CE-317 GIS/RS Application to Civil Engineering
Spring 2011

•Engr. Faisal ur Rehman

•Lecture 02: Data Models in GIS

Lecture Outline

- Definition
- Vector Data Model
- Raster Data Model
- Advance Data Model

Data Model Definition

- You have to show the computer how to think like you do as a map reader.
- Formulate a conceptual model (a picture in your mind) of how you plan to tell the computer all the information that you glean from a map.
- *Data Model is the representation of Spatial Information that computer can read and understand.*

Data Model Definition

- GIS depicts the real world through models involving geometry, attributes, relations, and data quality.
- Spatial information is presented in two ways: as vector data in the form of points, lines, and areas (polygons); or as grid data in the form of uniform, systematically organized cells (raster data).

Vector Data Model

- The basis of the vector model is the assumption that the real world can be divided into clearly defined elements consisting points, lines, or areas.
- In principle, every point on a map and every point in the terrain it represents is uniquely located using two or three numbers in a coordinate system, such as in the northing, easting, and elevation.
- Every object is a point or mathematically defined function line like curve of the road.

Coding Digital Data For Map Production

- Roads, contour lines, property boundaries, and other data indicated by lines are usually shown in lines of various widths and colors
- Symbols designate the locations of mosques, airports, and other buildings and facilities.
- In thematic coding, data are divided into single-topic groups, such as all property boundaries.

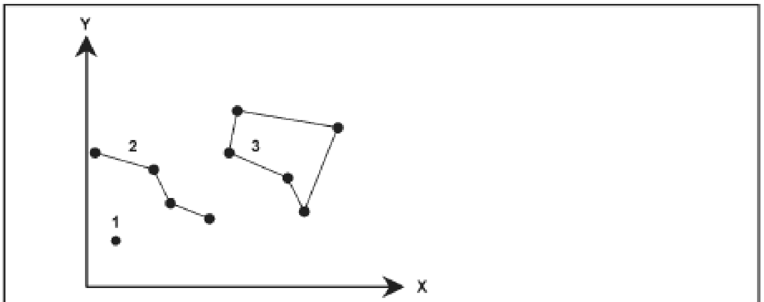
Coding Digital Data For Map Production

- Information on symbol types, line widths, colors, and so on, may be appended to each thematic code, and various combinations of themes may be drawn.
- Represented on a single Coordinate System

Coding Digital Data For GIS

- Point objects may easily be realized in a database because a given number of attributes and coordinates is associated with each point.
- Line and polygon objects are more difficult to realize in a database because of the variation in the number of points composing them.
- Object spatial information and object attributes are often stored in different databases to ease the manipulation of lines and areas, but in some systems they are stored together.

Coding Digital Data For Gis



Geometry	ID	Attributes		Coordinates
Point	1	A	B	(Single Point) 2,3
Line	2	C	D	(String) 1,7 3,6 5,5 7,4
Area	3	E	F	(Closed Polygon) 8,7 11,6 12,5 14,9 9,15

Figure 7.3: Each object is assigned attributes and coordinates. The number of coordinates for lines and polygons depends on the length or circumference of the object.

Storing Points and Lines

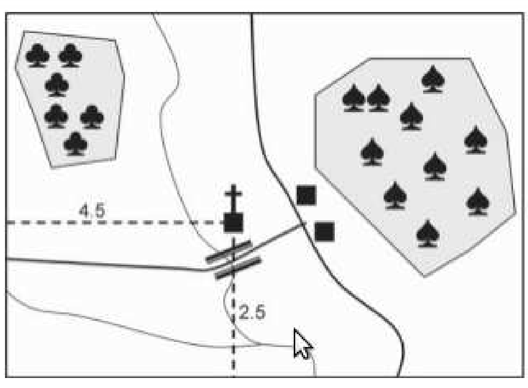


Figure 7.5: Imaginary map.

Table 7.1: Coordinate data for points.

Building 1	4.5	2.5
Building 2	5.8	2.9
Building 3	6.0	2.2

Storing Points and Lines

- we can do three things:
  1. Store the locational data from the map on a computer.
  2. Use this information to reproduce the map.
  3. Make simple calculations from the data.

Attributes of Points

Table 7.2: Attribute data for points.

	X Coordinate	Y Coordinate	Feature Code	Building Material	Name
Building 1	4.5	2.5	Temple	Stone	Krishna's
Building 2	5.8	2.9	House	Brick	Shaan's
Building 3	6.0	2.2	House	Stone	Sameer's

Attributes of Lines

- Again, the basic table of X Y coordinates for a line is a relatively simple file which could be stored in a spreadsheet or a database package.
- However, we also want to store attributes for this line, and this is when we begin to run into difficulties.

Attributes of Lines

Table 7.5: Adding locational information to attribute table for lines. Rows containing data for Khair Road are shaded light grey.

Name	Surface Quality	Peak Traffic Flow	X Coordinate	Y Coordinate
Khair Road	Fair	600	4.5	10.0
			4.5	5.7
			5.5	2.5
			6.5	0.3
			6.8	0.0
Anupshahar Road	Good	1000	0.0	1.5
			3.6	1.5
			5.5	2.5

Table 7.6: Alternative method of adding locational information to attribute table for lines.

Name	Surface Quality	Peak Traffic Flow	X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4	X_5	Y_5
Khair Road	Fair	600	4.5	10.0	4.5	5.7	5.5	2.5	6.5	0.3	6.8	0.0
Anupshahar Road	Good	1000	0.0	1.5	3.6	1.5	5.5	2.5				

Identifier or Unique Key

- Spatially defined objects without attributes need no identifiers
- But they are required for all objects that are listed in attribute tables

Identifier or Unique Key

Table 7.7: Locational data for buildings.

Building – ID	X Coordinate	Y Coordinate
1	4.5	2.2
2	5.8	2.9
3	6.0	2.2

Table 7.8: Attribute data for buildings.

Building – ID	Feature Code	Building Material	Name
1	Temple	Stone	Krishna's
2	House	Brick	Shaan's
3	House	Stone	Sameer's

Identifier or Unique Key

Table 7.9: Modified attribute table for roads.

Road – ID	Name	Surface Quality	Peak Traffic Flow
1	Khair Road	Fair	600
2	Anupshahar Road	Good	1000

Storing Area Object

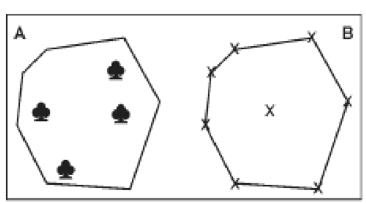


Figure 7.7: Storage of area feature.

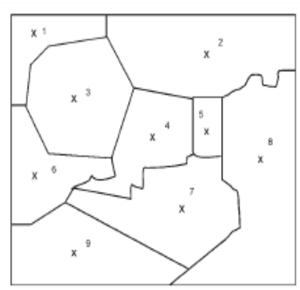


Figure 7.8: Land use map – an example of multiple areas.

Table 7.10: Attribute table for land use map.

Polygon ID	Land Use
1	Vacant Land
2	Vacant Land
3	Forest
4	Urban
5	Agriculture
6	Vacant Land
7	Vacant Land
8	Forest
9	Agriculture

Problems in Storing Areas

- The use of centroids means that to store a single area in our GIS, we actually need to store two things – the line defining the boundary and the point defining the centroid.
- Because the two lines do not coincide, there are small areas of overlap, and small gaps between the two areas.
- These mismatches areas are called sliver polygons, because they are usually very small and thin.
- Some times GIS analysis requires merging of polygon areas
  - This operation is called a polygon dissolve, and is quite common in GIS analysis.
  - However, it is difficult to do with the simple method of storing the area boundaries.

Problems in Storing Areas

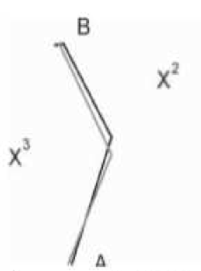


Figure 7.9: Sliver polygons as a result of digitizing the same line twice.

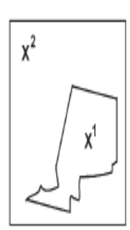


Figure 7.10: Map of urban and non-urban areas created using a polygon dissolve operation.

Spaghetti Data Model

- The simplest of vector data models is the spaghetti model
- A one-to-one translation of graphics into the computer without regard for spatial relationships
- Points are represented by search-able ID

Spaghetti Data Model

- It doesn’t work — meaning it doesn’t store information correctly
- Other operations vital in GIS, such as overlaying and network analysis, are intractable.
- Furthermore, unlinked data require an inordinate amount of storage memory

Spaghetti Data Model

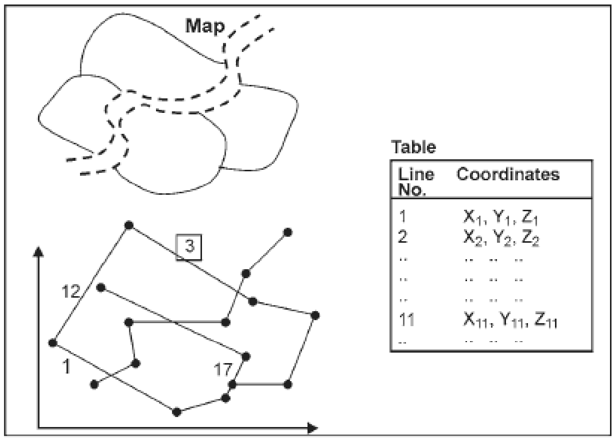


Figure 7.12: Spaghetti data is a term used to describe digital map data with crossing lines, loose ends, double digitization of common boundaries between adjacent polygons, etc. These data lie in a pile, just like spaghetti. Several line segments are found at odd places in the data file.

ESRI Shapefile (shp)

- Shapefile stores the geometry of each feature as a shape that contains the coordinates and links to the attributes.
- To keep track of the shapes and search them by type, it also has an index file that has a file extension .shx.
- Finally, the .dbf (or database) file contains all the attributes that describe the features.

Storing Raster Data Structures

- Using a raster GIS we could store a set of spatial data in the form of a grid of pixels.
- Each pixel will hold a value which relates to some feature of interest at that point in space.

Storing Raster Data Structures

- These values are normally one of three possible types:

I. Binary – A value which indicates the presence or absence of a feature of interest.

- For example, in a layer representing roads, we might use 1 for pixels which contained part of road, and 0 for pixels which did not.

II. Enumeration – A value from some classification.

- For example, a layer representing soils might contain codes representing the different soil types– 1 for alluvial, 2 for red soil etc.

Storing Raster Data Structures

III. Numerical – An integer or floating point number recording the value of a geographical phenomenon.

- In the soil example, we might have measurements of soil moisture content.
- A common example of this kind of raster layer is when the values represent the height of the land surface, in which case the layer is often referred to as a Digital Elevation Model (DEM).

Storing Raster Data Structures

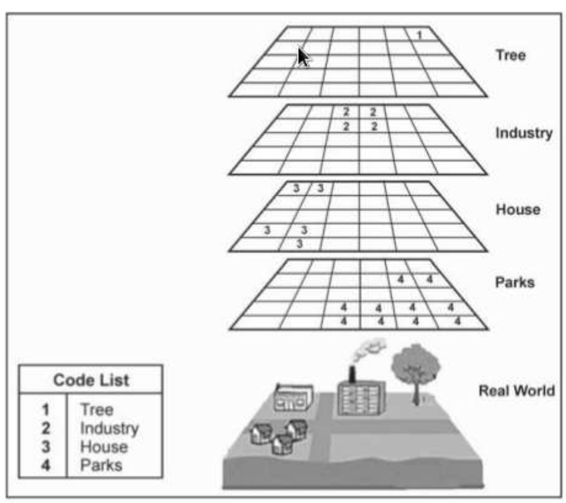


Figure 7.24: Only one attribute value is assigned to each cell. Objects that have several attributes are therefore represented with a number of raster layers, one for each attribute.

Saving Space

- Use of smallest unit of memory for saving.
- The Run Length Encoding And Quad-Trees

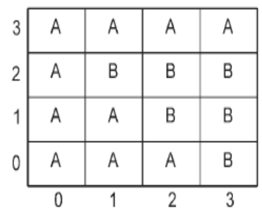


Figure 7.27: An example of simple raster array.

Table 7.21: Storage of a run length encoded layer in computer memory.

Address	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Value	3	A	1	B	2	A	2	B	1	A	3	B	4	A		

Saving Space

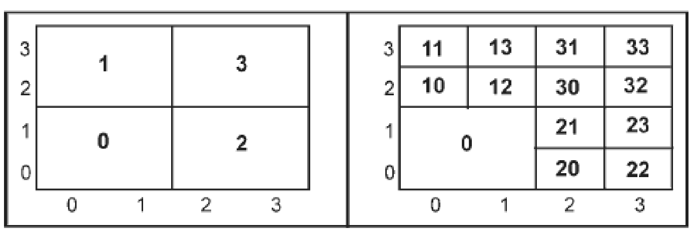


Figure 7.28: Quadtree subdivision of layer shown in figure 7.28.

Table 7.22: Storage of quadtree in memory.

Address	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Quadtree Address	0	1	2	3	10	11	12	13	20	21	22	23	30	31	32	33
Value	A	4	8	12	A	A	B	A	A	B	B	B	B	A	B	A

Saving Space

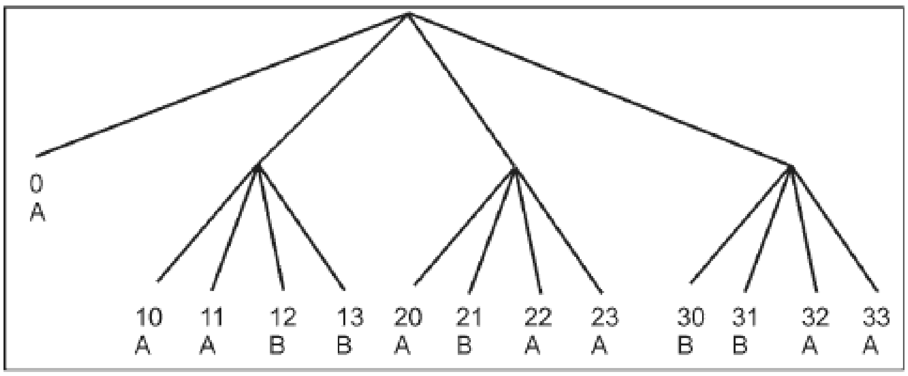


Figure 7.29: Graphical representation of quadtree.

Automatic Conversion Between Vector And Raster Models

- Raster data are converted to vector data through vectorization. The reverse process, which is just as common, is rasterization.
- In vectorization, areas containing the same cell values are converted to polygons with attribute values equivalent to the pre -conversion cell values.
- In the reverse process of converting polygons to cells, each cell falling within a polygon is assigned a value equal to the polygon attribute value

Rasterization & Vectorization

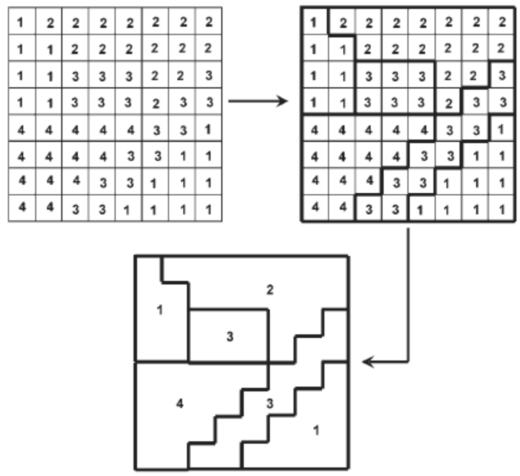


Figure 7.30: Conversion of raster data to vector data; first, each raster cell is assigned an attribute value; secondly, boundaries are set up between different attribute classes and finally, polygons are created by storing X and Y coordinates.

Rasterization & Vectorization

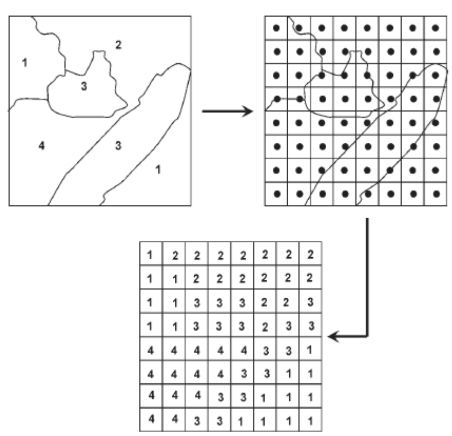


Figure 7.31: Conversion of vector data to raster data; first, polygons are coded; secondly, a grid with the right cell size overlays the polygons; here, the polygons that contain the centre of the individual cells are Identified; finally each cell is assigned the attribute code of the polygon to which it belongs.

Attributes

- Attribute are compressed, linked to geometry/pixel and stored in database.
- Final synthesized map is called Integrated Terrain unit (ITU)

Attributes

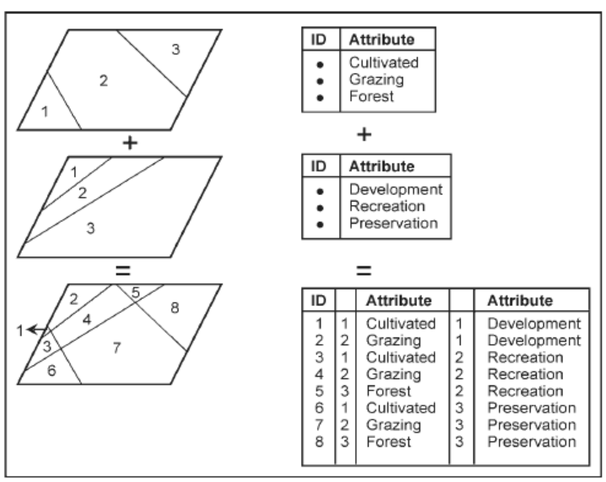


Figure 7.32: Attribute data can be made comparable by superimposing geometry from dissimilar geographical units to get the integrated data.

Advance Data Model

Storing Area Boundaries: The Topological Approach

- Topology is used to store features
- If we look at the land use map in Figure 7.13, we can see that each area boundary is made up of a series of line sections, which meet at junctions

Storing Area Boundaries: The Topological Approach

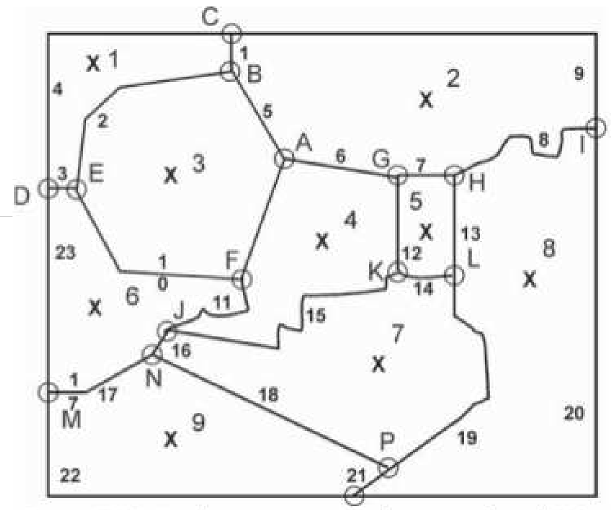


Figure 7.13: Land use map as stored in a topological GIS.

Advantages of Topology

- Without identifier a polygon cannot be stored so no chance of error
- Link + Point solves the problems of storing areas and lines
- It is easy to perform mathematical calculation line dissolving of polygon

7 Bridge Problem

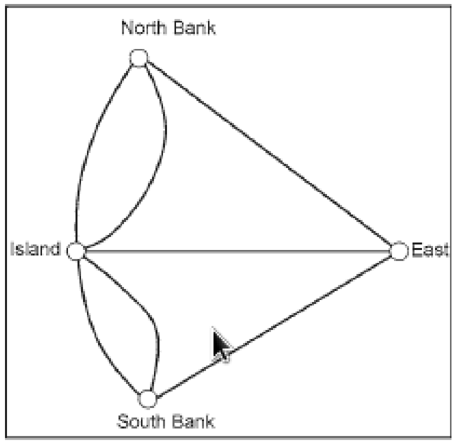


Figure 7.14: Sketch of the Konigsberg bridge problem.

DIME

- The computer to construct and test the graph around any block or junction.
- The system which was devised was called DIME – (Dual Independent Map Encoding)
- The DIME model stores the topology by creating explicit From and To nodes that define each link for each map
-

DIME

- DIME recognizes the three basic components of topology (contiguity, connectivity, and area)
- The primary purpose of this model is to allow census data users to link the tabulated census information to census geographic units (such as streets, blocks, districts, and so on)
- There is an automatic check that data will form a close loop.
- The major drawback to the DIME file is that its geography isn't very realistic. (curved street)

DIME

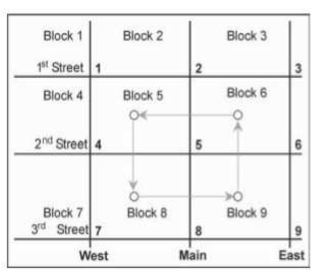


Figure 7.15: Fictitious city blocks illustrating Poincarés dual graph model of maps.

Table 7.12: DIME data structure for fictious city map.

Segment	From	To	Block Left	Block Right	Street Name
1	1	2	2	5	1st
2	2	3	3	6	1st
3	4	5	5	8	2nd
4	5	6	6	9	2nd
5	7	8	8	11	3rd
6	8	9	9	12	3rd
7	4	1	4	5	West
8	7	4	7	8	West
9	5	2	5	6	Main
10	8	5	8	9	Main

Table 7.13: Records from DIME file related to block 5. All records are modified so that block 5 is on right hand side of the street.

Segment	From	To	Block Left	Block Right	Street Name
1	1	2	2	5	1st
3	5	4	8	5	2nd
7	4	1	4	5	West
9	2	5	6	5	Main

TIGER

- Topologically Integrated Geographical Encoding and Referencing
- **Easy retrieval:** The TIGER model directly addresses points, lines, and polygons, so you can more easily retrieve a census block by retrieving the block number, rather than having to find the links first .
- **Improved geography:** TIGER files show the actual shape of Lake Drive. This feature makes TIGER files much more compatible with non-census-related research and other data files that are more geographically accurate.

ESRI Coverage Model

- Each theme covers the study area's geographic extent, and a series of these themes represents a more comprehensive view of the content of the study area.

ESRI Coverage Model

- Entities: Primary features such as points, arcs (the word ESRI uses for lines), and polygons
- Complementary information: Secondary features such as
  - Tics: Points that show the input device where the map coordinates are
  - Links: Computer code that ties the graphics to the descriptions
  - Annotations: The text of the descriptive information

Object-Oriented Representations

- OOP focuses on objects where an object is a set of computer code that can be copied from place to place in a computer program.
- In GIS, an object represents a type of geographic feature that you can move around and whose properties follow it (are inherited) no matter where you place it.
- Much of the power of OOP data models results from the objects' shared properties

Advanced Geographic Representations

- **Attribute tables** from typical **RDBMS**
- **Geographic features** that you might find in land-use maps.
- **Satellite and other digital imagery** such as the files you might get from the LANDSAT satellite.
- **Surface models** you could obtain from the USGS Digital Elevation Models (DEMs).
- **Survey data** from scientific vegetation samples, or telephone surveys, for example.

Advanced Geographic Representations

- This is known as geodatabase model
- Like the non-object-oriented coverage model, it has tables of data that include objects (rows in the database table) and features (objects that have explicitly defined geometry).
- It’s a completely topological model, like its predecessors.
- OO analogy makes it powerful and error free.

Advance Data Model

- Surface Representation
  - Digital terrain model (DTM) or a digital elevation model (DEM).
- Grid Model
- TIN Model (triangulated irregular network)

DEM

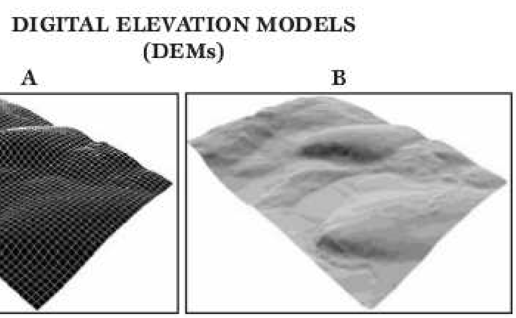


Figure 8.1: A DEM is an essential layer in the representation or analysis of any area with variable terrain.

Raster Grid

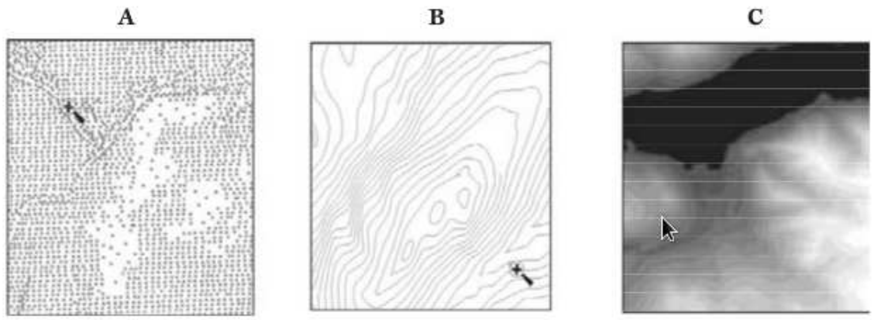
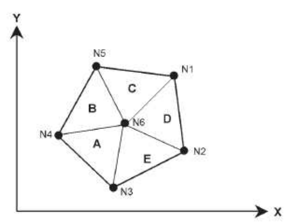


Figure 8.2: Elevation data are acquired through:

a. Mass points (lattices)
b. Contour lines
c. GRIDs (interpolated from points or lines; or created currently from digital imagery)

TIN



Triangle Table and Node

Triangle	Adjacent	Nodes
A	B, E	N3, N4, N6
B	A, C	N4, N5, N6
C	B, D	N1, N5, N6
D	C, E	N1, N2, N6
E	D, A	N2, N3, N6

Coordinate Table

Node	Coordinates
N1	X1, Y1, Z1
N2	X2, Y2, Z2
N3	X3, Y3, Z3
N4	X4, Y4, Z4
N5	X5, Y5, Z5
N6	X6, Y6, Z6

Figure 8.4: TIN model: the triangles are stored in a topological structure.

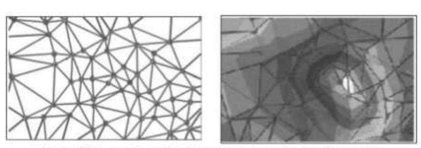


Figure 8.5: An example of topological structure in TIN model.

Isolines / Contours

- Isolines – continuous lines connecting points of the same elevation may represent terrain in much the same way as contour lines depict terrain on conventional maps

Isolines / Contours

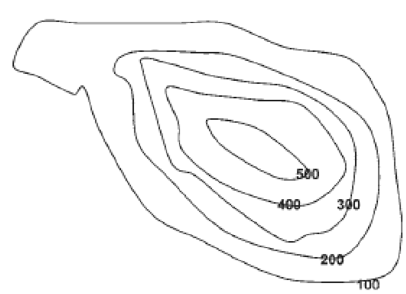


Figure 8.6: Lines that link points with the same terrain height are used to represent terrain surface, which corresponds to traditional elevation contours. However, this structure is poorly suited to the calculation of elevation values for new points.

Representation of Time

- Possible practical solutions will therefore be:
  - The attributes of the objects will be changed.
  - The geometry of objects is changed.

Other Models

- Models For Moveable Objects
- Model For Movement Over Surfaces
- Network Model
  - The network model comprises road systems, power grids, water supply, sewerage systems, and the like, all of which transport movable resources.
- Combination Of Models

Common network operations

- **Pathfinding** is the process to find the shortest, least cost, or most efficient path or tour on a network.
- **Tracing** is the process to determine a connected portion of a network that are either flow from this connected portion of the network to a given node or flow from a given node to this connected portion of the network.
- **Allocation** is the process to assign portions of a network to a location (e.g., a center) based on some given criteria.

Common Network Applications

- **Geocoding** is the process for building a relationship between locational data in a database and street address data that are normally in a tabular format
- **Location-allocation** is the process of determining the optimal locations for a given number of facilities based on some criteria and simultaneously assigning the population to the facilities.
- **Business logistics:** The optimization of vehicle routing and delivery scheduling is vital for many business operations.

Common Network Applications

- **Spatial interaction** and gravity modelling: The interaction between different locations in geographic space and the mathematical modelling of the interaction
- **Dynamic segmentation:** Dynamic segmentation is a particular network model used to represent, analyze, query, and display linear features.
  - Dynamic segmentation is commonly used to model linear features such as highways, river networks, power lines, city streets, and telephone lines.

Q & A