Chapter 2
Geographic information and
Spatial data types
Three types of phenomena

- Real World
- Computer representations
- Visualizations

This chapter studies geographic phenomena more deeply, and looks into the different types of computer representations of them.
Spatial Data modeling

Geographic phenomena → Computer representations → Visualization
Spatial Data modelling
This chapter

• This chapter studies geographic phenomena more deeply, and looks into the different types of computer representations of them.

• Any geographic phenomenon can be represented in various ways; how do we determine which representation we are going to use?

• What original raw data is available

• What sort of data manipulation does the application want to perform.
We might define a geographic phenomenon as something of interest that:

- Can be named or described
- Can be georeferenced
- Can be assigned a time (interval) at which it is/was resent

What if one of them is missing?
Different types of geographic phenomena

Observation 1: The representation in a GIS requires us to state *what* it is, and *where* it is.

The reason why we forget about temporal issues is that current GIS do not provide much automatic support for time-dependent data.
Different types of geographic phenomena

Observation 2: some phenomena manifest themselves essentially everywhere in the study area, while others only occur in certain localities.

1. A (geographic) **field** is a geographic phenomenon for which, for every point in the study area, a value can be determined. (temperature, barometric pressure and elevation)

2. (Geographic) **objects** populate the study area, and are usually well distinguishable, discrete, bounded entities. The space between them is potentially empty.
Spatial Data modeling

Geographic phenomena
- Object
- Field

Computer representations

Visualization
Geographic fields

- Fields can be *discrete* or *continuous*, and if they are continuous, they can even be differentiable.

In a **continuous** field, the underlying function is assumed to be continuous. (temperature)

Continuity means that all changes in field values are gradual.

In a differentiable field we can determine a measure of change per unit of distance anywhere. (for elevation, this measure would be slope)
In a **discrete** field,

- *Discrete fields*, cut up the study space in mutually exclusive, bounded parts, with all locations in one part having the same field value. (land classification, geological units).

- One may note that discrete fields are a step from continuous fields towards geographic objects:
  - *Discrete fields as well as objects make use of ‘bounded’ features.*
  - *A discrete field still assigns a value to every location.*
Continuous fields - Discrete fields

Landuse

Elevation
Different types of geographic phenomena

Fields
- Continuous
  - Temperature

Objects
- Discrete
  - Landuse
  - Buildings
Spatial Data modeling

Geographic phenomena
- Object - Discrete
- Field – Discrete or Continuous

Computer representations

Visualization
Different kinds of data values

- **Nominal data values**, values that provide a name or identifier so that we can discriminate between different values. Also called categorical data. (Landuse)

- **Ordinal data values**, values that we can put in a natural sequence, could be assigned as ‘low’, ‘average’ or ‘high’. (Course satisfaction)

- **Interval data values** do allow computation. It knows no arithmetic zero value, and does not support multiplication or division. (Temperature)

- **Ratio data values**, do allow computation, know arithmetic zero value and do allow multiplication or division. (Population)
Different kinds of data values

Data: ESRI documents – Using ArcMap
Different kinds of data values

- Earthquakes
  - Magnitude
  - 0-5
  - 5-6
  - 6-7
  - 7-8

Ref: ESRI documents – Using ArcMap

Page Number: 42
Different kinds of data values

- Critical
- Endangered
- Relatively Stable

Major Habitat Types:
- Deserts and xeric shrublands
- Mediterranean scrub
- Montane grasslands
- Temperate broadleaf and mix
- Temperate grasslands, savann
- Tropical and subtropical grass
- Tropical and subtropical moist
- Tundra

Ref: ESRI documents – Using ArcMap
Geographic objects

- Their position in space is determined by a combination of one or more of the following parameters:
  - Location (where is it?)
  - Shape (what form is it?)
  - Size (how big is it?)
  - Orientation (in which direction is it facing)

How we want to use the information determines which of the parameters is required to represent it.
Geographic objects

**Shape**

- 0-dimensional
- 1-dimensional
- 2-dimensional
- 3-dimensional

**Size**

**Orientation**
Collections of objects

- We usually do not study geographic objects in isolation but whole *collections of objects*.
  - Different objects do not occupy the same location.
  - Observe that collections of geographic objects can be interesting phenomena at the higher aggregation level.
- Studies of multi-scale approaches are also conducted
Collections of objects

Plot - object

Block – collection of objects.

Collection of blocks will form a suburb (higher aggregation level)
Boundaries

Which phenomena have boundaries:
- Discrete fields
- Objects

Two different types of boundaries:
- *Crisp* boundaries
- *Fuzzy* boundaries
Fuzzy boundaries (natural)

Different colors of algae
Fuzzy boundaries (man-made)
Fuzzy boundaries (dynamic)
Spatial Data modeling

Geographic phenomena

- **Object – Discrete**
  - Location
  - Shape
  - Size
  - Orientation
  - **Boundaries: Crisp - Fuzzy**

- **Field**: Discrete or Continuous

Computer representations

Visualization
**Computer representation of a continuous field**

**Option 1**: store as many points as possible

**Option 2**: find a symbolic representation

\[(3.0678x^2 + 20.08x - 7.34y)\]

**Spatial autocorrelation**: locations that are close are more likely to have similar values than locations that are far apart.

**Finite number of locations with “Elevation”**

**Interpolation function**
Computer representations of geographic information

- Regular tessellations (mosaic)
  - Is a partition of space into mutually exclusive cells that together make up the complete study area.
  - In a regular tessellation, the cells are the same shape and size.
  - The field attribute assigned to the cell is associated with the entire area occupied by the cell.
- Square, regular tessellations are known under various names: raster, grid
- The size of the area that a grid cell represents is called the raster’s resolution
Most common regular tessellation types

Square cells  
Hexagonal cells  
Triangular cells
Regular tessellation

- The field value of a cell can be interpreted as one for the complete tessellation cell (in which case the cell is discrete).
- Some convention is needed to state which value prevails on cell boundaries.

• Lower and left boundaries belong to the cell.
To improve on the continuity issue, we can do two things:

- Make the cell size smaller, so as to make the ‘continuity gaps’ between the cells smaller
- Assume that a cell value only represents the value for one specific location in the cell, and to provide a good interpolation function for all other locations that has the continuity characteristic.
Regular tessellation

- Make the cell size smaller to make the ‘continuity gaps’ smaller.
Regular Tessellation

- Assume that a cell value only represents the value for one specific location in the cell, and to provide a good interpolation function for all other locations.

Cell centroid

Left lower corner
Regular Tessellation

- Assume that a cell value only represents the value for one specific location in the cell, and to provide a good interpolation function for all other locations.

- This technique is considered too costly for most rasters.
Advantages - Disadvantages of regular tessellation

Advantages:

- We know how they partition space. This leads to fast algorithms.

Disadvantages:

- They do not adapt to the spatial phenomenon we want to represent.
- Cell boundaries are both artificial and fixed: they may or may not coincide with the boundaries of the phenomenon of interest.
Irregular tessellation

Elevation 100 meters

All values are 100
Irregular tessellations

- Again, these are partitions of space into mutually exclusive cells, but now the cells vary in size and shape, allowing them to adapt to the spatial phenomena they represent.
- One example: region quadtree.
Region quadtree

- It splits up the area into four quadrants. This procedure stops when all the cells in a quadrant have the same field value.
Region quadtree
Region quadtree

- When a conglomerate of cells has the same values, they are represented together in the quadtree.
- Quadtrees provide nested tessellation:
  - Quadrants are only split if they have two or more values.
- An interesting characteristic of quadtrees:
  - Quick computation is possible because square nodes at the same level represent equal area sizes.
Spatial Data modeling

Geographic phenomena

- Object – Discrete
  - Location
  - Shape
  - Size
  - Orientation
- Boundaries: Crisp - Fuzzy

- Field
  - Discrete or Continuous

Computer representations

Tessellation – Raster

- regular tessellation
- Irregular tessellation

Visualization
Vector representations

In vector representations, an attempt is made to associate georeferences with the geographic phenomena explicitly.

Coordinate pair from some geographic space also known as a vector

Note: rasters do not explicitly store georeferences of the phenomena they represent.
Georeferences

Raster

<table>
<thead>
<tr>
<th>18</th>
<th>16</th>
<th>14</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>15</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

georeference

Vector

y1
y2
y3

x1 x2 x3
Triangulated Irregular Network (TIN)

- It is built from a set of locations for which we have a measurement.
- The locations can be arbitrarily scattered in space (not a nice regular grid)
- Observe that in three-dimensional space, three points uniquely determine a plane, as long as they are not positioned on the same line.
- If we restrict the use of a plane to the area between its three anchor points, we obtain a triangular tessellation of the complete study area.
TIN

- The circle through its three anchor points does not contain any other anchor points.

- Triangles are as equal-sided as they can be.

**Delaunay triangulation**

**Triangulation with many stretched triangles**
TIN

- A plane fitted through the anchor points has a fixed aspect and gradient.
  - Slope consists of two parts gradient and aspect.
  - Slope can be used to compute an approximation of elevation of other locations.
TIN
Does TIN belong to vector?

- Yes, each anchor point has a stored georeference.
- No, as the chosen triangulation provides a tiling of the entire study space. However, the cells of the tiling do not have an associated stored value as is typical of tessellations.
Points are defined as single coordinate pairs \((x,y)\) when we work in 2D or coordinate triplets \((x,y,z)\) when we work in 3D.
Line representations

- Used to represent one-dimensional objects (roads, railroads, canals, rivers...)
- Line is defined by 2 end nodes and 0-n internal nodes.
- An internal node or vertex is like a point that only serves to define the line.

![Diagram showing line representation with begin and end nodes and internal nodes connected by arcs.]

Page Number: 51
Line representation

- Network is a collection of connected lines
Area representations

- When area objects are stored using a vector approach, the usual technique is to apply a *boundary model*. This means that each area feature is represented by some arc/node structure that determines a polygon as the area’s boundary.
Geographic phenomena

- Object – Discrete
  - Location
  - Shape
  - Size
  - Orientation
  - Boundaries: Crisp - Fuzzy

- Field – Discrete or Continuous

Computer representations

Vector:
- point
- line (network)
- Area

TIN

Tessellation – Raster
- regular tessellation
- Irregular tessellation

Visualization
Area representation

- A simple but naïve representation of area features would be to list for each polygon simply the list of lines that describe its boundary.

Why is this not working?

- Redundancy

- If we want to find neighboring polygons, we have to do time-consuming analysis (topology)
Area representation

Redundancy: data duplication
Topology and spatial relationships

- Topology deals with spatial properties that do not change under a transformation
  - E (polygon, area) is still inside D (another polygon or area)
  - The neighborhood relationships between the areas remain, and the boundary lines have the same start and end points.
  - The areas are still bounded by the same boundary lines, only the shapes and lengths of their perimetry have changed.
## Topology and spatial relationships

<table>
<thead>
<tr>
<th>Topological</th>
<th>Non-topological</th>
</tr>
</thead>
<tbody>
<tr>
<td>A point is at an end-point of an arc</td>
<td>Distance between two points</td>
</tr>
<tr>
<td>An arc is a simple arc (the arc does not cross over itself)</td>
<td>Bearing of one point from another point</td>
</tr>
<tr>
<td>A point is on the boundary of a region</td>
<td>Length of an arc</td>
</tr>
<tr>
<td>A point is in the interior of a region</td>
<td>Perimeter of an area</td>
</tr>
<tr>
<td>A point is in the exterior of a region</td>
<td>Area of an area</td>
</tr>
<tr>
<td>An area is open (excludes all its boundary)</td>
<td></td>
</tr>
<tr>
<td>An area is closed (includes all its boundary)</td>
<td></td>
</tr>
<tr>
<td>An area is simple (has no holes)</td>
<td></td>
</tr>
<tr>
<td>An area is connected (given any two points in the area, it is possible to</td>
<td></td>
</tr>
<tr>
<td>follow a path from one point to the other such that the path is entirely</td>
<td></td>
</tr>
<tr>
<td>within the area)</td>
<td></td>
</tr>
<tr>
<td>An area is within a loop</td>
<td></td>
</tr>
</tbody>
</table>
Mathematical properties of the geometric space

- The space is a three-dimensional *Euclidean space* where for every point we can determine its three-dimensional coordinates as a triple \((x, y, z)\) of real numbers.
- The space is a *metric space*, which means that we can always compute the distance between two points according to a given distance function. Such a function is also known as a *metric*.
- The space is a *topological space*. For every point in the space we can find a neighbourhood around it that fully belongs to that space as well.
- *Interior* and *boundary* are properties of spatial features that remain invariant under topological mappings. When transformed, the interior and the boundary of a feature remains unbroken and intact.
Mathematical properties of the geometric space

- Euclidean Space
- Metric Space
- Topological Space

Neighbourhood

Boundary is unbroken

1.56 km
Simplices

A tool to obtain some ‘topological sensitivity’

Simple building blocks with which more complicated representations can be constructed.
Topology of two dimensions

- We use the topological properties of interior and boundary to define relationships between spatial features.

Region A → Boundary A
   Interior A

Region B → Boundary B
   Interior B

When the two regions meet this is defined as the Boundary of A intersects the boundary of B. The interiors do not intersect, and the boundary of one does not intersect the interior of the other.
The mathematical definition of *meets*.

A meets B = \( \text{interior} (A) \cap \text{interior} (B) = \emptyset \) \( \land \)

\( \text{interior} (A) \cap \text{boundary} (B) = \emptyset \) \( \land \)

\( \text{boundary} (A) \cap \text{interior} (B) = \emptyset \) \( \land \)

\( \text{boundary} (A) \cap \text{boundary} (B) \neq \emptyset \)

There are 7 other spatial relationships that can be defined in a similar way.
Spatial Relationships

- disjoint
- meet
- equal
- inside
- covered by
- contains
- covers
- overlap
1. Every 1-simplex (‘arc’) must be bounded by two 0-simplices (‘nodes’, namely its begin and end node)

2. Every 1-simplex borders two 2-simplices (‘polygons’, namely its ‘left’ and ‘right’ polygons)

3. Every 2-simplex has a closed boundary consisting of an alternating (and cyclic) sequence of 0- and 1-simplices.

4. Around every 0-simplex exists an alternating (and cyclic) sequence of 1- and 2-simplices.

5. 1-simplices only intersect at their (bounding) nodes.
2½-D

- Rules of two-dimensional topology apply
- Nodes have a z-value
- A node with an x- and y- coordinate can have only one z-value.

Cliffs of Dover

Ground surface

Tree
$2^{\frac{1}{2}}$-D
Map scale can be defined as the ratio between distance on a paper map and distance of the same stretch in the terrain.

1:50,000 → 1 cm on map = 50,000 cm in reality
1 cm on map = 500 meters in reality

Large-scale → much detail of a small area
Small-scale → few detail (world map)
Scale and resolution

- The size of the area a pixel cell represents is called *resolution*.
Representation of geographic fields

- Tessellation
- TIN
- Vector representation
Representation of geographic objects

Vector

Raster
Spatial Data modeling

Geographic phenomena
- **Object** – Discrete
  - ✓ Location
  - ✓ Shape
  - ✓ Size
  - ✓ Orientation
  - ❏ Boundaries: Crisp - Fuzzy
- **Field** – Discrete or Continuous

Computer representations
- **Vector**:
  - point
  - line (network)
  - Area
  - TIN

Visualization
- **Tessellation** – Raster
  - regular tessellation
  - Irregular tessellation
Organizing one’s spatial data

- Spatial data layer = field or collection of objects
- (In ArcGIS a data layer is called a layer)
- Data is organized by kind (buildings, rivers, lakes)
- A data layer contains spatial data and attributes
- Data layers can be overlaid with each other so as to study combinations of geographic phenomena (spatial correlation)
The temporal dimension

Besides having geometric (where), thematic (what) and topological properties geographic phenomena also have temporal characteristics (when)

This area of work is called *change detection*.

Change detection addresses issues like:

- Where and when did change take place?
- What kind of change occurred?
- With what speed did change occur?
- What else can be understood about the pattern of change?
Properties of the time dimension

- Time density, *discrete or continuous scale*
- Dimensions of time, *valid time (world time) or transition time (database time)*
- Time order, *linear time, branching time, cyclic time (seasons)*
- Measure of time, *chronon (the shortest non-decomposable unit of time) granularity (precision of time value)*
- Time reference, *absolute (fixed time) or relative (implied time)*
## Properties of the time dimension

### Time Density

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Discrete time**
- **Continuous time**

### Dimensions of time

- **Valid time**
- **Transaction time**

### Time order

- **Past**: linear
- **Future**: branching
- **Cyclic**: s, s, a, w

---

Page Number: 63-64
Properties of the time dimension

Measures of time
- chronon
- granularity

Time reference
- yesterday
- tomorrow
- Oct 19, 2001
- Oct 21, 2001

Relative time
Absolute time