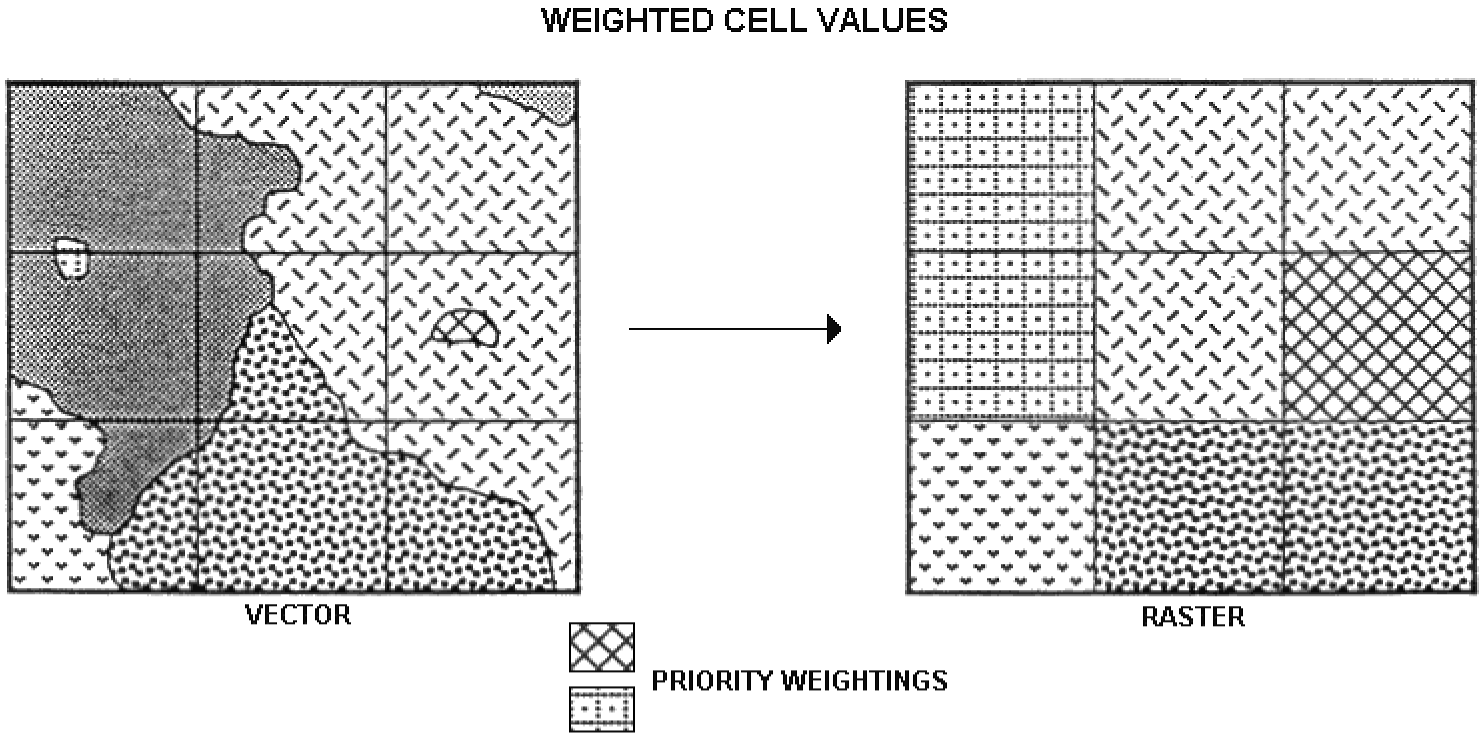
2. Use majority weighting within each cell



- The value covering the **majority** of the area is assigned to a cell
- This is a **“fairer”** representation than cell centers

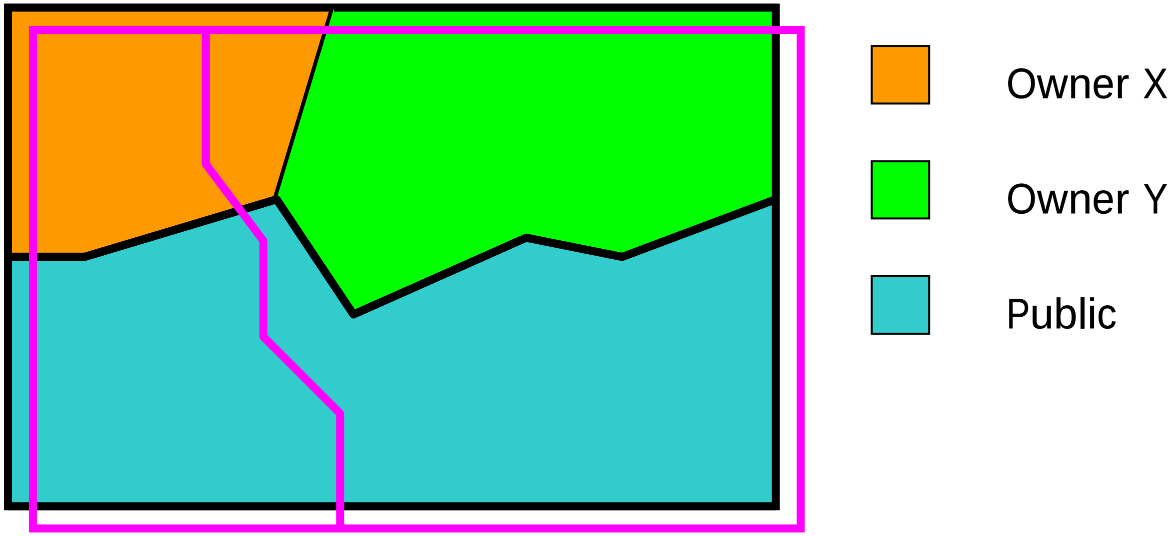
Rasterization – Cell Values

3. Calculate weighted values



- Priority weights are based upon the **importance** of different values
  - The **“most important”** value present is assigned to a cell
- This ensures the representation of **crucial** geographic phenomena

Polygon Overlay, Field Case



- A layer representing a field of **land ownership** (symbolized using colors) is overlaid on a **layer of soil type** (layers offset for emphasis). The result after overlay will be a single layer with **5 polygons**, each with land ownership information and a soil type

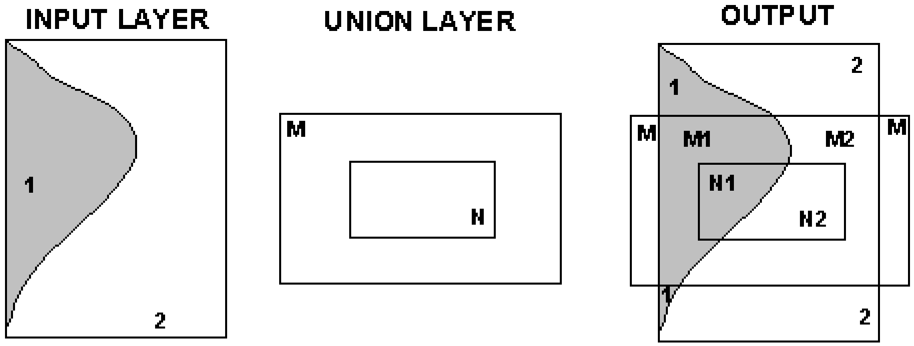
Polygon Overlay Analysis

- Overlay polygon layer (A) with polygon layer (B)
  - **What** are the **spatial polygon combinations** of A and B?
    - » Generate a **new data layer** with **combined polygons**
      - **attributes from both** layers are included in output
- **How** are polygons combined (i.e. what geometric rules are used for combination)?
  - UNION (Boolean OR)
  - INTERSECTION (Boolean AND)
  - IDENTITY
- Polygon overlay will generally result in a **significant increase** in the **number of spatial entities** in the output
  - can result in output that is **too complex** too interpret

Polygon Overlay Analysis

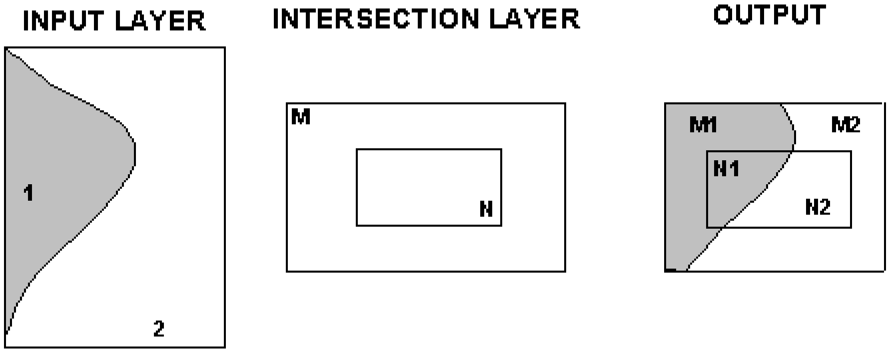
UNION

- overlay polygons and **keep areas from both layers**



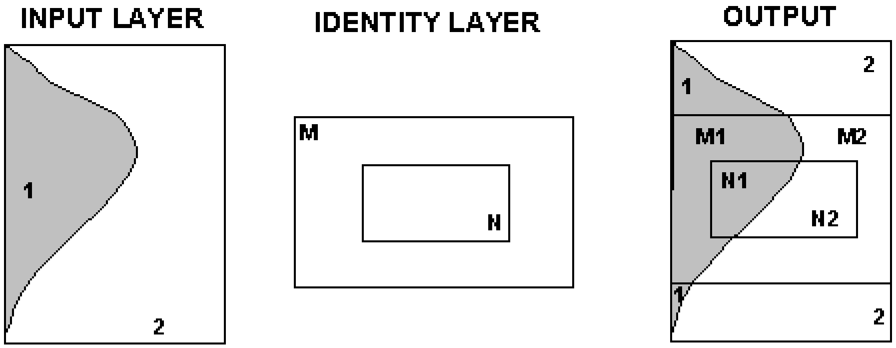
INTERSECTION

- overlay polygons and **keep only areas** in the input layer that fall **within the intersection** layer



IDENTITY

- overlay polygons and **keep areas from input layer**

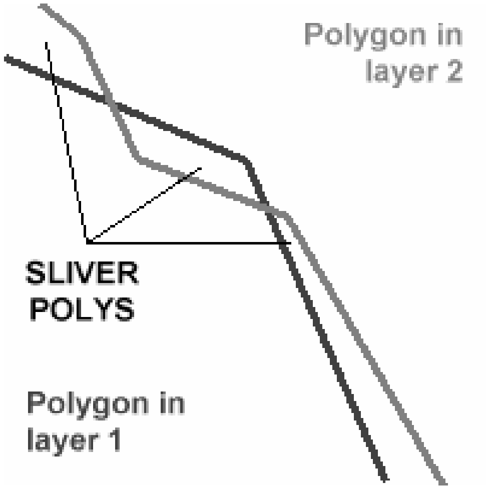


Problems with Vector Overlay Analysis (esp. Polygon)

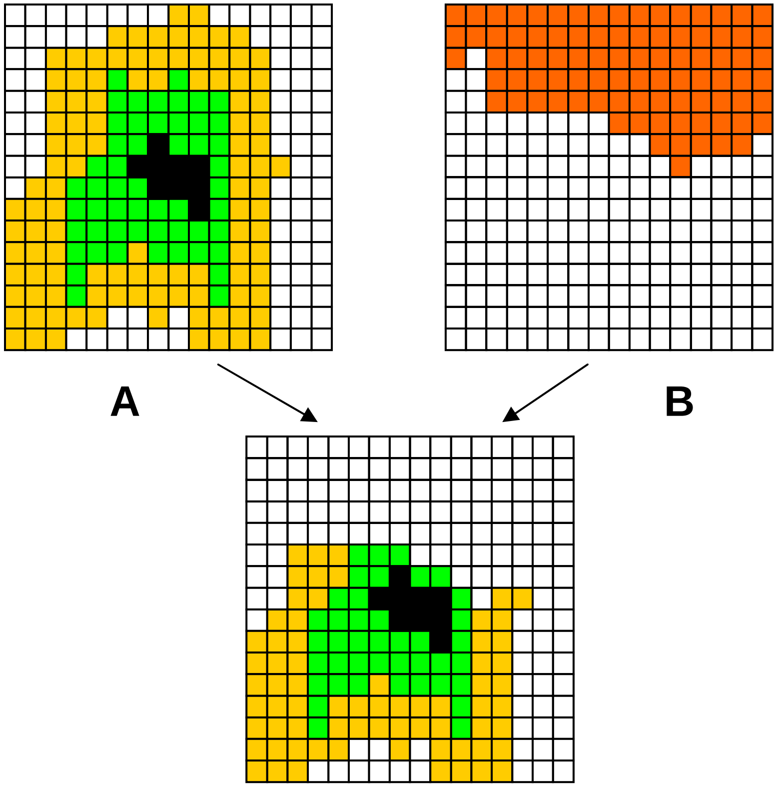
- Overlay analysis using the vector spatial data model is **highly computationally intensive**:
  - Complicated input layers can tax even current processors
- There is a tradeoff between the **complexity** vs. the **interpretability** of results
  - Complex input layers with many polygons can result in 100s or 1000s of resulting polygon combinations … can we make sense of all those combinations?

Problems with Vector Overlay Analysis (esp. Polygon)

- There are often **spatial mismatches** between input layers
  - This is a **common problem** of vector geospatial data sets
    - » Overlay results in **spurious sliver polygons**
  - We can "filter" out spurious slivers by querying to select all polygons with AREA less than some **minimum threshold**
    - It is **difficult to choose** a threshold to avoid deleting 'real' polygons



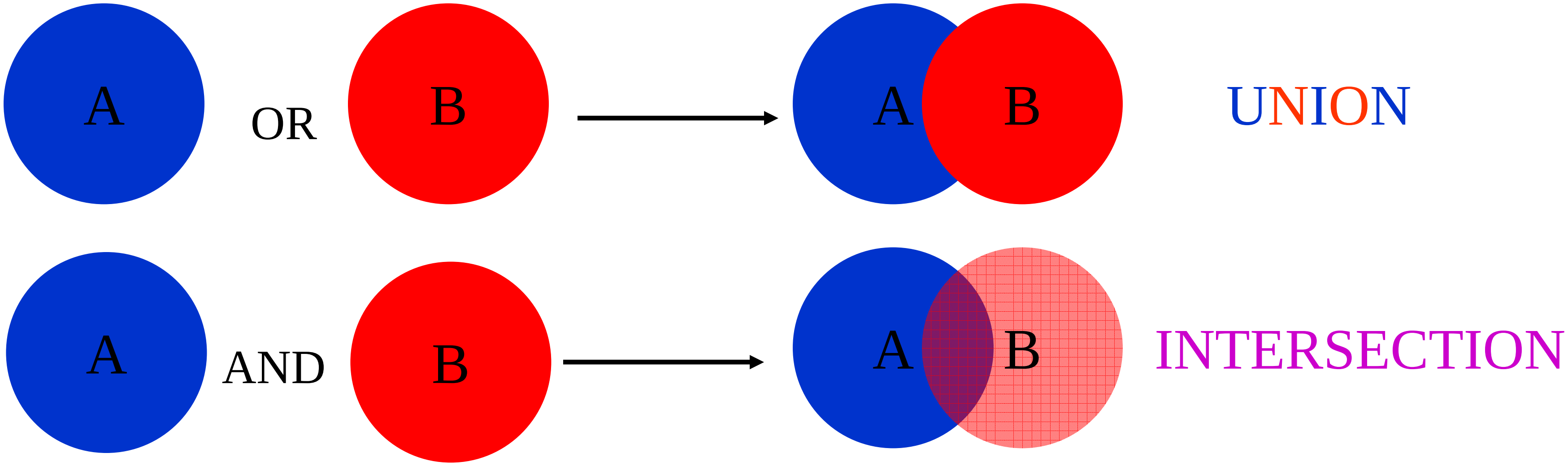
Overlay of Fields Represented as Rasters



The two input data sets are maps of (A) travel time from the urban area shown in black, and (B) county (red indicates County X, white indicates County Y). The output map identifies **travel time to areas in County Y only**, and might be used to compute average travel time to points in that county in a subsequent step

Boolean Operations

- In both **Venn probability diagrams** and **vector overlay analysis**, we used **UNION & INTERSECTION** operations, corresponding to Boolean operations of **OR & AND**



- We can apply these concepts in the **raster** spatial data model as well, but on a **per cell basis** with two input layers that contain true/false or 1/0 data:

Boolean Operations with Raster Layers

- The AND operation requires that the value of cells in **both** input layers be **equal to 1** for the output to have a value of 1:

0	1	1
0	0	1
1	0	1

AND

0	0	0
1	1	1
0	0	1

=

- The OR operation requires that the value of a cells in **either** input layer be **equal to 1** for the output to have a value of 1:

0	1	1
0	0	1
1	0	1

OR

0	0	0
1	1	1
0	0	1

=

Algebraic Operations w/ Raster Layers

- We can **extend** this concept from Boolean logic to **algebra**
- Map algebra:
  - Treats input layers as **numeric inputs** to mathematical operations (each layer is a separate numeric input)
  - The result of the operation on the inputs is calculated on a **cell-by-cell basis**
- This allows for **complex overlay analyses** that can use as many input layers and operations as necessary
- A common application of this approach is **suitability analysis** where multiple input layers determine suitable sites for a desired purpose by **scoring cells** in the input layers according to their effect on suitability and combining them, often **weighting layers** based on their importance

Simple Arithmetic Operations

Summation

0	1	1
0	0	1
1	0	1

+

0	0	0
1	1	1
0	0	1

=

0	1	1
1	1	2
1	0	2

Multiplication

0	1	1
0	0	1
1	0	1

×

0	0	0
1	1	1
0	0	1

=

0	0	0
0	0	1
0	0	1

Summation of more than two layers

0	1	1
0	0	1
1	0	1

+

0	0	0
1	1	1
0	0	1

+

0	0	0
1	1	1
0	0	1

=

0	1	1
2	2	3
1	0	3

Raster (Image) Difference

The difference between two layers

5	4	3
6	5	6
7	1	5

-

3	5	6
1	4	5
3	2	7

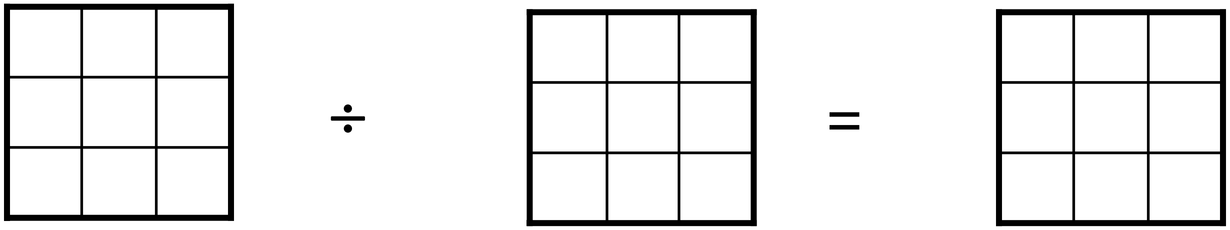
=

2	-1	-3
5	1	1
4	-1	-2

- An application of taking the differences between layers is **change detection**:
  - Suppose we have **two raster layers** that each show a map of the **same phenomenon** at a particular location, and each was generated at a **different point in time**
  - By taking the **difference** between the layers, we can **detect changes** in that phenomenon over that interval of time
- Question: **How** can the locations where changes have occurred be identified using the difference layer?

Raster (Image) Division

Question: **Can we** perform the following operation? Are there any **circumstances** where we **cannot** perform this operation? Why or why not?



More Complex Operations

Linear Transformation

$$\mathbf{a}\begin{bmatrix}1 & 2 & 4\\ 3 & 2 & 1\\ 5 & 3 & 2\end{bmatrix} + \mathbf{b}\begin{bmatrix}1 & 0 & 0\\ 5 & 1 & 1\\ 2 & 0 & 1\end{bmatrix} + \mathbf{c}\begin{bmatrix}0 & 0 & 0\\ 1 & 1 & 1\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

- We can multiply layers by **constants** (such as a, b, and c in the example above) before summation
- This could applied in the context of computing the results of a **regression model** (e.g. output y = a*x1 + b*x2 + c*x3) using raster layers
- Another application is **suitability analysis**, where individual **input layers** might be **various criteria**, and the **constants** a, b, and c determine the **weights** associated with those criteria