GE2215 Lecture 3
GIS Data Model II: Raster Data Model

Dr. Yan Yingwei

Department of Geography

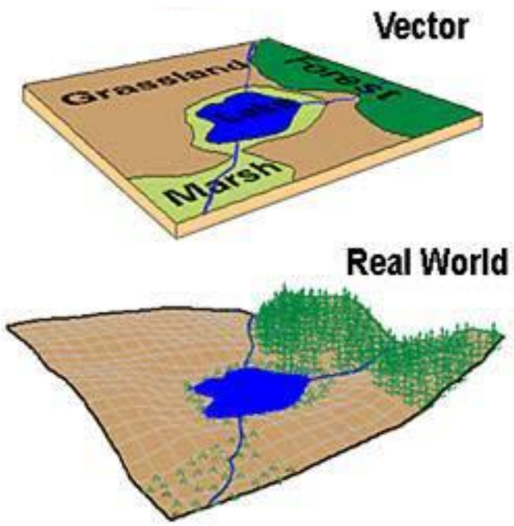
National University of Singapore

Recap: Basics of GIS Data Model

- Concept of GIS data model
- Characteristics of GIS data model
  - **Spatial (geometric and topological) characteristics**
  - **Attribute characteristics**
  - **Temporal characteristics**
- Types of GIS data
  - Spatial data: Graphic data and Image data
  - Non-spatial data: Attribute data and Statistical data

Recap: Vector Data Model

- Vector data model, also called the discrete object model, uses discrete objects to represent spatial features



Recap: Compression of vector data

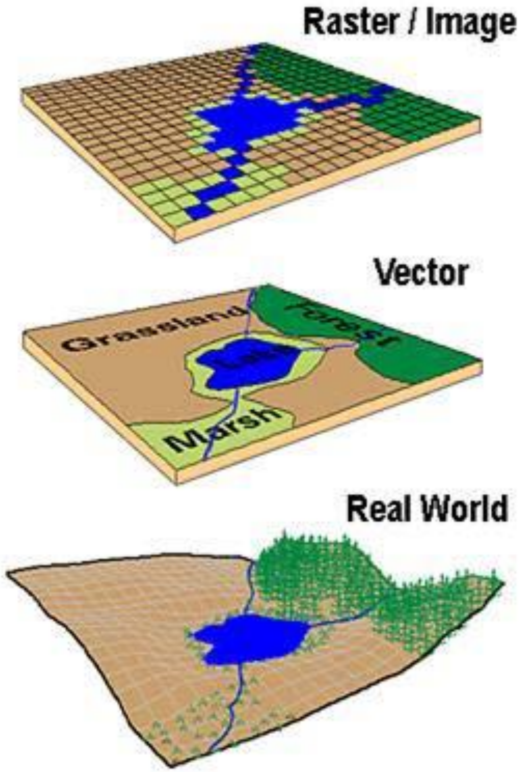
- In vector data model, data compression is often necessary to reduce **storage** space, improve **transfer speed** and **processing** efficiency
- Three commonly used algorithms for vector data compression
  - Vertical distance
  - Deflection angle
  - Douglas Peucker

Recap: Topological Relations

- Topology
- Topological relations
  - Topological adjacency – same category
  - Topological correlation – different category
  - Topological containment – same category but different level
- Topological and Non-topological (Spaghetti) data model

Recap: Spatial data model

- Vector data model
  - — Represented by point, line and polygon, with their coordinates (x, y) being recorded
  - — Object-based
  - — Ideal for discrete objects
- Raster data model
  - — To divide region into small regular cells, which are often called as pixels.
  - — Location-based. The location of a pixel can be identified by its row and column number
  - — Ideal for continuous objects

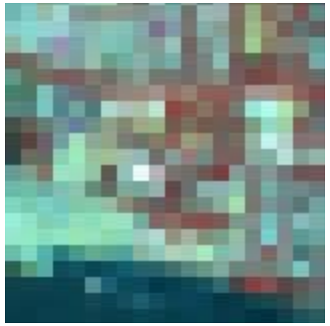


Outline of this lecture

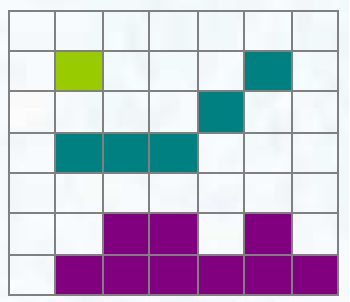
- Concept of the raster data model
- Elements of the raster data model
- Raster data structure
- Data conversion between raster and vector data

Concept of the raster data model

A raster



A raster image after zooming in

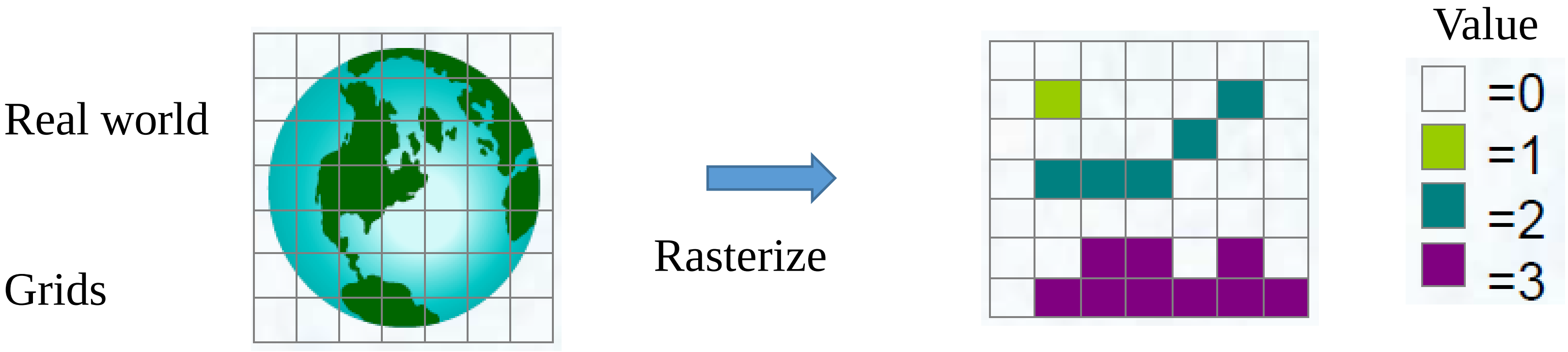


Digital representation of a raster image

- A raster is also called a grid or an image
- The basic unit in raster data is pixel or cell, and its numerical value can be used to describe the existing phenomenon in the real world
- The changes in the cell value reflect the spatial variation of the phenomenon

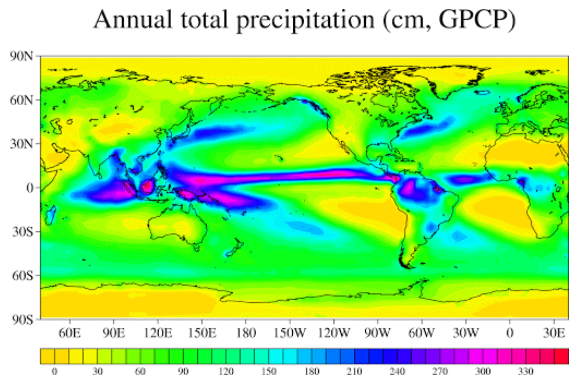
Concept of the raster data model

The essence of raster image data is a **pixel array**. The location of each pixel is identified by its **row** and **column** number, and has **intensity values** associated with it.

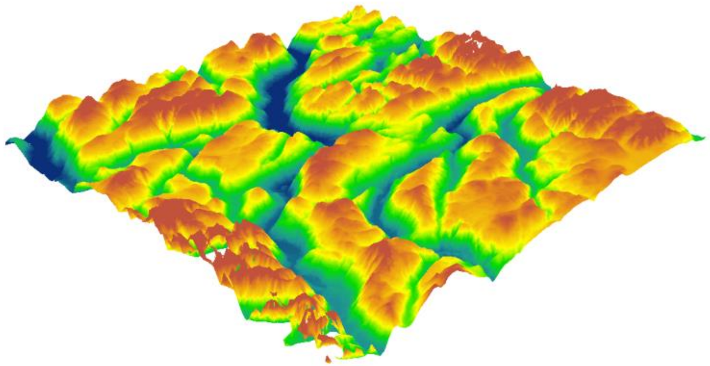


Raster data is an approximate representation of the real world.
The level of approximation is related to the cell size

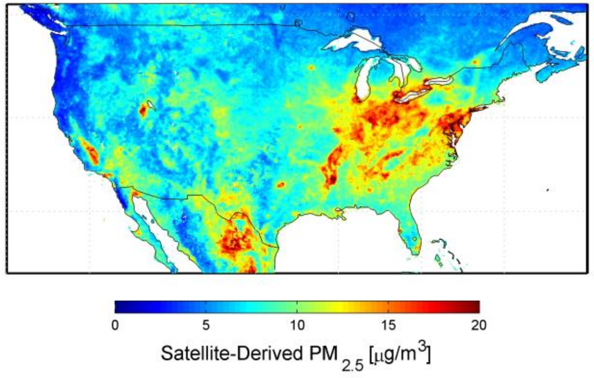
Why raster data model?



Precipitation distribution



Elevation model



PM 2.5 concentration

- The vector data model does not work well with spatial phenomenon that vary **continuously** over the space such as precipitation, elevation, and PM 2.5 concentration.
- A better option for representing continuous phenomenon is the raster data model

Can you give another example that a raster data model is a better option for representation?

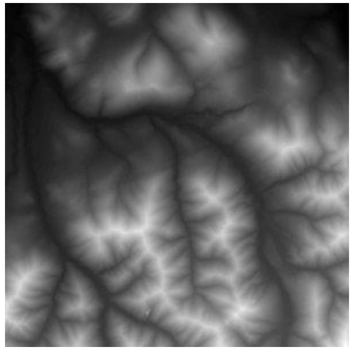
Types of raster data

- In practice, a wide variety of data used in GIS are encoded in raster format
  - Scanned paper maps
  - Maps captured on computer screen
  - Satellite images (one of the most important remote sensing data)
  - Digital elevation model (DEM)

Scanned paper map

Screenshot of OneMap

Satellite image

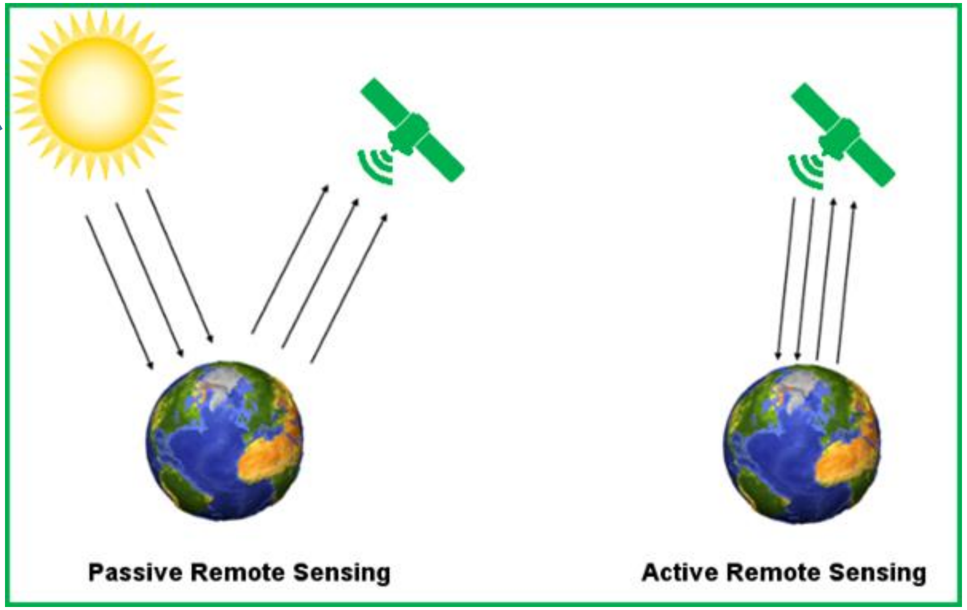


DEM

Activate and passive remote sensing

- Passive sensors record electromagnetic radiation that is reflected or emitted from the terrain
- Active sensors such as microwave (RADAR), Light Detection and Ranging (LIDAR), or sound navigation ranging (SONAR) bathe the terrain in machine-made electromagnetic energy and then record the amount of radiant flux scattered back toward the sensor system.

The Sun represents the initial source of most of the electromagnetic energy recorded by remote sensing systems (except RADAR, LIDAR, and SONAR)



Slides for education purposes only

Satellite images

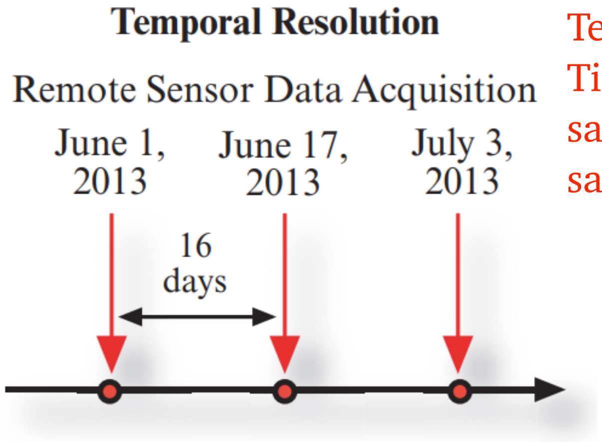
There are many different satellites with different **spatial** and **temporal** resolutions. They are designed by different **countries**, and have multiple **numbers of bands** with different wavelength ranges.

Satellite	Country	Bands	Resolution (m)	Spectral range (um)	First Launch time
Landsat	USA	~ 8	15 - 120	0.45 – 2.35	1972
Spot	France	4	1.5 - 20	0.50 – 0.89	1986
GeoEye	Digital Globe/USA	5	0.4 – 1.6	0.45 – 0.92	2008
MODIS	NASA/USA	36	250 - 1000	0.62 – 14.4	1999

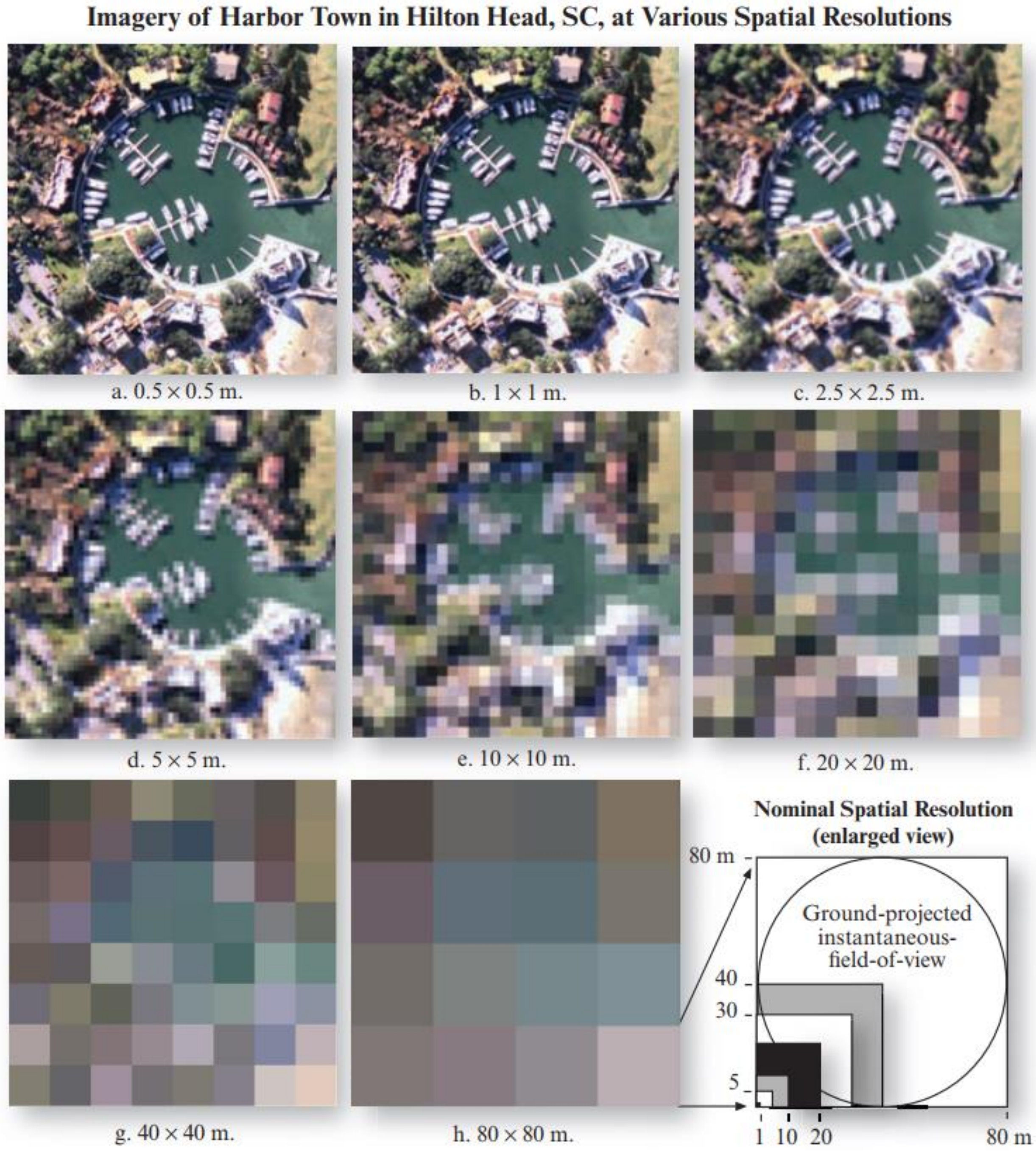
High-resolution image

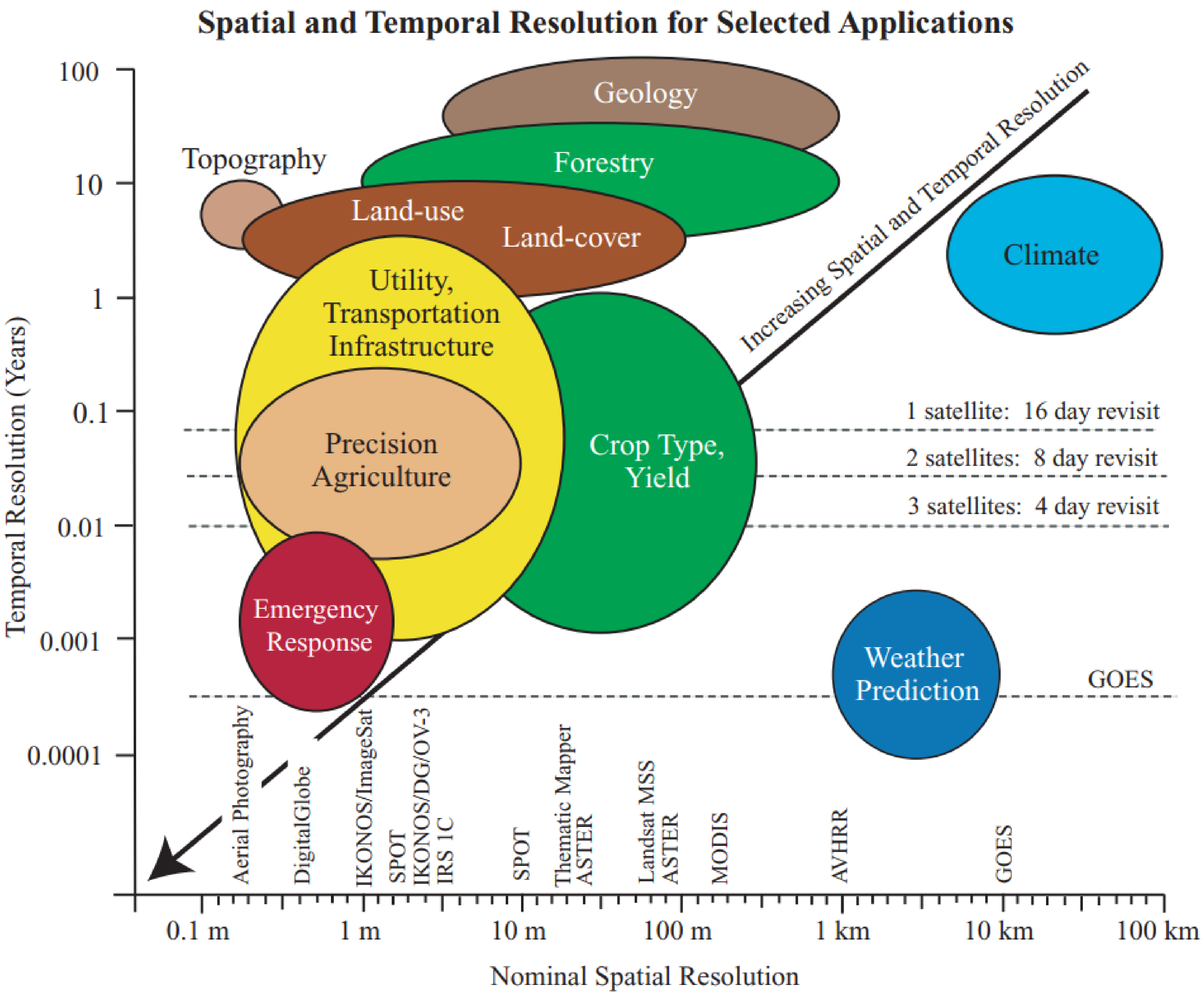
The different satellite parameters determine that they were designed for different purposes and used in different cases

Spatial and temporal resolutions



Temporal Resolution: Time interval that the satellite revisits the same area

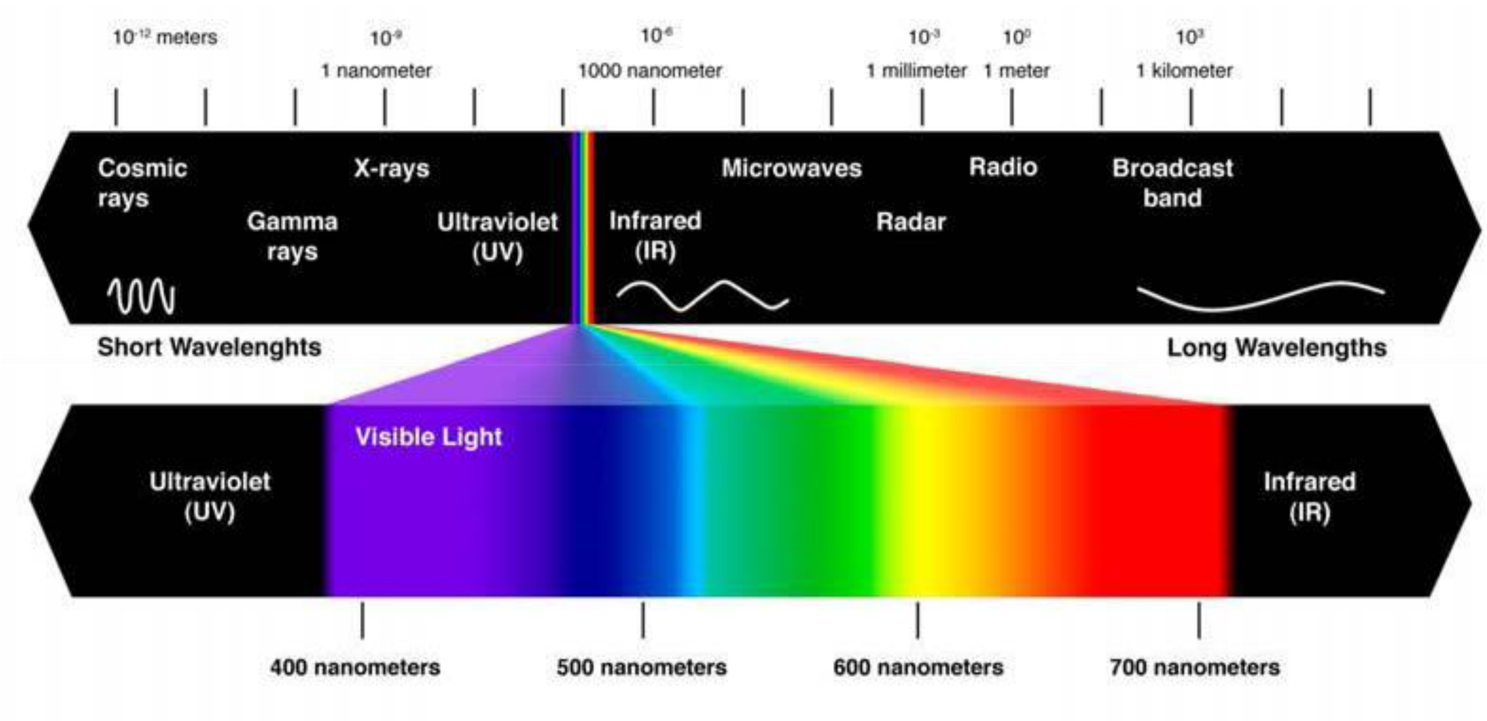


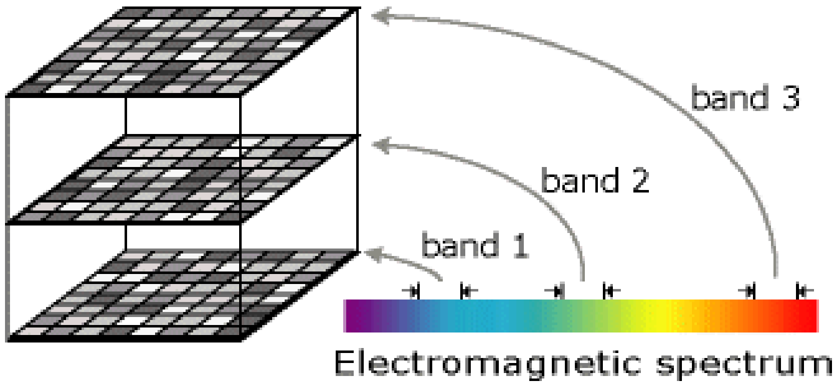


des for education purposes only

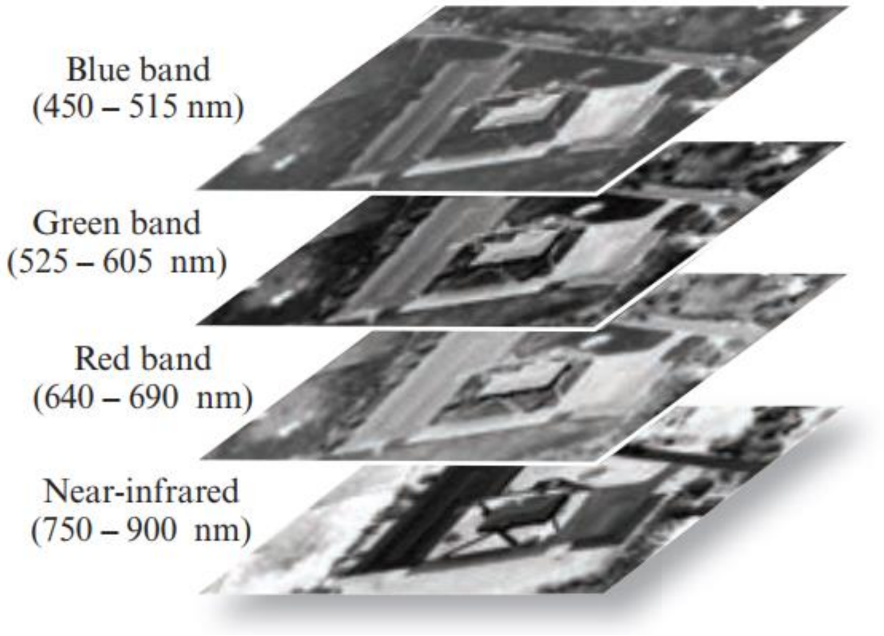
The electromagnetic spectrum

- We often specify a particular region of the electromagnetic spectrum (e.g., red light) by identifying a beginning and ending wavelength and then attaching a description.
- This wavelength (or frequency) interval in the electromagnetic spectrum is commonly referred to as a **band**, channel, or region.

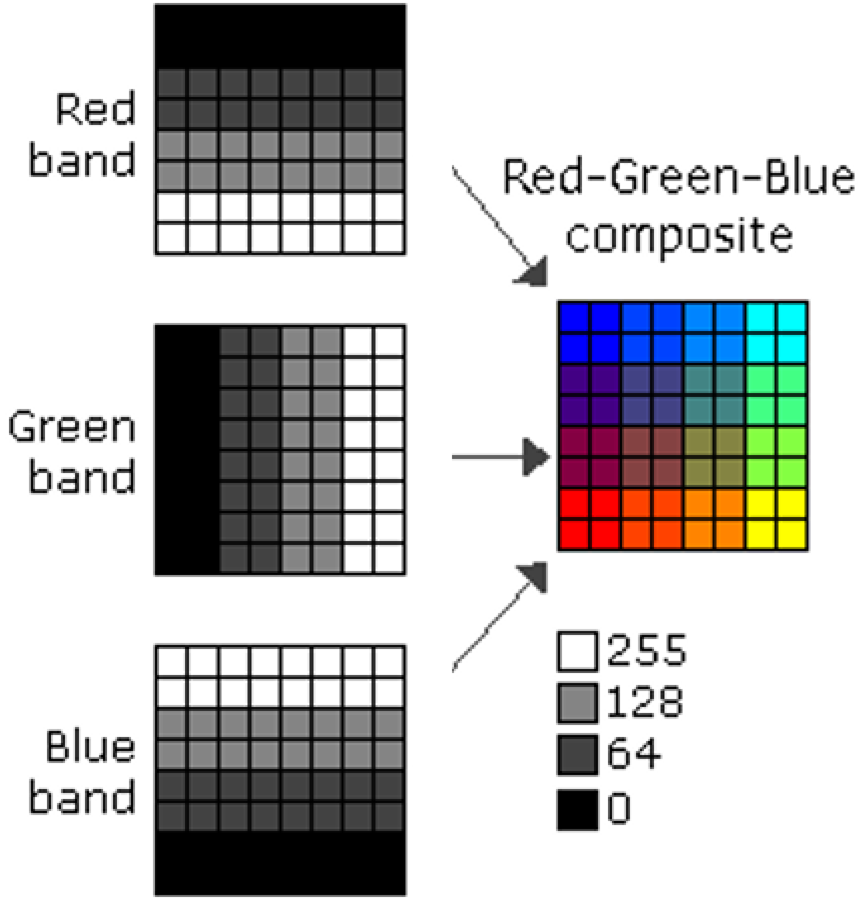




Multispectral remote sensing systems

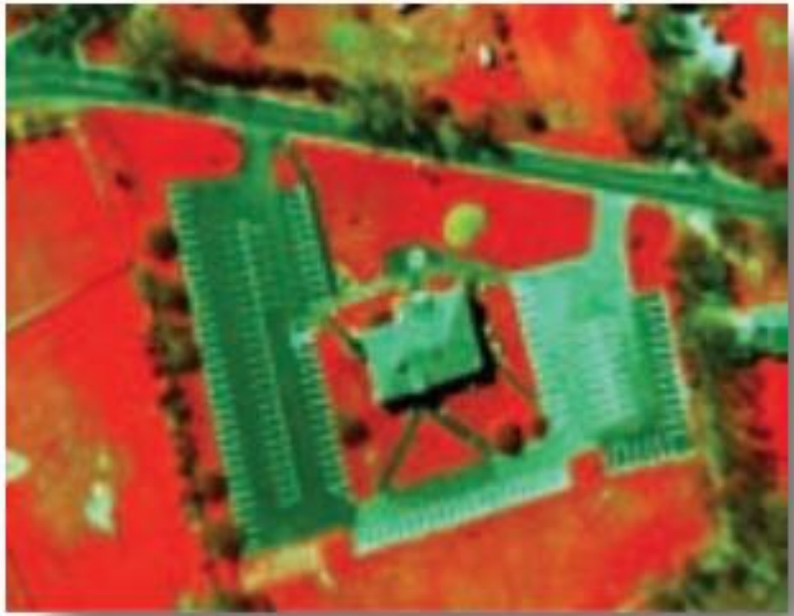


Individual multispectral band images.



Multispectral remote sensing systems

d. Natural color-composite created using blue, green, and red bands.

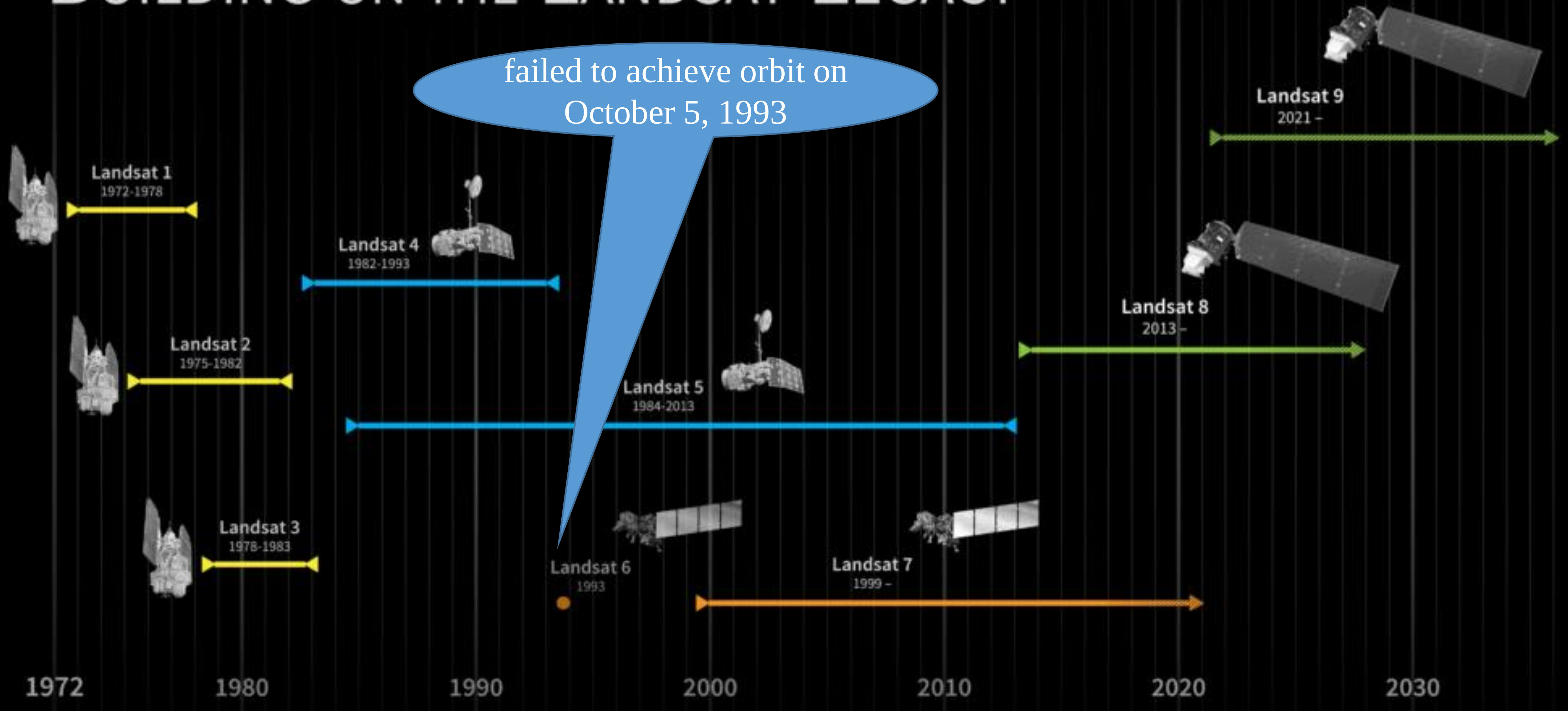


e. Color-infrared color-composite created using green, red, and near-infrared bands.

Landsat

- The US **Landsat** program, started by National Aeronautics and Space Administration (**NASA**) and the US Geological Survey (**USGS**) in 1972 has produced the most widely used imagery worldwide
- The Landsat images are free for download (http://landsat.usgs.gov/)
- **Landsat 8**, launched in 2013 with two sensors: Operational Land Imager (**OLI**), and the Thermal Infrared Sensor (**TIS**). The TIS provides two thermal bands (band 10 and band 11), both with 100 m spatial resolution (for measuring the surface temperature)

Building on the Landsat Legacy



failed to achieve orbit on October 5, 1993

SPOT

- The French **SPOT** satellite series began in 1986. Each **SPOT** satellite carries two types of sensors, capturing multi-spectral bands and the panchromatic band

Satellite	Launch Year	No. of MS bands [nominal resolution]	Panchromatic resolution (nominal)	Altitude (km)	Revisit time (days)
SPOT 1	1986	3 [20 m]	1 [10 m]	832	2-3
SPOT 2	1990	3 [20 m]	1 [10 m]	832	2-3
SPOT 3	1993	3 [20 m]	1 [10 m]	832	2-3
SPOT 4	1998	4 [20 m]	1 [10 m]	832	2-3
SPOT 5	2002	4 [10 m]	1 [2.5-5 m]	822	2-3
SPOT 6	2012	4 [6 m]	1 [1.5 m]	694	1
SPOT 7	2014	4 [6 m]	1 [1.5 m]	694	1

A panchromatic image is a single-band grayscale image with a high spatial resolution that “combines” the information from the visible R, G, and B bands.

GeoEye

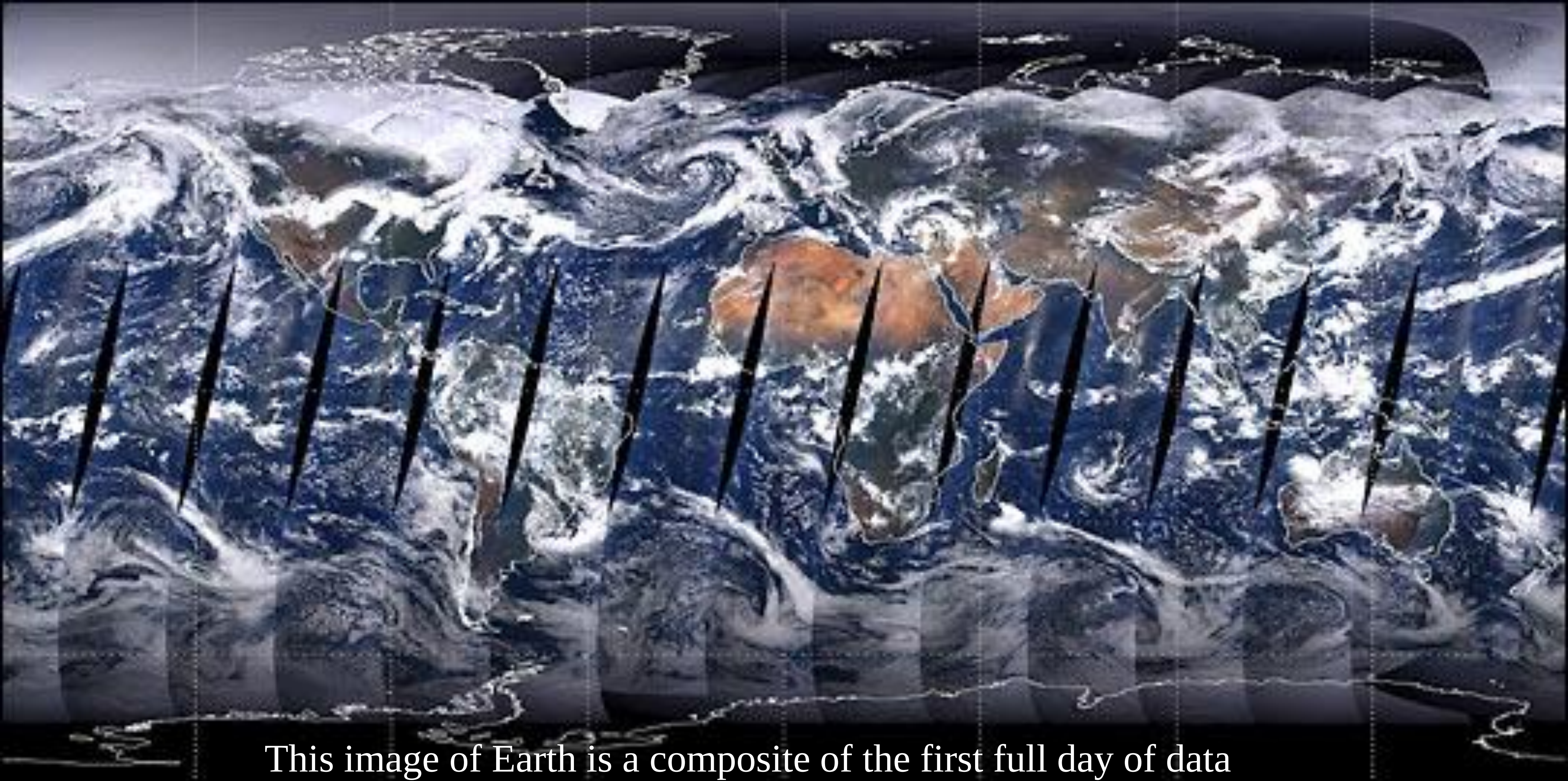
- **GeoEye** was an American commercial company, and was merged into **Digital Globe** company in 2013.
- **Digital Globe** has expanded the products, including Ikonos, QuickBird, GeoEye-1, and WorldView (1-4). They all belong to high-resolution satellite images
  - One of the Google earth image providers

Digital Globe

IKONOS*		GeoEye-1	
Panchromatic 82 cm	Multispectral 4 m	Panchromatic 41 cm	Multispectral 1.65 m
QuickBird*		**WorldView-4**	
Panchromatic 61 cm	Multispectral 2.4 m	Panchromatic 31 cm	Multispectral 1.24 m

MODIS

- The **Moderate Resolution Imaging Spectroradiometer** (MODIS) was launched by **US NASA** in 1999.
- It captures data in 36 spectral bands ranging in wavelength from 0.4 um to 14.4 um and varying spatial resolution (2 bands at 250 m, 5 bands at 500 m, and 29 bands at 1 km)

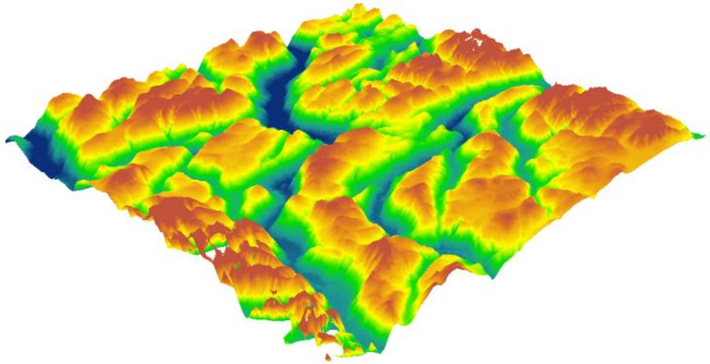


This image of Earth is a composite of the first full day of data gathered by the Moderate Resolution Imaging Spectroradiometer on the Terra spacecraft.

Digital Elevation Model (DEM)

- A **digital elevation model** (DEM) consists of an array of uniformly spaced elevation data
- DEMs are a data source between 2-dimensional (**2D**) and **3D**
  - They represent the **terrain surface** in a 3D environment
  - They are often called **2.5D** data
    - Data only exists on terrain surface
    - Data does not fill up the entire 3D space/environment
- DEMs are a primary data source for terrain mapping and analysis

2.4	2.5	2.3	2.6	2.7	3.1	3.3
2.5	2.6	2.5	2.5	2.8	3.0	4.1
2.6	2.6	2.5	2.7	3.1	3.3	3.8
2.5	2.2	2.6	2.5	2.5	2.8	3.0
2.9	2.8	2.7	2.9	2.4	3.1	3.3
2.9	2.5	2.7	2.7	2.8	3.3	3.4
2.6	2.6	2.5	2.7	3.1	3.3	3.8



Digital Elevation Model (DEM)

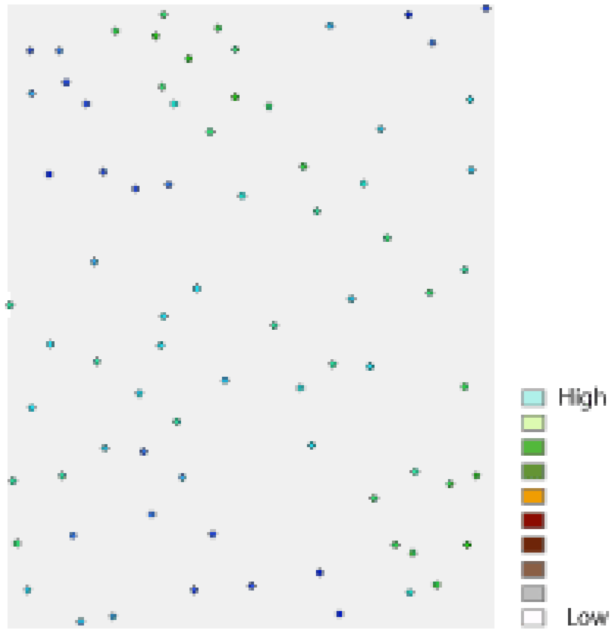
- Techniques for DEM generation
  - **Surveying** (total station & GPS)

GPS

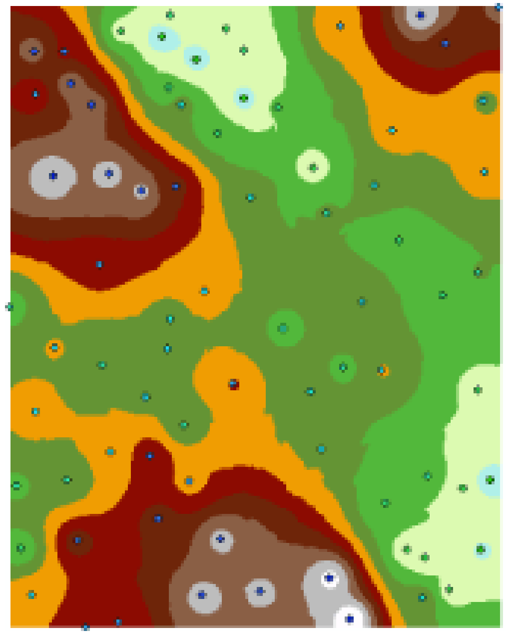
Total station

Digital Elevation Model (DEM)

- Interpolation



Input elevation point data



Interpolated elevation surface

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/understanding-interpolation-analysis.htm

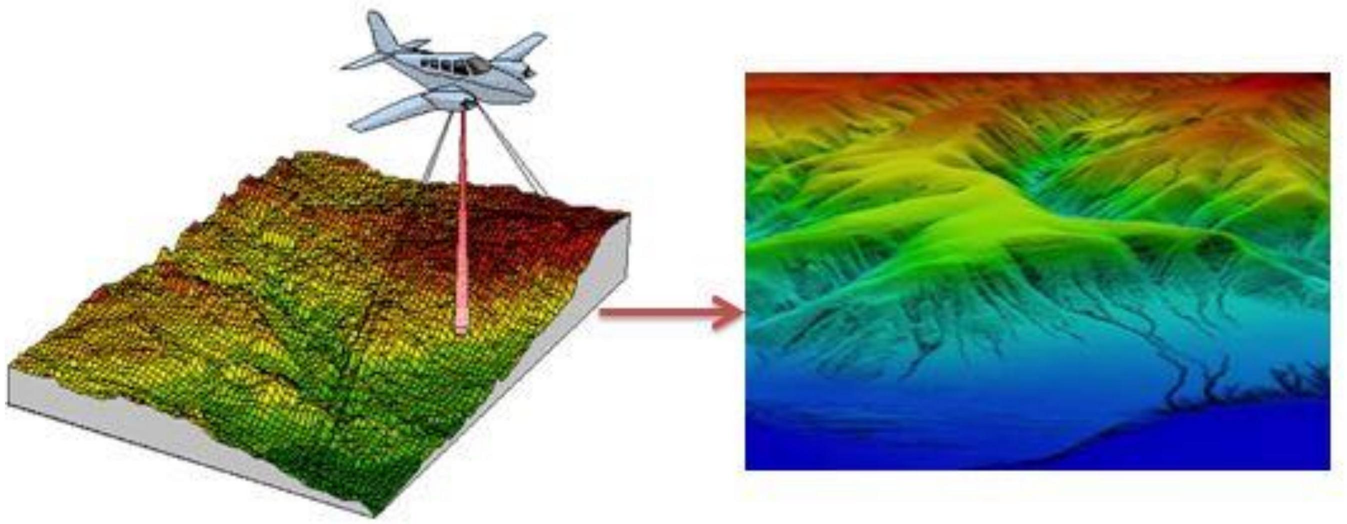
- TIN To Raster

https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-tin-to-raster-3d-analyst-works.htm

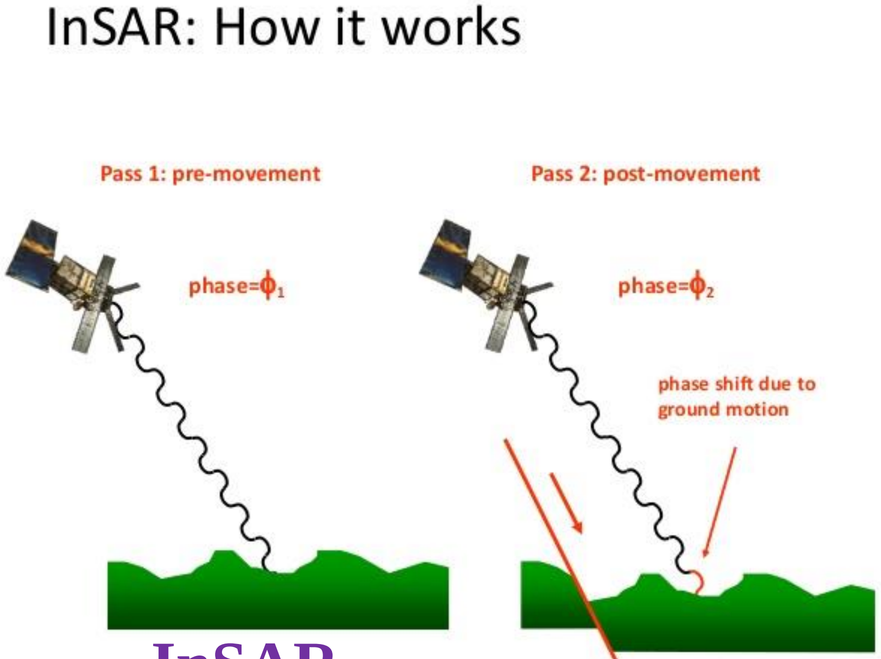
Please read this page: https://storymaps.arcgis.com/stories/7c293adf1b54436eaa07065e04db0f27

Digital Elevation Model (DEM)

- Techniques for DEM generation
  - **LiDAR** (Light Detection and Ranging)
  - **InSAR** (Interferometric synthetic-aperture radar)



LiDAR

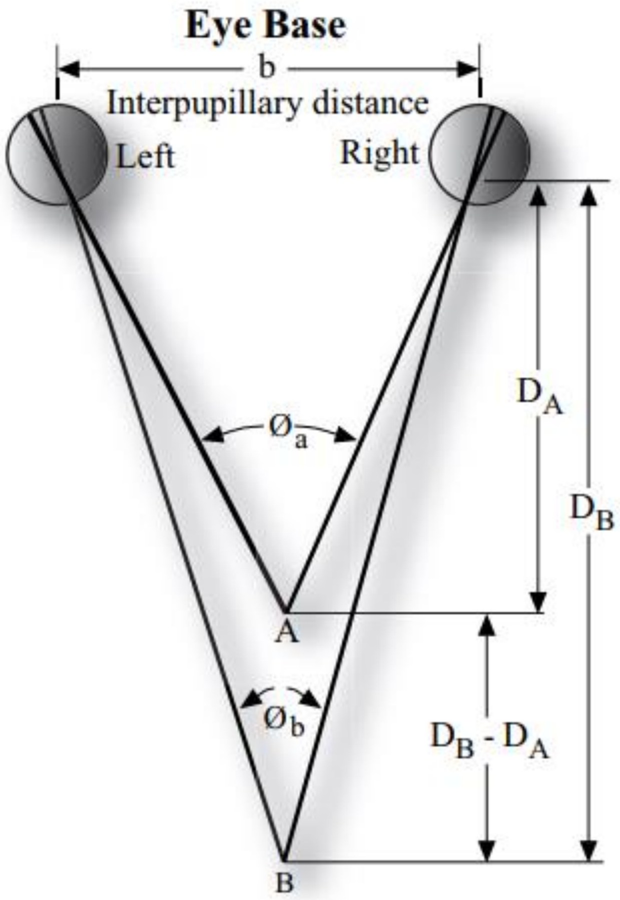


InSAR

Digital Elevation Model (DEM)

- Techniques for DEM generation
  - **Stereo pairs** of optical images (Photogrammetry)

Stereo images



Outline of this lecture

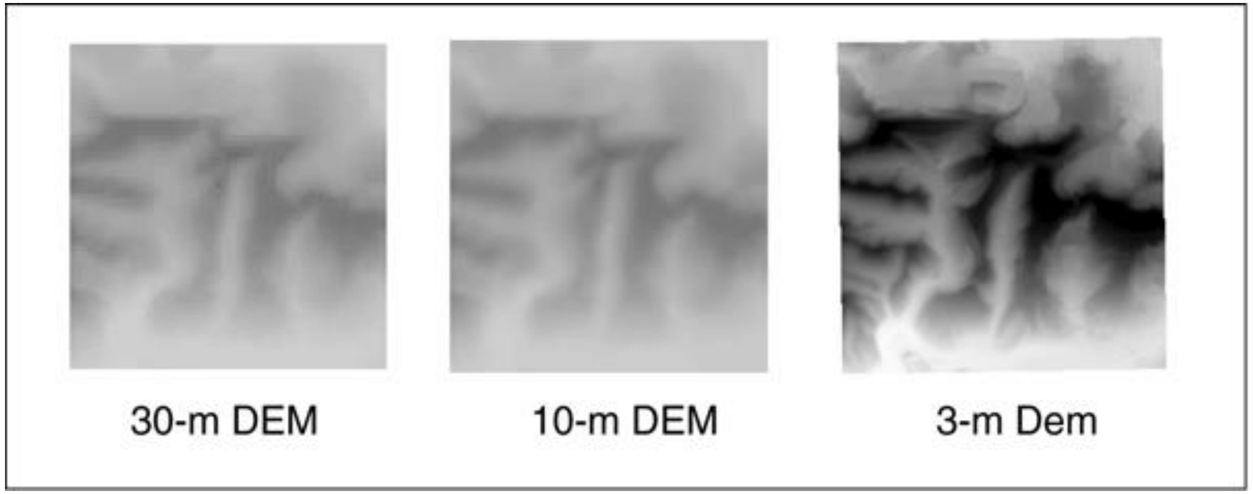
- Concept of the raster data model
- Elements of the raster data model
- Raster data structure
- Data conversion between raster and vector data

Cell size

- The cell size of a raster refers to the size of the area represented by a single cell (Area)
- A smaller cell size means a higher precision of location
- A larger cell size may increase the chance of having mixed features such as forest, pasture, and water in the same cell, which means the cell is not pure

Spatial resolution

- The spatial resolution refers to the size of the smallest feature that can be detected
- The spatial resolution equals to the length represented by a cell

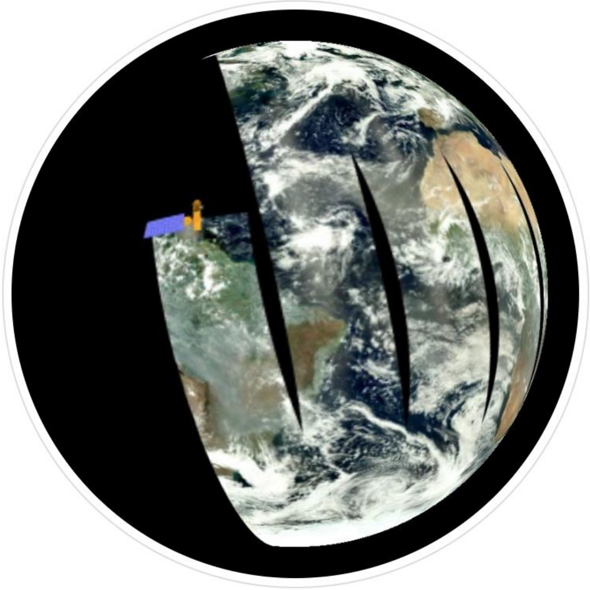


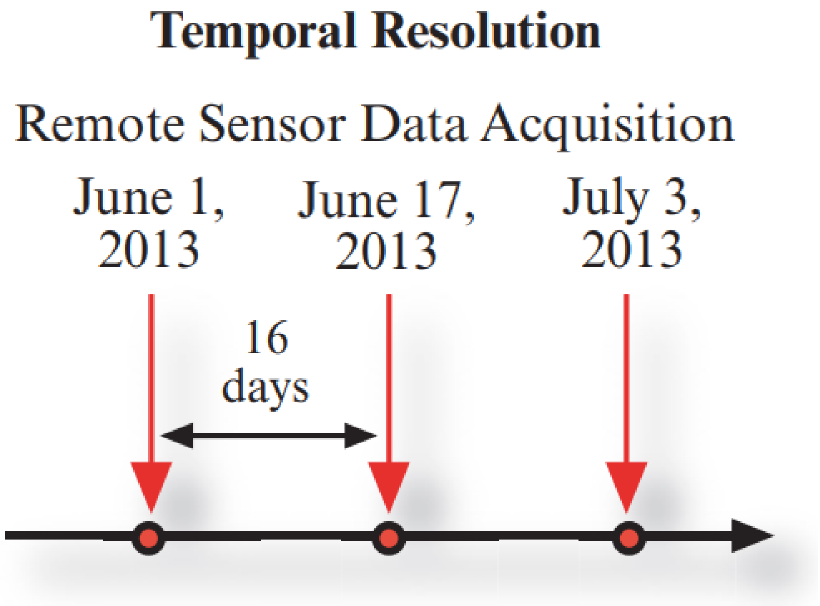
Shorter lens means higher resolution

Which DEM has the highest spatial resolution?

Resolution of satellite images

- Apart from spatial resolution, satellite images also have temporal resolution and spectral resolution
- Temporal resolution – refers to the time interval that the satellite revisit the same geographic area





Resolution of satellite images

- Apart from spatial resolution, satellite images also have temporal resolution and spectral resolution
- Spectral resolution – refers to the ability of a satellite sensor to measure specific wavelengths
  - Number of MS bands
  - Interval of the wavelength

	Bands	Wavelength (micrometers)	Resolution (meters)
Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Launched February 11, 2013	Band 1 - Coastal aerosol	0.43 - 0.45	30
	Band 2 - Blue	0.45 - 0.51	30
	Band 3 - Green	0.53 - 0.59	30
	Band 4 - Red	0.64 - 0.67	30
	Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
	Band 6 - SWIR 1	1.57 - 1.65	30
	Band 7 - SWIR 2	2.11 - 2.29	30
	Band 8 - Panchromatic	0.50 - 0.68	15
	Band 9 - Cirrus	1.36 - 1.38	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100

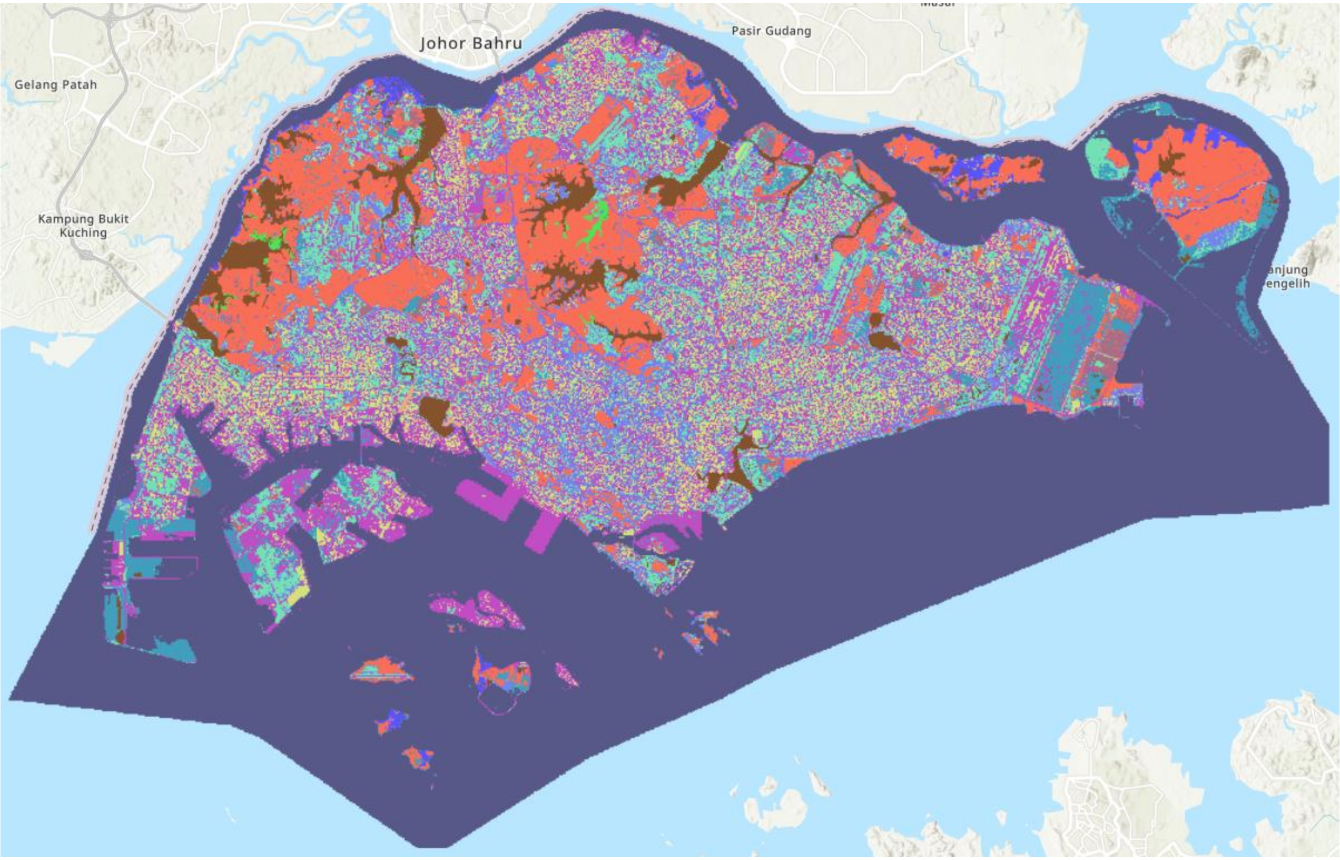
Temporal resolution 16 days

Cell value

- Cell values in a raster can be categorical or numerical

Categorical values

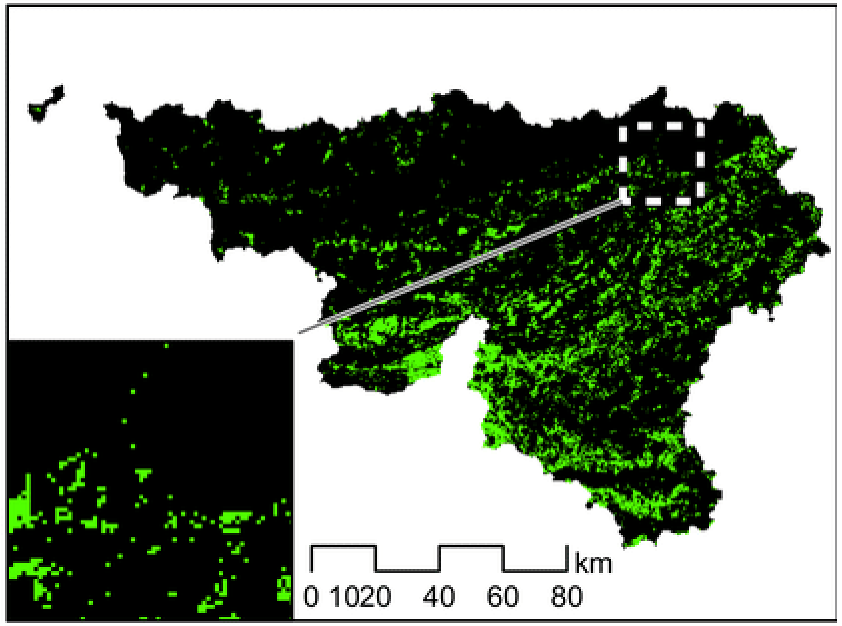
Each value represents one land cover type

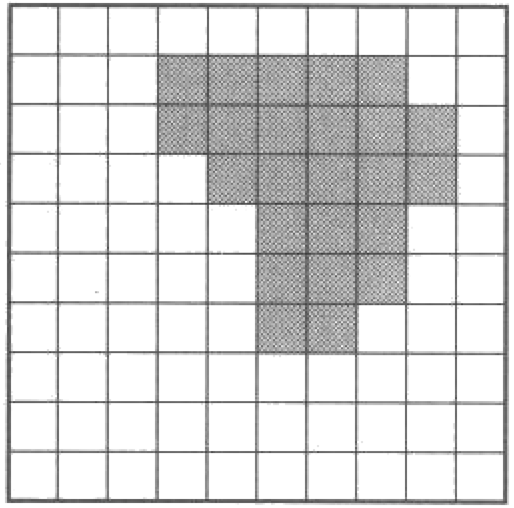


Land cover map

Cell value

- Bi-level map Only 2 categories





(a)

0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	0	0
0	0	0	1	1	1	1	1	1	0
0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

(b)

Non-forest (background)

Forest (foreground)

Bi-level map

Cell value

- Numerical values
  - Integer
  - Float/double
    - Float – 32-bit floating point precision (has a precision of 6 decimal places)
    - Double – 64-bit floating point precision (has a precision of 15 decimal places)
- Usually, a **float/double** raster requires more computer memory than an **integer** raster

https://www.w3schools.in/difference-float-double-data-types/

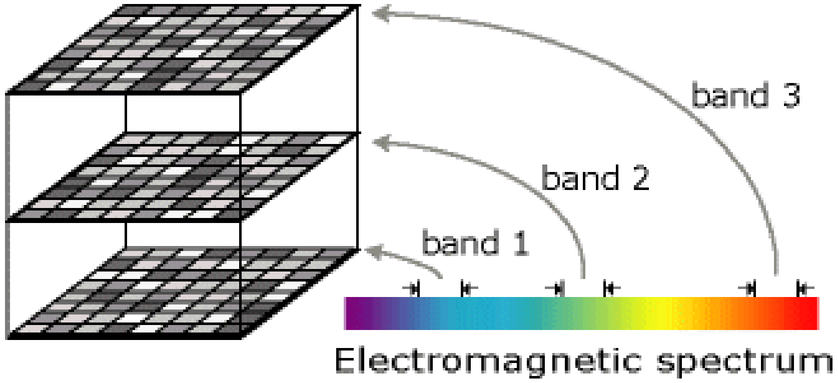
Cell depth

- The bit depth (pixel depth) of a pixel determines the range of values that a particular raster file can store, based on the formula 2^n (where n is the bit depth)
  - A **deeper** pixel depth means a **wider** range of values
  - An 8-bit raster can store 256 (2^8) values (0-255)
  - A 16-bit raster can store ~~65,356~~ 65,536 (2^{16}) possible values (0 - 65535)
  - **One byte** equals to **8 bits**

An 8-bit DEM raster

Raster bands

- A raster may have a single band or multiple bands
  - A DEM raster has only one band, with one elevation value at each cell location
  - A Quickbird satellite image has 5 bands: red, green, blue, near infrared, and a panchromatic band
  - Landsat 8 has 11 bands
- Each cell in a single-band raster has only one cell value



Spatial reference

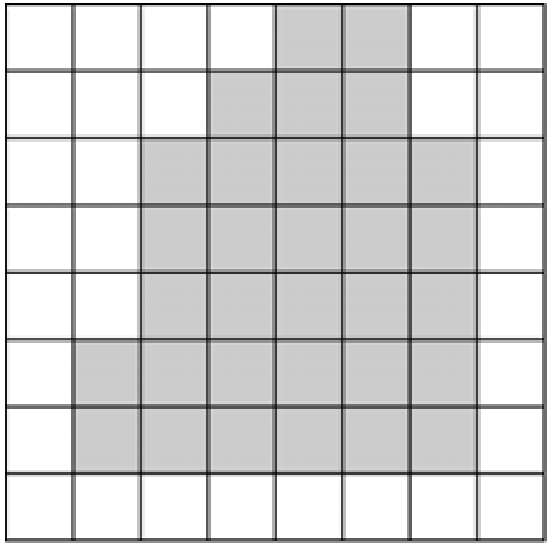
- Raster data must have the spatial reference information so that
  - The geo-location of each cell can be known
  - The raster can align spatially with other data sets in GIS, e.g., to superimpose a DEM on a land use map
- The spatial reference is the main item that distinguishes raster images from common images
  - Common image format: JPEG, TIFF, PNG…
  - Geographic raster image format: GeoTIFF, Esri Grid…
  - More about spatial reference will be introduced in Lecture **5**

Outline of this lecture

- Concept of the raster data model
- Elements of the raster data model
- Raster data structure
- Data conversion between raster and vector data

Raster data structure: cell-by-cell encoding

- Raster data structure refers to the method by which raster data are encoded and stored in the computer
- **Cell-by-cell** encoding
  - Provides the simplest raster data structure
  - A raster is stored as a matrix
  - DEMs and satellite images are all encoded cell-by-cell

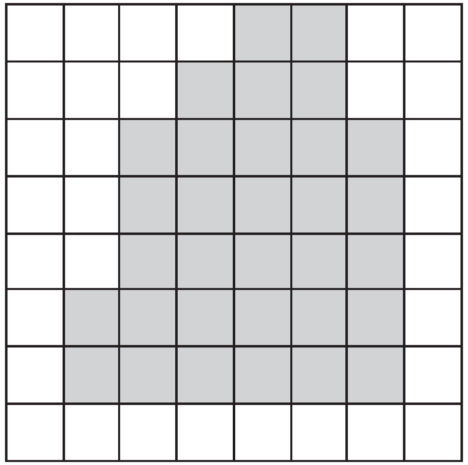


Row 1: 0 0 0 0 1 1 0 0
Row 2: 0 0 0 1 1 1 0 0
Row 3: 0 0 1 1 1 1 1 0
Row 4: 0 0 1 1 1 1 1 0
Row 5: 0 0 1 1 1 1 1 0
Row 6: 0 1 1 1 1 1 1 0
Row 7: 0 1 1 1 1 1 1 0
Row 8: 0 0 0 0 0 0 0 0

Raster data structure: run-length encoding

- **Run-Length Encoding** (RLE)
  - Cell-by-cell encoding becomes **inefficient** if a raster contains many redundant cell values
  - **RLE** records cell values by row and group. A group refers to a series of adjacent cells with the same cell value
  - For each group, its **starting** and **ending** cell number are recorded

A bi-level forest map

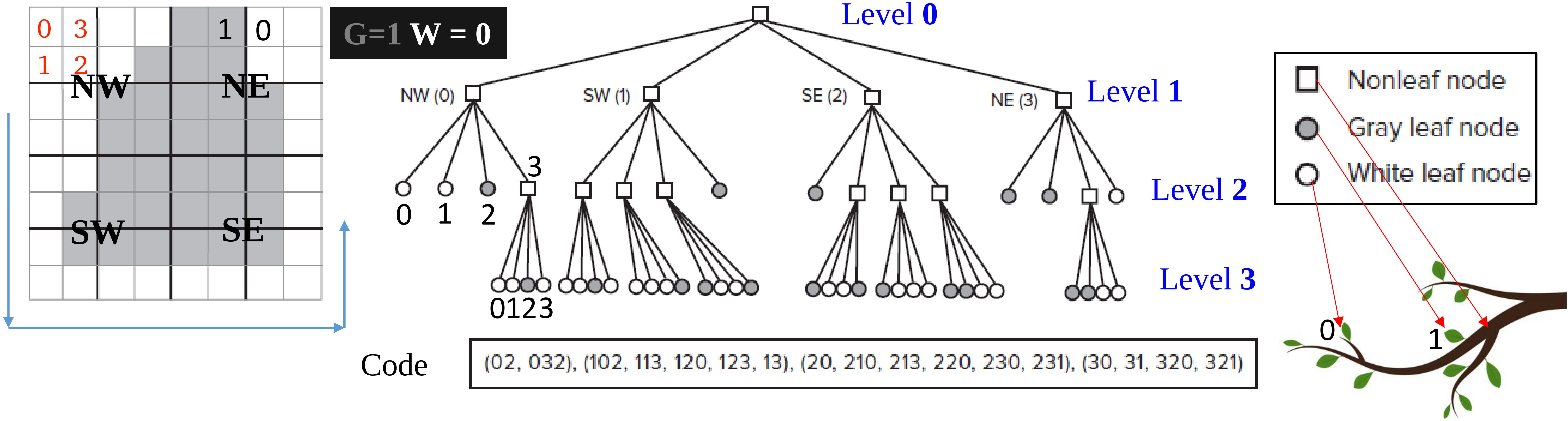


Row 1 : 5 6
Row 2: 4 6
Row 3: 3 7
Row 4: 3 7
Row 5: 3 7
Row 6: 2 7
Row 7: 2 7

RLE is efficient if there are many repetitive cell values

Raster data structure: Quadtree encoding

- Quadtree uses recursive decomposition to divide a raster into a hierarchy of quadrants until every quadrant in a quadtree contains only one cell value



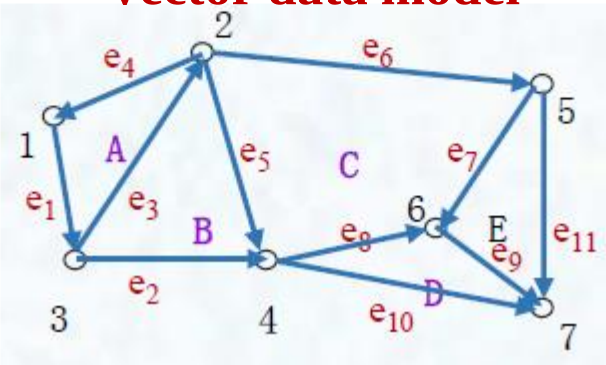
It is straightforward in representing nested relations

Outline of this lecture

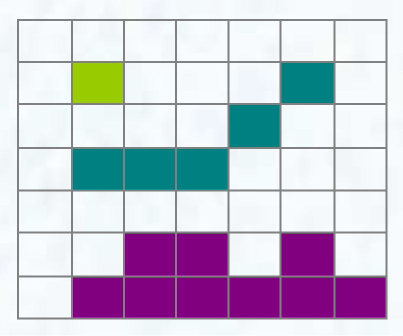
- Concept of the raster data model
- Elements of the raster data model
- Raster data structure
- Data conversion between raster and vector data (optional – for your interest only)

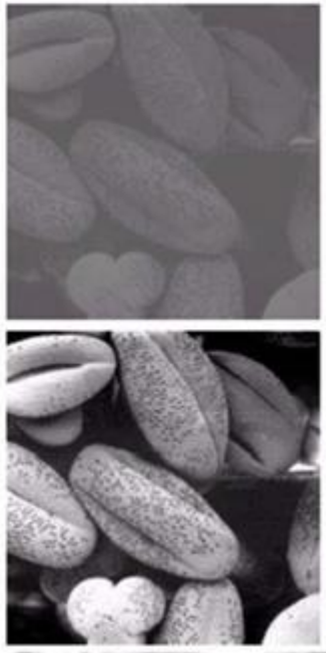
Vector and raster data model: **pros and cons**

Vector data model



Raster data model

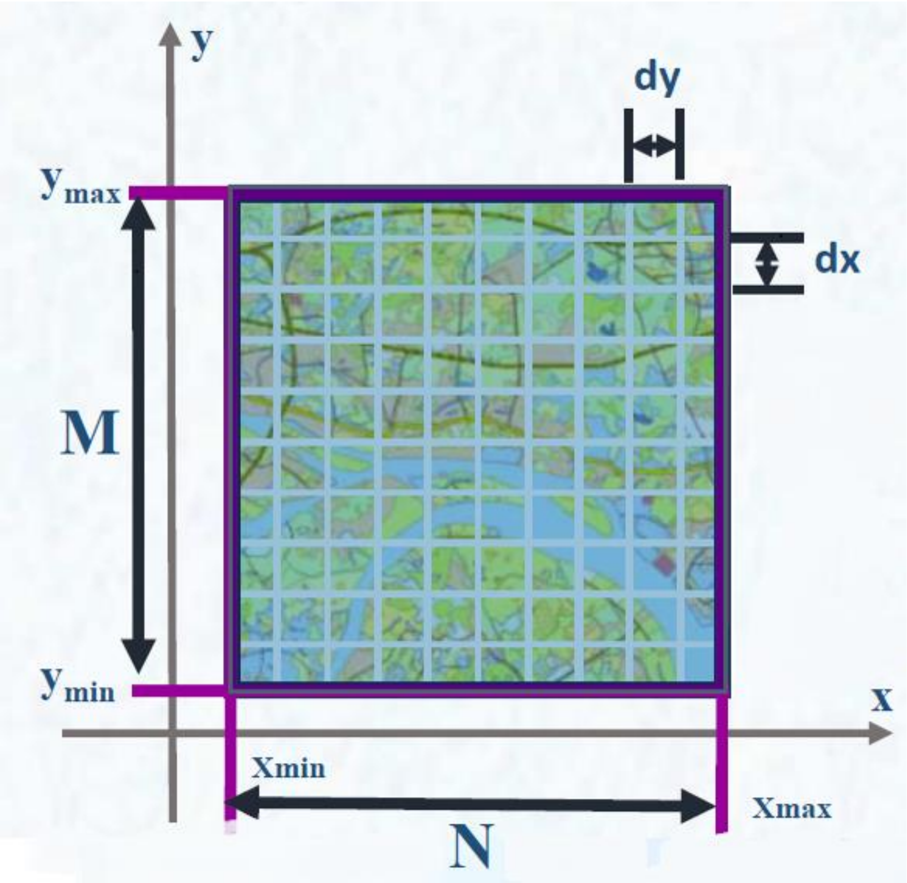




	Vector data model	Raster data model
Advantages	• The structure is stricter • Easy for topological process (e.g., query) • More delicate output (not affected by zooming in and out)	• The structure is simple • Easy to perform overlaying analysis • Easy to express spatial changes • Enhancement can be easily conducted
Disadvantages	• The structure is more complicated • Not convenient to perform overlaying analysis • Poor in expressing spatial changes	• Hard to express topological relations • The data volume is large • The output is affected by zooming in and out

From vector to raster data: Rasterization

First step: Set up a raster with a specified **pixel size**

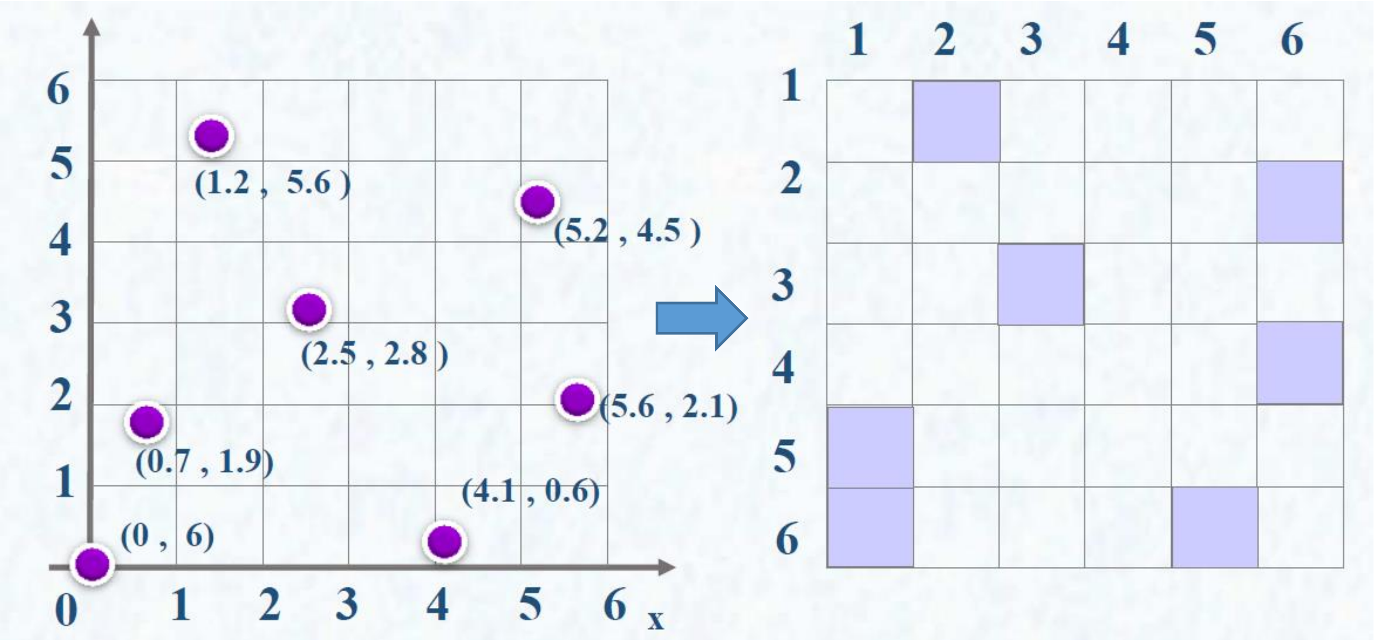


- x_{min}, x_{max}, y_{min}, y_{max}: the boundary extent of the graphic data
- d_x, d_y: the specified **pixel size** of the raster along x and y axis
- M, N: the number of rows and columns after conversion

$$M = |y_{max} - y_{min}|/d_y$$
$$N = |x_{max} - x_{min}|/d_x$$

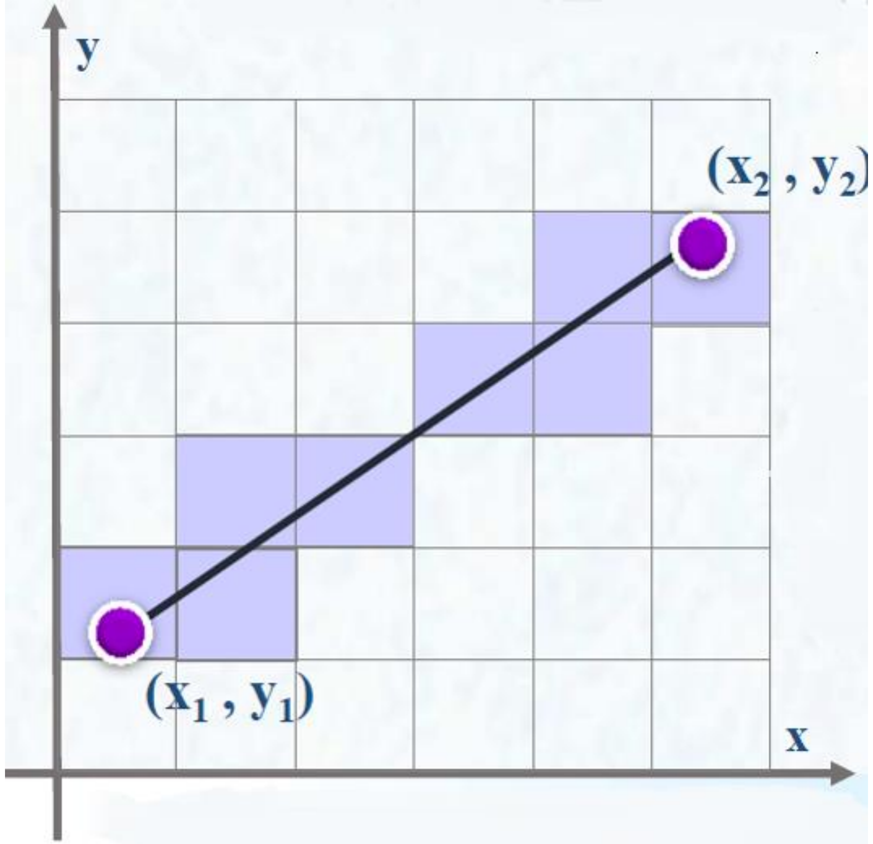
Rasterization: Points

Each point corresponds to one pixel in the raster data. Assume that the coordinate of a point is (x, y), and its corresponding cell location is (I, J), then $I = 1 + INT\left[\frac{y_{max} - y}{d_y}\right]$; $and\ J = 1 + INT\left[\frac{x - x_{min}}{d_x}\right]$



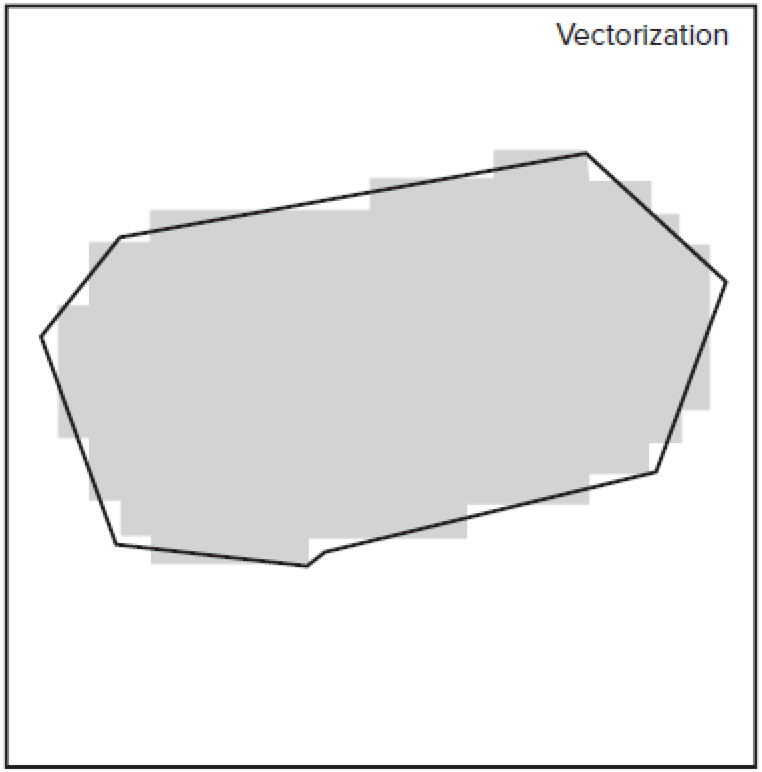
Note that the origin in the vector and raster data model is slightly different

Rasterization: Lines



- A line can be deemed as a set of points by interpolating points at a certain interval
- $x_{i+1} = x_i + d_x$
- $y_{i+1} = f(x_{i+1})$, where $y = f(x)$ is the function of the line
- So, the line can be represented as a set of points $[(x_1, y_1), \ldots (x_i, y_i), \ldots (x_2, y_2)]$
- Rasterize all the points in the point set

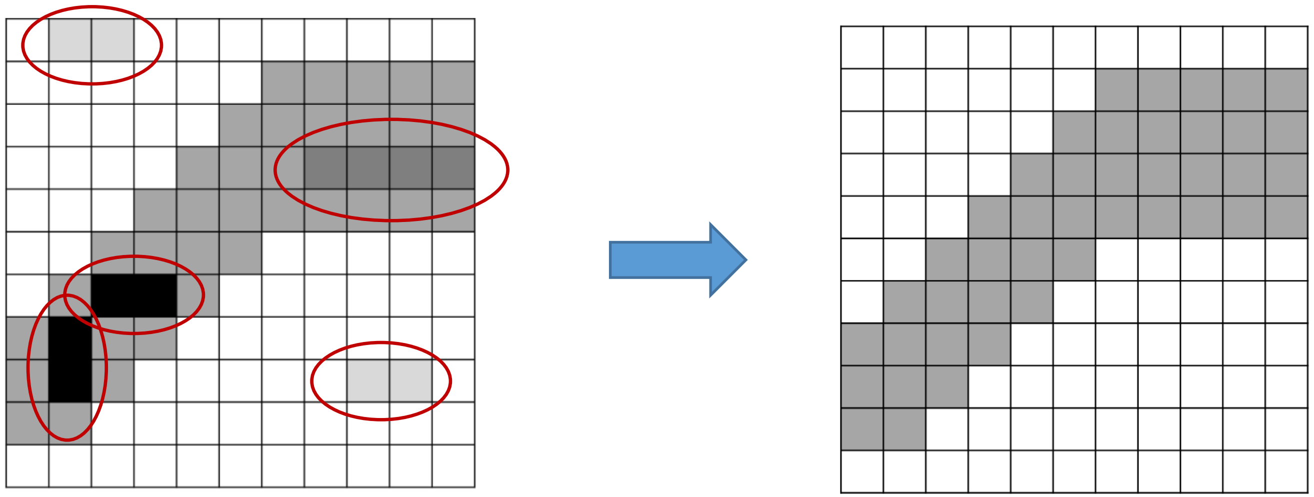
Rasterization: Polygons



- Extract **boundary outlines** of the polygon
- Rasterize the boundary lines
- Fill the **interior** of the boundary outline with the same **cell value**

From raster to vector data: Vectorization

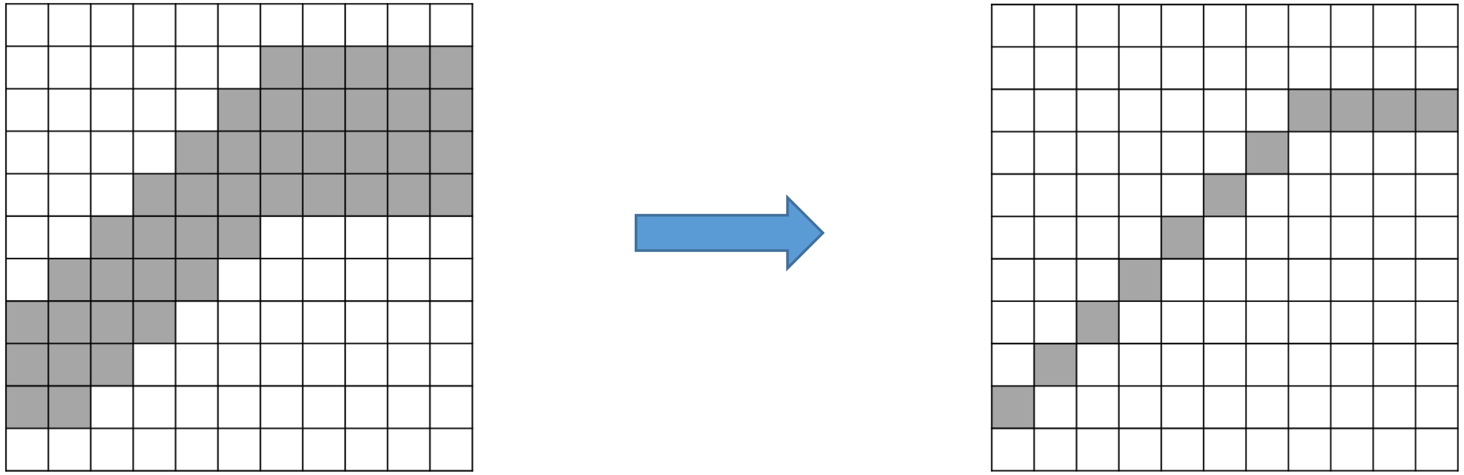
Image Binarization → **Line Thinning** → **Line Tracking** → **Topological Reconstruction**



$$\text{if } G(i,j) \geq T,\ \text{then } B(i,j)=1 \text{ else } B(i,j)=0$$

From raster to vector data: Vectorization

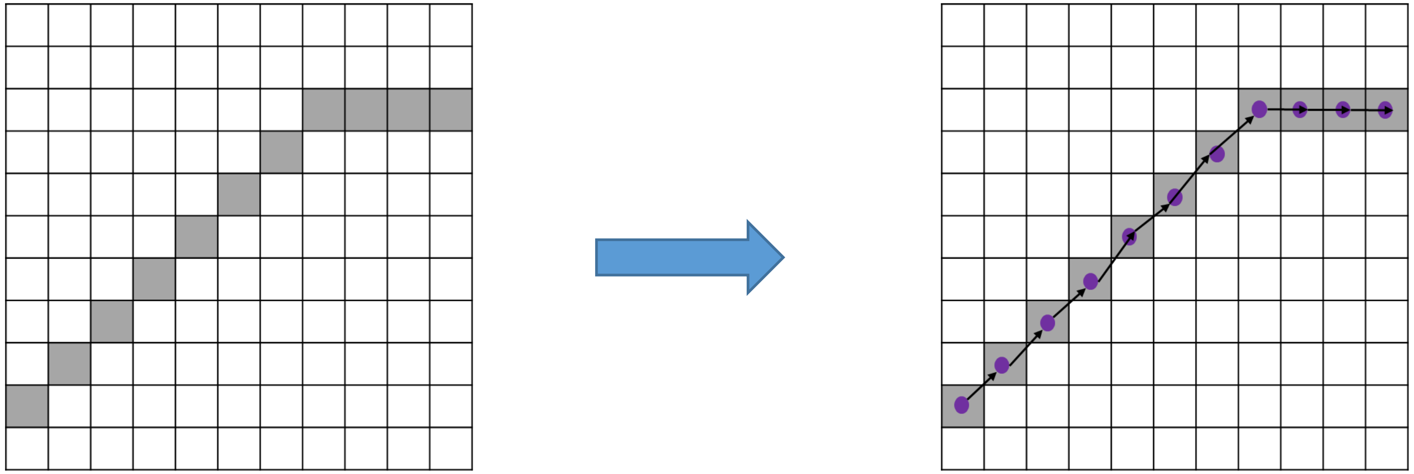
Image Binarization → **Line Thinning** → Line Tracking → Topological Reconstruction



Lines in the vector data model have length but no width.
Therefore raster lines must be thinned to a 1-cell width

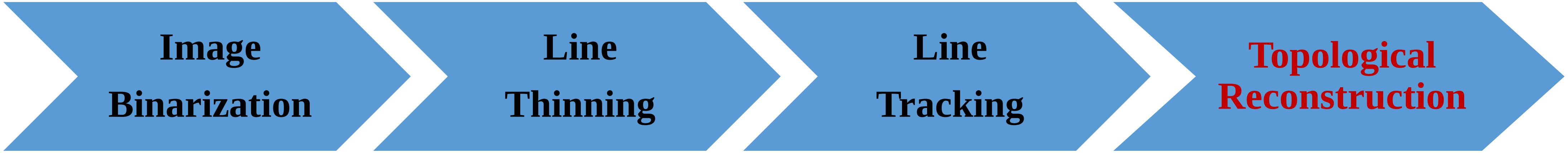
From raster to vector data: Vectorization

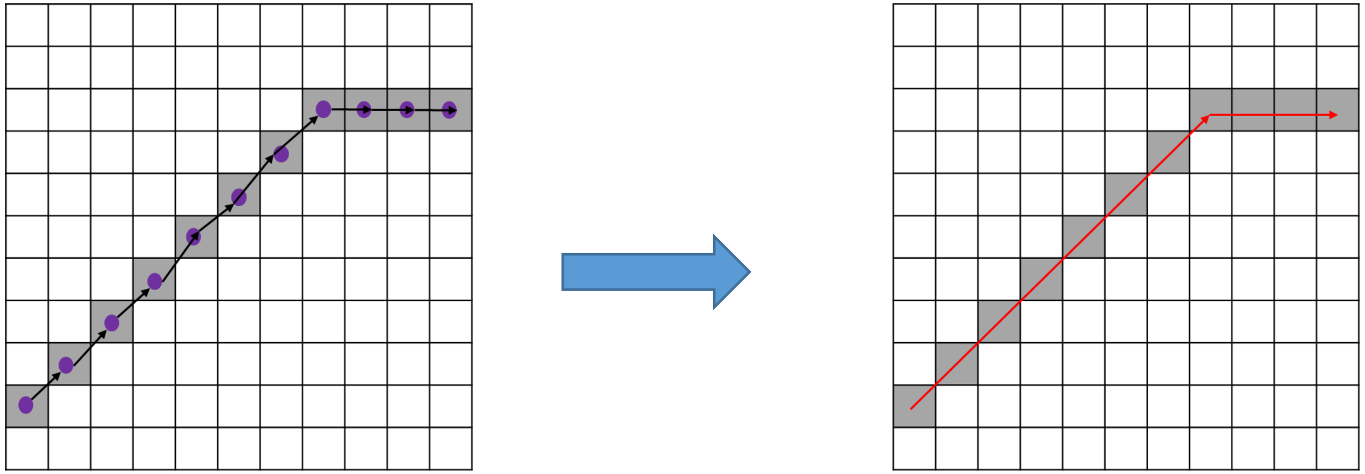
Image Binarization → **Line Thinning** → **Line Tracking** → **Topological Reconstruction**



Line tracking is the process of determining where individual lines begin and end

From raster to vector data: Vectorization





Topological reconstruction connects tracked lines, and examines the topological relation, e.g., remove middle points, redundant points

Summary

- Concept of the raster data model
  - Basic unit: cell or pixel
  - Why raster data model?
  - Types of raster: Scanned maps, satellite images , DEMs
- Elements of the raster data model
  - Cell size, resolution, cell value, cell depth, raster band, spatial reference
- Raster data structure
  - Cell-by-cell, run-length, and quadtree encoding
- Data conversion between raster and vector data
  - Advantages and disadvantages of raster and vector data model
  - Rasterization and Vectorization

THANK YOU