People’s Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

Hassiba Benbouali University of Chlef
Faculty of Exact Sciences and Computer Science
Département of Computer Science

COURS PACK

Geographic information system (GIS)

FOR :

Master’s Degree M2 : Advanced information system Engineering

by :

GHERBAOUI Radhwane

Last updated : january 2026

Preamble

This handout is the result of years of teaching the module « Geographic Information Systems »at Hassiba Benbouali University of Chlef (UHBC). This module is delivered to second-year Master's students in computer science, specifically those in the « ISIA »track. In Geographic Information Systems, several disciplines are involved such as artificial intelligence, imaging, and simulation, etc. Therefore, in this handout, we have emphasized the structuring of geographic data and certain processing aspects that are rarely addressed in other fields.

This handout is mainly intended for second-year Master's students and computer engineers in order to enable them to design and develop GIS prototypes. The objective is to familiarize future computer science specialists with the basic concepts for the modeling and implementation of GIS.

For continuous improvement of this handout, please feel free to contact me for any suggestions, criticism, remarks, or corrections at the following email : r.gharboui@univ-chlef.dz
The last update of this handout was made in March 2023.

Table des matières

Préambule **1**

General Introduction **6**

Chapitre 1
Presentation of Geographic Information Systems

1.1 Introduction 9
1.2 Definitions 9
1.3 Multidisciplinarity of GIS 11
1.4 General Objectives 12
1.5 Components of a GIS 13
1.6 Acquisition of Geographic Data 14
1.7 Conclusion 14

Chapitre 2
Spatial Data Structures

2.1 Introduction 18
2.2 Geographic database 18
2.2.1 Advantages of using databases 19
2.2.2 Database models 20
2.3 Modes of representation of geographic data 22
2.4 Data structures in vector mode 23
2.4.1 Non-topological structures 24
2.4.2 Topological structures 25
2.5 Structure specific to elevation information 30
2.5.1 Irregular triangulated network structure : RTI (Triangulated Irregular Network : TIN) 31

2.5.2 Topological structure for contour lines 31
2.6 Raster mode 35
2.7 Conclusion 38
2.8 The Route Table 47
2.9 The Building Table 47
2.10 The Wilaya Table 47
2.11 Attribute Queries 48
2.12 Spatial Queries 48
2.13 Conclusion 49
2.13.1 Attribute queries 52
2.13.2 Spatial queries 52
2.14 Conclusion 53

General Conclusion **56**

Bibliographie **57**

Table des figures

1.1 Example of business functionalities. 10
1.2 Structure of a GIS. 11
1.3 Multidisciplinarity of GIS. 11

2.1 The hierarchical model of a country. 20
2.2 A map M composed of two regions. 21
2.3 The hierarchical model of map M. 21
2.4 The relational model of map M. 22
2.5 Classes of the object-oriented model of map M. 22
2.6 Example of formalization of a map M_1 according to a model inspired by graph theory. 24
2.7 Sequential structure of map M_1. 25
2.8 Sequential structure with point dictionary of map M_1. 26
2.9 Nodes of map M_1 according to DIME structure. (a) map M_1 ; (b) node file. 27
2.10 Edges of map M_1 according to DIME structure. (a) map M_1 ; (b) edge file. 28
2.11 Nodes of map M_1 according to POLYVRT structure. (a) map M_1 ; (b) node file. 29
2.12 Intermediate points file and zones file of POLYVRT structure for map M_1. (a) intermediate points file ; (b) zones file. 29
2.13 Chain file of POLYVRT structure for map M_1. 30
2.14 Construction of triangles from points. 31
2.15 Irregular triangulated network structure. 32
2.16 Triangle-based data storage. 32
2.17 Vertex-based data storage. 32
2.18 Projection of points onto the (XY) plane. 33
2.19 Contour lines. 33
2.20 Entity file of the previous map. 34
2.21 Coordinate file. 35
2.22 Lower neighborhood file. 35
2.23 Upper neighborhood file. 36
2.24 Three commonly used grid types : square, triangular, and hexagonal grids. 36
2.25 RLC method (Run Length Coding). Left : bitmap image. Right : RLC format. 37
2.26 Square block method applied to image matrix. 37
2.27 Carte C 39

2.28 Hierarchical model of map C 40
2.29 Network model of map C 40
2.30 Simple sequential structure of map C 41
2.31 Map C 42
2.32 Node file of map C 43
2.33 Edge file of map C 43
2.34 Zone file of map C 43
2.35 Node file of map C 44
2.36 Intermediate point file of map C 44
2.37 Chain file of map C 44
2.38 Contour map C-3 45
2.39 Contour map C-4 45
2.40 Contour map C-2 46
2.41 Three files of the contour structure. (a) Lower neighbors file. (b) Upper neighbors file. (c) Point file. 46
2.42 Building Table 50
2.43 Route Table 50
2.44 District Table 50
2.45 Wilaya Table 50
2.46 Building Table 54
2.47 Road Table 54
2.48 Daira Table 54
2.49 Wilaya Table 54

General Introduction

Geographic Information Systems (GIS), as their name indicates, are information systems dedicated to geographically referenced information. The evolution of information systems has led to the evolution of GIS to the point that they are currently considered advanced systems. As a result, their characteristics and objectives continue to grow increasingly numerous. In the early seventies, the architecture of this system was confined to a processing unit, a storage disk, a display screen, and an acquisition scanner. More recently, this architecture has undergone significant evolution such that storage and processing disks have been replaced by entire servers. Information acquisition means have been developed so that they include data servers, satellites, radars, aircraft, etc. Meanwhile, display tools have gone far beyond a simple screen to be integrated into vehicles (GPS), mobile terminals, electrical network display panels, etc.

The evolution of Geographic Information Systems is not limited only to the hardware aspect ; software tools have also evolved. Entire disciplines have been integrated into GIS such as imaging, geometry, decision support systems, simulation, etc. GIS appear in several domains including institutions such as land registry offices, municipalities, meteorological centers, etc., as well as companies such as business corporations, maritime and road transport companies, vehicle manufacturing plants, etc. The need of institutions and companies for Geographic Information Systems continues to grow ; day after day, these systems are being used in new domains.

GIS are above all information systems. Therefore, the implementation of these systems benefits from standard Database Management Systems (DBMS) and existing models to reduce development costs. Consequently, extensions of DBMS have been adopted to support spatial data. To facilitate GIS design, geographic data is divided into two parts : an attribute part handled by DBMS and a spatial part that has a geographic reference. Thus, other approaches combine DBMS with spatial management systems to further reduce development costs. The increasing emergence of GIS in various domains and disciplines has led researchers to design completely new models to facilitate the modeling and processing of geographically characterized data.

This manual is organized as follows. In Chapter 1, definitions framing Geographic Information Systems are presented. In this chapter, the architecture and objectives for which GIS are designed will be illustrated. Methods for acquiring geographic data will be detailed. In Chapter 2, modes of representation and structuring of geographic data will be explained with examples. Chapter 3 is devoted to the analysis of geographic information.

1

Presentation of Geographic Information Systems

Sommaire

1.1 Introduction 9
1.2 Definitions 9
1.3 Multidisciplinarity of GIS 11
1.4 General Objectives 12
1.5 Components of a GIS 13
1.6 Acquisition of Geographic Data 14
1.7 Conclusion 14

1.1 Introduction

In recent decades, Geographic Information Systems have experienced unprecedented evolution. This evolution is mainly due to the development of computational processing capabilities, communication and information exchange technologies, and spatial data acquisition methods. Indeed, the needs of companies and organizations for spatial processing tools remain the most important factor driving designers to develop such systems. In this chapter, we introduce readers to Geographic Information Systems by presenting definitions from the literature that frame the concept of these systems. A description of the main components and associated functions is provided, as well as the objectives for which these systems are designed.

1.2 Definitions

Several definitions have been proposed in the literature identifying the concept of a "Geographic Information System".

- Definition by Clark (1986) [7] : « *"A system for the development and expansion of the use of maps"* ».
- Definition by Aronoff (1989) [1] : « *"A set of procedures used to store and process geographically referenced information"* ».
- Definition by Burrough (1998) [4] : « *"A set of powerful tools for the acquisition, storage, editing, as well as transformation and analysis of spatial data related to the real world".* »
- Definition by Dider [8] : « *"A set of spatially referenced data structured in such a way as to easily extract useful summaries for decision-making"* ».
- Definition by the Federal Inter-agency Coordinating Committee for Digital Cartography in the United States (1988) : « *A GIS is a computer system composed of hardware, software, and processes designed to allow the collection, management, manipulation, analysis, modeling, and display of spatially referenced data in order to solve complex planning and management problems* ».
- Definition by Konecny [11] : « *A GIS, in a narrow definition, is a computer system for the capture, manipulation, storage, and visualization of digital spatial data. In a broader definition, it is a digital system for the acquisition, management, analysis, modeling, and visualization of spatial data for the purposes of planning, administration, and control of the natural environment and socio-economic applications* »

According to these definitions, we conclude that a Geographic Information System is an information system that performs processing on spatial data. Thus, spatial concepts are reformulated in these definitions through maps, geography, or space. In other words, GIS are extended information systems for processing spatial information. The definitions of Burrough and the U.S. federal committee include the fundamental components that must be part of a GIS architecture, which will be detailed in the following sections.

A GIS can also be defined by answering the questions Where ?, What ?, How ?, When ?, and If, as illustrated in Figure 1.1.

The functions of a GIS can also be summarized by what is called the five « A ».

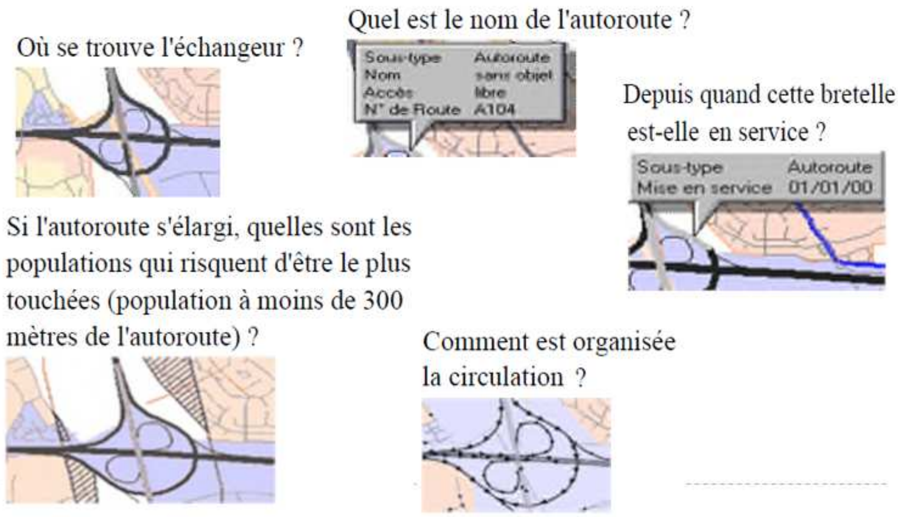


FIG. 1.1 – Example of business functionalities.

1. Abstraction : Information modeling
2. Acquisition : retrieval of existing information.
3. Archiving : storing data in a way that allows easy retrieval and querying.
4. Analysis : answering queries. Analysis may involve decision support or artificial intelligence subsystems. The analysis subsystem affects the quality of the GIS ;
5. Display : graphical output ; maps and attribute data are clearly displayed so that potential users can easily extract knowledge about the targeted areas or objects.

These five functions effectively represent the main modules of a GIS structure (see 1.2).

To better understand GIS, the nature of data known as geographic information (GI), which is processed by GIS, must be specified.

GI is spatially referenced data providing the following information :

1. place name,
2. geographic coordinates,
3. postal address or other.

In general, GI can be shared among multiple clients and servers, duplicated, enriched, and linked with other GI. Geographic information undergoes all the functionalities of the « five A »of a GIS. It is represented by the following two parts.

1. By its nature and attributes (characteristics).
2. By its location on Earth (or another celestial body !).

The first part of GI is known as "attribute information", frequently processed by database management systems. The second part represents the spatial component of GI.

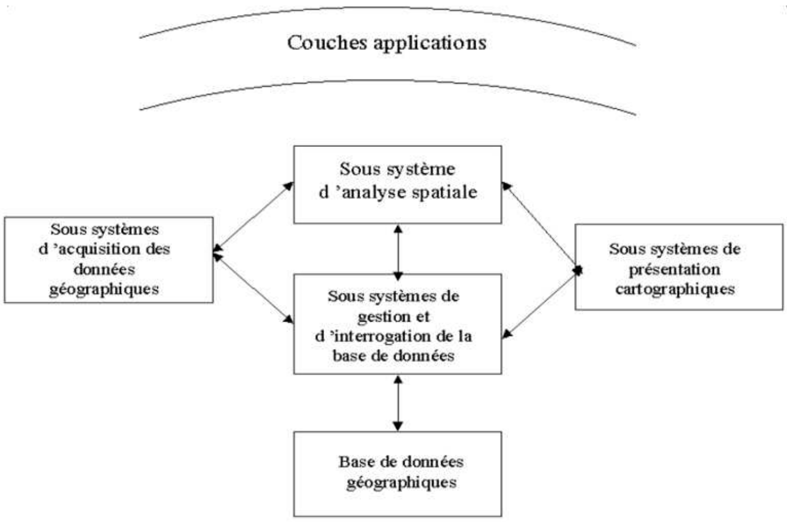


FIG. 1.2 – Structure of a GIS.

1.3 Multidisciplinarity of GIS

A GIS can be viewed as a tool for analyzing information in the spatial domain. It allows the simultaneous analysis of multiple spatial parameters on multi-source and multimedia datasets. When there are several mono-disciplinary maps over a well-defined study area, combining multidisciplinary information becomes essential for comprehensive analysis (see Fig.1.3). This operation can be performed in two ways : manually and computer-assisted [3].

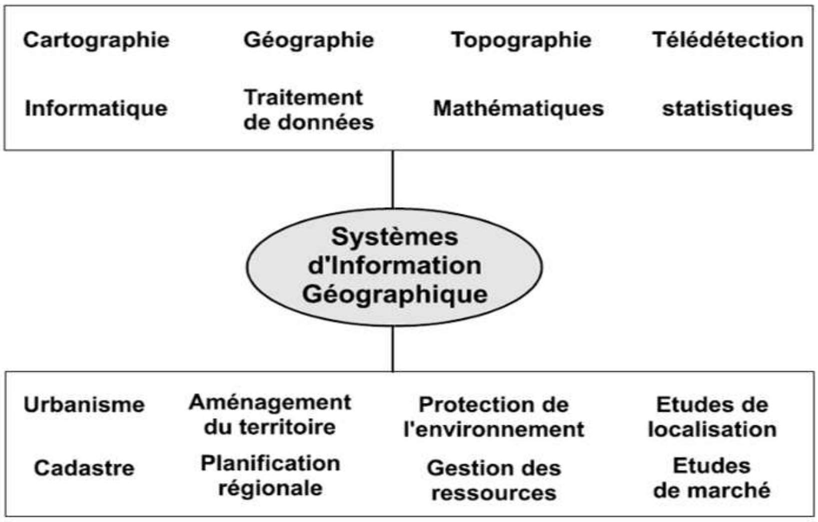


FIG. 1.3 – Multidisciplinarity of GIS.

1. Manual method : this is a classical procedure introduced by Mc Harg (1969). It is

carried out simply by overlaying transparent copies of maps on a light table. The disadvantages of this method are mainly :
– The limited number of maps that can be superimposed simultaneously for visual analysis.
– The difficulty of obtaining all transparent copies at the same scale.

2. Computer-assisted method : data are stored in digital form. Programs are developed for analysis operations, allowing what Mc Harg achieved with transparent overlays to be performed in a faster and reproducible manner.

1.4 General Objectives

The objectives for which GIS are designed are diverse and involve several research axes.

Digital capture and storage of plans and maps : The primary objective of GIS is the digital storage of geographic data, whether two- or three-dimensional. However, the difference between a system that stores objects with both graphical and descriptive information and a system that only stores drawings without semantic content is significant.

Information structuring : Like any database management system, a GIS requires modeling of the real world and structuring of information. This structuring is often difficult because it deals with objects that may have multiple forms, both graphical and descriptive, and is strongly related to the functions applied to these objects.

Metric calculations, technical calculations, positioning and geographic projections : Metric calculations involve measuring distances and areas, while technical calculations include visibility, volumes, and operations research. GIS make it easy to calculate areas, distances, and volumes from object location data. Geographic projection transformations are also easily accessible. Operations research (mainly path calculations in graphs) finds in GIS all the necessary data.

Management and processing of object collections : This is one of the fundamental objectives of GIS. After structuring the information, it must be captured and managed by the system. GIS often delegate the management of descriptive data to classical relational DBMS (such as ACCESS, ORACLE, SQL Server, DBase, etc.) and focus mainly on object location and the links between graphical and descriptive parts.

These tasks are ensured through proper management of information flows, updates, modifications, especially for the graphical component.

Administrative management and data sharing : When data are distributed among multiple users, as is often the case in administrative applications such as cadastre, GIS are designed to manage this sharing and optimize data access among users.

Spatial management and analysis : GIS must handle all types of geographic objects, from points to pixels, including areas and networks. The goal is to build a georeferenced database by linking different objects regardless of their type.

This connectivity ensures spatial analysis, meaning that location is taken into account in data analysis. A large number of processes involving object location are implemented in GIS (distance-based selection, operations research, spatial aggregation and scaling, geojoining, interpolation, vectorization, proximity classification, etc.).

Statistics and geostatistics : Building a geographic database often aims at studying all objects within a territory. GIS must therefore provide easy access to statistical calculations, whether exploratory or methodological. Some GIS include statistical modules, while others are connected to specialized software. Geostatistical methods are also important objectives since spatial management facilitates their use.

Spatio-temporal management : Introducing time into GIS allows queries combining space and time, enabling the management of object history and system states at a given time. However, practical implementations remain limited due to the complexity of managing temporal evolution.

Simulation and modeling : Implementing models for process simulation is now considered an objective of GIS. An interface must exist between modeling programs and the geographic database.

Remote sensing, georeferencing and image processing : GIS are intended to manage all types of geographic objects. Remote sensing provides an important data source. Imaging techniques ensure proper georeferencing and facilitate access to large datasets.

Cartographic drawing and 3D mapping : GIS aim to display data resulting from queries. Automatic mapping modules are essential. 3D representation is now integrated, although its maintenance remains complex.

Internet and remote accessibility : The Internet offers new possibilities for remote data access. GIS must exploit this capability to provide simple access and mapping tools online.

1.5 Components of a GIS

According to Burrough [4], a GIS consists of three main components :

Hardware : This includes a computer (CPU), storage devices, digitizers, scanners, and output devices (screen, plotter, printer, etc.).

Software : The software part includes modules for :

1. Data input and validation ;
2. Database management ;
3. Data display ;
4. Data transformation ;
5. User interaction (HCI).

Organizational aspect : Hardware and software alone do not guarantee efficiency. Proper organization involving users, operators, and managers is required.

1.6 Acquisition of Geographic Data

Several methods exist :

Field surveys :
1. Equipment : total station, electronic field notebook.
2. Accuracy : depends on equipment and measurement errors.
3. Coordinates : Cartesian or geodetic.

GPS measurements :
1. Equipment : GPS station.
2. Accuracy : depends on measurements and equipment.
3. Coordinates : Cartesian/geodetic.

Photogrammetry :
1. Equipment : analytical/digital restitutor.
2. Accuracy : depends on scale and measurements.
3. Coordinates : Cartesian/geodetic.

Remote sensing :
1. Equipment : onboard scanner.
2. Accuracy : depends on resolution.
3. Coordinates : image coordinates.

1.7 Conclusion

Geographic Information Systems are widely used in many domains. They can now be considered distributed systems handling spatial information.

In this chapter, we presented definitions, structures, architectures, and functionalities of GIS. The following chapters will detail components, functions, and data modeling methods.

Questions

1. Give the five basic application functionalities of a GIS, the "five A".
2. Give the basic business functionalities of a GIS.
3. Give the different ways to locate an object.
4. There are two important models for implementing a GIS. Name them.
5. Give the two formats of geographic information.

Answers

1. The five "A" of a GIS :
   - Abstraction
   - Acquisition
   - Archiving
   - Analysis
   - Display
2. Business functionalities :
   (a) Where
   (b) What
   (c) How
   (d) When
   (e) If
3. Ways to locate objects :
   - Indirect
   - Direct
4. Database modeling and cartographic modeling.

2

Spatial Data Structures

Sommaire

2.1 Introduction **18**
2.2 Geographic database **18**
2.2.1 Advantages of using databases 19
2.2.2 Database models 20
2.3 Modes of representation of geographic data **22**
2.4 Data structures in vector mode **23**
2.4.1 Non-topological structures 24
2.4.2 Topological structures 25
2.5 Structure specific to elevation information **30**
2.5.1 Irregular triangulated network structure : RTI (Triangulated Irregular Network : TIN) 31
2.5.2 Topological structure for contour lines 31
2.6 Raster mode **35**
2.7 Conclusion **38**
2.8 The Route Table **47**
2.9 The Building Table **47**
2.10 The Wilaya Table **47**
2.11 Attribute Queries **48**
2.12 Spatial Queries **48**
2.13 Conclusion **49**
2.13.1 Attribute queries 52
2.13.2 Spatial queries 52
2.14 Conclusion **53**

2.1 Introduction

Historically, spatial data have been represented in the form of maps. They include areas and a small amount of semantic data that briefly describe the concerned localities. With the advancement of hardware and software systems for data processing, the linkage between semantic and spatial data has been strengthened. However, data acquisition, filtering, and updating remain tasks that continue to evolve. Data resources are often complex and contain inaccuracies. Therefore, data must be reorganized and restructured so that they can be easily reused. In general, data structures are designed to avoid redundancy and facilitate processing operations ; in other words, the allocated memory space and, above all, the computation time must be optimized. Indeed, complex data contain inaccuracies that are due to

1. information redundancy : redundant information may result from direct duplication, for example when the same locality address appears in several files, or by induction. For example, both object coordinates and distances may be stored, even though distances can be calculated from coordinates
2. fuzzy semantics : for example, if distance calculations on a map are based on an improperly chosen scale.
3. poor-quality media, such as badly scanned texts.

2.2 Geographic database

It is a collection of spatial and non-spatial data structured and organized in such a way that they can be queried and analyzed interactively or automatically. A geographic database often targets a defined area. It has the following characteristics.

1. It is managed and administered by a GIS software.
2. It includes the data themselves as well as their metadata.
3. A database is a set of representations of reality in the form of interrelated data, as consistent as possible, stored and structured in a way that facilitates their use to satisfy a wide variety of applications.

Initially, spatial data were represented in the form of graphical data. Maps (topographic maps, thematic maps, cadastral plans, etc.) provide a large amount of spatially referenced information that is open to a wide variety of applications. These cartographic documents form graphical databases. However, it is easy to understand the limitations of these graphical databases, which are mainly

1. information loss,
2. the essential reliance on graphics specialists,
3. the static nature of the map, which makes it difficult to infer the dynamics of the phenomenon to be represented,
4. printing costs, etc.

With the evolution of distributed systems and information systems in general, graphical data are digitized and structured in such a way as to :

1. preserve and store original data : data are recorded independently of graphical representation,
2. enable dynamic management : allowing rapid and continuous updates over time ;
3. facilitate processing and manipulation : this is ensured through scale or projection changes, combination of themes, etc.,
4. enrich graphical processing : the digital mode considerably reduces the cost of cartographic production.

As indicated in the first chapter, geographic information systems must be designed to be capable of managing standard (semantic) and spatial databases as well as the functional dependencies between spatial and semantic characteristics. In practice, the implementation of GIS relies on existing database management systems (DBMS) through three approaches.

1. DBMS extension : Data are fully accessible through the management system and must be structured according to the models designed by the DBMS developer. Thus, existing data structures are adapted to properly handle spatial characteristics.
2. Hybrid solution : The hybrid solution is based on the coexistence of two models for spatial and thematic data. Semantic data are managed through the system (often attribute tables and relationships) because they conform to DBMS models. Data that model spatial information (geometry and location data) are handled by the spatial system because they do not conform to DBMS models.
3. Integration of the spatial domain into the DBMS : Development of DBMS including the necessary functionalities to manage spatial entities. This is the case of object-oriented DBMS that are currently under development.

2.2.1 Advantages of using databases

We list some advantages of using databases and the approaches corresponding to systems dedicated to management and administration. These advantages are associated with functional characteristics.

1. minimization of duplicated data (redundancy) : conceptual data models aim to minimize data as much as possible in order to optimize memory space usage. Experience gained and extensive research in modeling in this field clearly contribute to better data structuring.
2. ensuring data integrity and improving quality : structural models ensure association links and provide simple and understandable visibility.
3. preservation of semantic definitions with data (metadata) : although data are not often replicated, semantic characteristics and functional and logical links are preserved.
4. reduction of software development cost (many fundamental operations are handled by the DBMS) and avoiding modification of application software each time the database structure changes ;

5. minimization of data access and preservation of information security.

The mentioned benefits mainly concern the functional characteristics provided by databases, but non-functional characteristics such as availability, reliability, security, and others are also ensured. In short, the evolution of information systems and distributed systems leads to better formalization of data structures.

2.2.2 Database models

Different database models can be used to build a geographic database. In practice, some models sometimes lead to the same data structures such as the network, entity-relationship, and relational models ; however, conceptual views differ, and the scalability of the model determines the choice. The types of database models practically used are as follows.

1. Tabular model : data are stored in the form of a simple table of rows and columns, as in spreadsheet software.
2. Hierarchical : data are stored in the form of a hierarchy. A child can have only one parent, while a parent can have several children. To identify relationships between nodes, the hierarchy must be respected. The example in Figure 2.1 illustrates the simplicity of using this model to represent the geographic information of a country. The map in Figure 2.2 presents a locality composed of two zones. Zone I consists of four vertices (1, 2, 3, and 4) and four lines. Zone II also consists of four vertices (4, 3, 5, and 6) and four lines. The line connecting vertices (3, 4) is a boundary line between the two zones. Figure 2.3 illustrates the hierarchical model associated with map M. At the top, the name of the locality « M »is shown. Then, map M has two children : zone I and zone II. The zones have segments (lines) as children. The vertices define the children of the segments. The coordinates of the vertices can be added as children of the vertices.

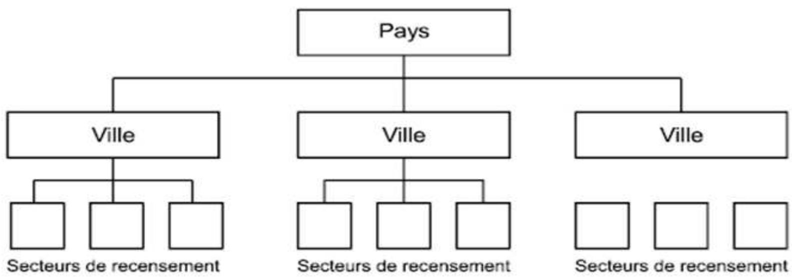


FIG. 2.1 – The hierarchical model of a country.

3. Network model : data are organized to build networks between the represented entities. An entity can have multiple parents and multiple children at the same time. In fact, the network model is much more flexible than the hierarchical model and supports complex spatial relationships. On the other hand, the network model significantly reduces data duplication. In Figure **??**, the network model is applied.

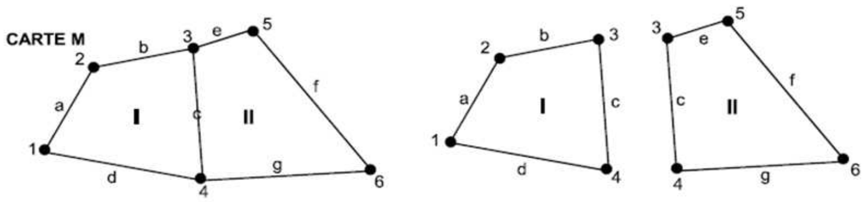


FIG. 2.2 – A map M composed of two regions.

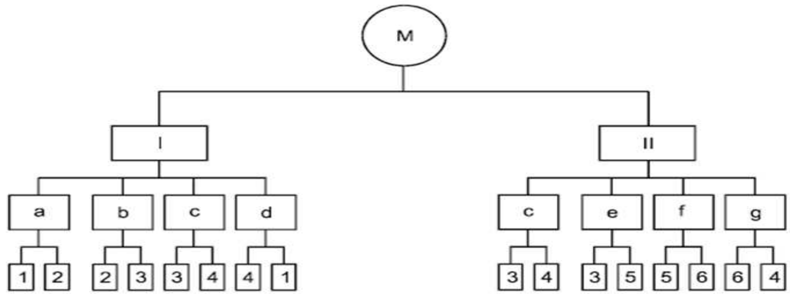


FIG. 2.3 – The hierarchical model of map M.

The line « c »has two parents : zone I and zone II. Thus, it is easy to extract common segments between zones. Unlike the hierarchical model where a large part of the structure must be traversed to extract common segments. Moreover, segment c is duplicated in the hierarchical model.

4. Relational : data are linked through relationships that establish correspondence between information distributed across several tables. The relational model was first defined by E. F. Codd, a researcher at IBM, in 1970. This model provides a flexible approach to links between records, allowing better modeling of complex spatial relationships that may exist between objects represented in the database. Data are arranged in simple records containing an ordered collection of attribute values. These collections are grouped into two-dimensional tables. Each table is often stored in a separate file. Relationships between tables are ensured by keys, where each record is identified by a key. The following three criteria must be respected.
   (a) Two records (rows) are never identical.
   (b) The order of rows is not important.
   (c) The order of columns must be respected.

   Figure 2.4 represents the relational model corresponding to map « M ».

5. The object-oriented model : data structures are modeled around the concept of objects that can be assimilated to real-world entities. They are less widespread because relationships through pointers between objects are difficult to store in persistent memory. However, the power of these models in terms of design, fast processing, and modularity leads them to play an important role in GIS. The need for new concepts

Les sommets

Id_sommet	Coordonnée X	Coordonnée y
1	10	10
2	15	30
3	46	35
4	50	2
5	65	40
6	80	6

Les zones

ID_zone	Nom du zone	....
I	Région est	....
II	Région ouest	....

Les segments (lignes)

ID_seg	sommet_1	Sommet_2	Zone_gauche	Zone_droite	....
a	1	2	-----	I	....
b	2	3	------	I	....
c	3	4	I	II	....
d	1	4	------	I	....
e	3	5	------	II	....
f	5	6	II	---------	....
g	4	6	------	II	....

Fig. 2.4 – The relational model of map M.

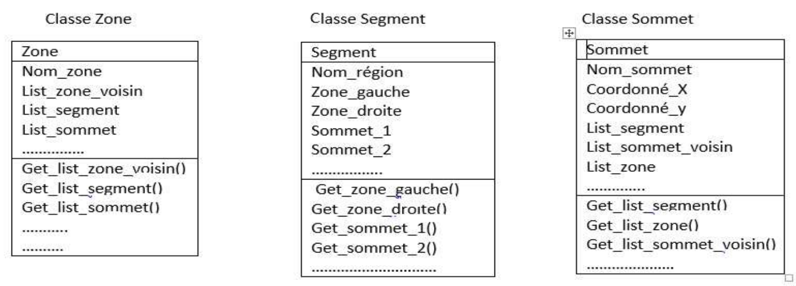


Fig. 2.5 – Classes of the object-oriented model of map M.

and tools that go beyond the limits of relational models and the emergence of object-oriented programming languages represent additional factors pushing designers to adopt these models. Data and processing are no longer separable. Each object has its own functions. Relationships between objects can be implemented in several ways (sets, lists, etc.). Indexing objects can be done directly without passing through join operations. Figure 2.5 represents the objects associated with the map components : region, zones, segments (lines), and points. The vertices composing a zone can be easily extracted using the « get_list_sommet »function of the « zone »class. Whereas in the relational model, a search across the three relational tables is necessary.

2.3 Modes of representation of geographic data

Geographic data are represented in two modes : vector mode and raster mode [12] [15] [4] [1]. The basic logical unit in the vector model is the line (boundary of an area, river, contour line, etc.). Indeed, a series of points (defining locations by coordinates X, Y) are stored along the line. A point can be defined as a line of zero length. In raster mode, the basic unit is the pixel (picture element) : a cell that represents a portion of the surface. Thus, areas are subdivided into adjacent cells of the same size. The shape of these cells is often rectangular, although other shapes can be used (hexagonal or triangular). In raster mode, the approaches and methods used come from the field of image processing. In practice, geographic information systems use both modes simultaneously to exploit the advantages of each. Indeed, each mode complements the other.

To design localized geographic information, three levels of information are mainly used : the geometric level, which concerns the geometry of geographic objects ; the semantic level, dedicated to the nature of geographic objects ; and the topological level, which represents connectivity and linkage relationships between geometric objects [12][10].

1. Geometric data consist of points, lines, and polygons. A line can be modeled by a set of segments (a segment is a straight line composed of 2 points), a set of more complex curves (cubic, etc.), or both, i.e., a set of segments and curves. Whereas a polygon is composed of a closed line. Any polygon can contain other polygons inside it (a park in a city, an island in a lake, a courtyard inside a house, etc.).
2. Semantic data represent the non-geometric characteristics of geographic data. They can describe simple entities (intersection, building, road segment) or more complex entities (road, interchange, group of buildings, etc.). Indeed, the semantic characteristics of an object such as name, address, or width can be associated with the concerned object.
3. Topological data describe connectivity relationships between geometric objects, namely points, lines, and polygons. Geometric data can be modeled as a graph. A graph consists of nodes (points) and arcs (directed or undirected lines). A node lies between two nodes, one being the initial node and the other the final node. Arcs meet at a node ; arcs directed toward the node are incoming arcs, and arcs originating from the node are outgoing arcs. The intersection between lines is represented by a node. Connectivity relationships in a graph are called topological relationships.

The example in Figure 2.6 represents a formalization based on concepts used in graph theory.

2.4 Data structures in vector mode

To better design storage models for spatial data in vector mode, data structures are categorized into two main classes : topological and non-topological structures [14] [15] [4] [1]. Non-topological models do not store connectivity relationships between geometric objects ; therefore, these relationships are derived by software. In contrast, in topological models, topological relationships are explicitly stored.

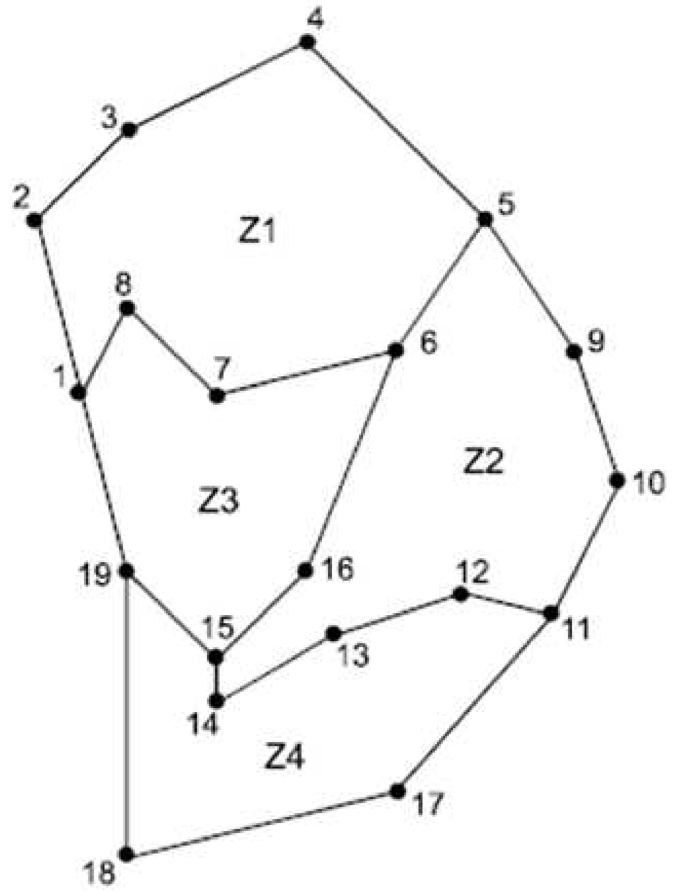


FIG. 2.6 – Example of formalization of a map M_1 according to a model inspired by graph theory.

2.4.1 Non-topological structures

In the simplest models, only point and line notions are stored. A data dictionary is used to store sequences of points. Two successive points represent a straight or curved line. Only the size and shape of spatial objects can be easily derived from point sequences. These models lead to what is known as the « spaghetti structure ». The « simple sequential structure »is considered a non-topological structure. The database is stored as a file of sequences of pairs of point coordinates : X, Y. The structure organization is as follows.

1. A sequence of two pairs of points represents a line.
2. A polygon spatial entity is described by a series of pairs.
3. A polygon is independent of others ; in other words, each series of pairs is dedicated to a single polygon.
4. A spatial coordinate pair is used to separate series (polygon corresponding to a zone). closing of polygons is implicit by repeating the coordinate pair of the initial point

X	Y
x1	**y1**
x2	y2
x3	y3
x4	y4
x5	y5
x6	y6
x7	y7
x8	y8
x1	**y1**
-1	*-1*
x5	**y5**
x9	y9
x10	y10
x11	y11

X	Y
x12	y12
x13	y13
x14	y14
x15	y15
x16	y16
x6	y6
x5	**y5**
-1	*-1*
x1	**y1**
x8	y8
x7	y7
x6	y6
x16	y16
x15	y15

X	Y
x19	y19
x1	**y1**
-1	*-1*
x19	**y19**
x15	y15
x14	y14
x13	y13
x12	y12
x11	y11
x17	y17
x18	y18
x19	**y19**
-1	*-1*

FIG. 2.7 – Sequential structure of map M_1.

Figure 2.7 shows an example of using the sequential model to store zones of map M_1. The region of map M_1 is composed of 4 closed zones. In the sequential file, zones are separated by -1. The coordinate pair of the initial point is repeated (for zone z_1, it is x1, y1). Common coordinates between zones are repeated for each zone. To reduce duplicate input of points, a point dictionary is used (see Fig. 2.8).

Sequential structures are easy to design, but they overcome some drawbacks, mainly :

1. storage space : duplicated data require additional memory space.
2. development cost : to implement basic functions such as searching for common segments between zones, significant effort is required.
3. processing time during use is also important : to obtain segments between two zones, for example, the file must be scanned to identify the concerned zones. Then segments must be determined by examining successive pairs in both sequences. Finally, a comparison is performed.

2.4.2 Topological structures

More advanced structures are called topological structures because most topological relationships are explicitly stored. These structures are designed as graphs ; therefore, a set of rules derived from graph theory is applied to model topological relationships properly. These rules are listed as follows.

Fichier des zones

Pointeurs		
	p11	p11
p1	p12	p17
p2	p13	p18
p3	p14	**p19**
p4	p15	*-1*
p5	p16	**p1**
p6	p6	p8
p7	**p5**	p7
p8	*-1*	p6
p1	**p19**	p16
-1	p15	p15
p5	p14	p19
p9	p13	**p1**
p10	p12	*-1*

(a)

Dictionnaire

Etiquette	X	Y
p1	x1	y1
p2	x2	y2
p3	x3	y3
p4	x4	y4
p17	x17	y17
p18	x18	y18
p19	x19	y19

(b)

FIG. 2.8 – Sequential structure with point dictionary of map M_1.

1. Graph nodes or vertices model map background points,
2. Arcs or edges connecting two vertices describe segments.
3. Zones are represented by a sequence of arcs and thus appear as cycles in the graph.
4. Separate files are created to distinguish graph elements.
5. Pointers between files describe topological relationships between graph elements forming the main map entities (zones, lines, and points).

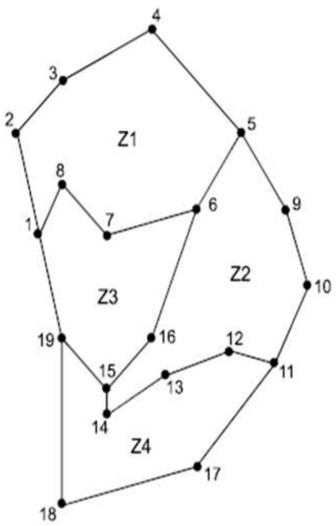
The background points that represent intersections between two lines on the map (roads, rivers, boundaries between zones) are represented by graph vertices. Lines are described by arcs or edges and stored in another file. Cycles in the graph model zones. For each entity (points, lines, zones), a separate file is usually dedicated. In most structures, the topological relationship zone-line is represented by pointers between the line file and the zone file. The point-line relationship is also represented by pointers. The zone-point relationship is derived by software. The point-point relationship is ensured by ordering in vertex sequences. Other topological relationships (« zone-zone », « line-line ») depend on the structure used (they are either pointers or derived by software).

DIME structure : Dual Independent Map Encoding

DIME is considered the first and most famous operational system that allows modeling a spatial data structure. This system was designed for visualization and management of a database representing urban road networks and intra-urban statistical units used for US censuses since 1970. The segment associated with a face of an urban block is the basic element of this structure. Complex lines such as railways and streets are described by a series of successive segments. The arrangement of edges identifies blocks, from which census data are gathered. A node label file and a node dictionary file are used to define zones.

Indeed, the database consists of a coordinate file storing graph vertices and an edge file. The node file consists of records. Each record contains a node label and the two coordinates of the vertex : X and Y . Therefore, no two nodes have the same label and coordinates. Each record of the second file is associated with an edge. It contains a label, a pointer to the start node, a pointer to the destination node, and two codes identifying the two blocks on each side of the edge. Other semantic (attribute) information, such as street name, can be added. To better control the data acquisition stage, the following criteria must be respected.

1. Two arcs do not have the same label.
2. The start node pointer is not empty or null.
3. The destination node pointer is not empty or null.
4. Start and destination node pointers are not the same and do not point to the same arcs (in most systems).
5. At least one of the two neighbor codes is not empty.
6. Codes indicating neighboring zones of an arc are not identical.
7. Nodes represent overlaps between arcs ; in other words, two arcs do not intersect.

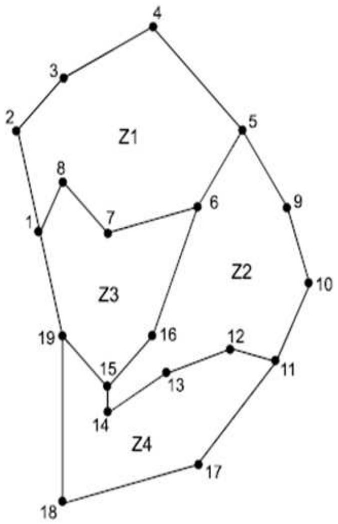


Etiquette	X	Y
N1	X_{N1}	Y_{N1}
N2	X_{N2}	Y_{N2}
N3	X_{N3}	Y_{N3}
...	...	...
N17	X_{N17}	Y_{N17}
N18	X_{N18}	Y_{N18}
N19	X_{N19}	Y_{N19}

(a) (b)

FIG. 2.9 – Nodes of map M_1 according to DIME structure. (a) map M_1 ; (b) node file.

Conceptually, these criteria are respected because this model is based on graph theory supported by a solid mathematical foundation. However, during data acquisition (input, radar acquisition, etc.), a single coordinate error for a node is enough to violate these criteria. Verification is done using non-automatic methods such as displaying the map from the structure. A GIS at a country level requires distributed databases and a large number of acquisition tools (including human agents). It is important to implement software tools



(a)

Varianles Urbanist.	Noeud de Départ	Noeud de Destinat..	Zone gauche	Zone droite
...	N1	N2	0	Z1
...	...	...	...	...
...	N5	N6	Z2	Z1
...	...	...	...	...
...	N18	N19	0	Z4
...	N19	N15	Z3	Z4
...	N19	N1	0	Z3

(b)

FIG. 2.10 – Edges of map M_1 according to DIME structure. (a) map M_1 ; (b) edge file.

for structure verification.

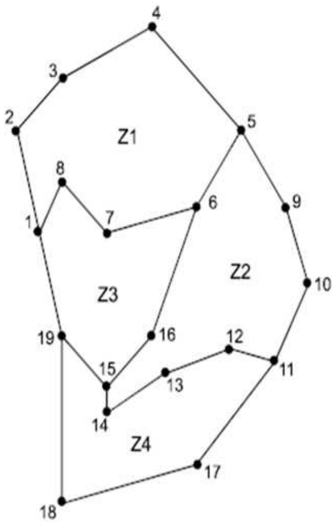
Figures 2.9 and 2.10 illustrate map modeling using the DIME model. Map M_1 is composed of 19 nodes and 22 arcs. The node file (see Fig. 2.9 (b)) consists of three attributes : node label, X coordinate, and Y coordinate. Conventionally, the attributes of the edge file (see Fig. 2.10 (b)) are ordered as : Label, start node, destination node, left zone, right zone.

POLYVRT structure (POLYgone conVeRTer)

This structure uses a more detailed modeling of spatial components that form the map background. This structure is considered a reference for several systems designed later. The line is also a basic element on which it is based. A chain is defined as consisting of one or more successive segments. It can alone represent a boundary such as an island. Segments are connected by intermediate points. A chain is bounded by two nodes, while a segment lies between two intermediate points or between an intermediate point and a node if it is at the end of a chain. Therefore :

1. an intermediate point belongs to only one chain ;
2. nodes represent meeting points between chains and may belong to several chains ;
3. a chain must belong to at least one zone and at most two adjacent zones ;
4. zones are defined by a sequence of chains forming a closed loop ;
5. a chain must contain at least and at most two nodes at its ends.

The POLYVERT structure is composed of four files : a zone file, a chain file, an intermediate point file, and a node file. If semantic information is excluded, the zone and chain files contain only pointers. Records in each file are organized as follows.



(a)

Etiquette	X	Y
N1	XN1	YN1
N2		
N3		
N4		
N5		
N6	XN6	YN6

(b)

FIG. 2.11 – Nodes of map M_1 according to POLYVRT structure. (a) map M_1 ; (b) node file.

Etiquette	Liste des coord. des points
C1	$x_{p1}, y_{p1}, x_{p2}, y_{p2}, x_{p3}, y_{p3}$
C3	$x_{p4}, y_{p4}, x_{p5}, y_{p5}$
C4	$x_{p6}, y_{p6}, x_{p7}, y_{p7}$
C5	$x_{p8}, y_{p8}, x_{p9}, y_{p9}$
C6	x_{p11}, y_{p11}
C9	$x_{p12}, y_{p12}, x_{p13}, y_{p13}$

(a)

Etiquette	Liste de pointeurs de chaînes
Z1	C1 C2 C3
Z2	C4 C5 C6 C2
Z3	C3 C6 C7 C8
Z4	C9 C7 C5

(b)

FIG. 2.12 – Intermediate points file and zones file of POLYVRT structure for map M_1. (a) intermediate points file ; (b) zones file.

1. Node file : each record stores a label identifying the node, and fields for X and Y coordinates.
2. Intermediate point file : the chain label is stored after the list of coordinates of intermediate points. Coordinates are separated by commas or another non-numeric character.
3. Chain file : each record contains, in order :
   (a) pointer to the label representing the intermediate point list of the chain,

Etiquette	Nombre de points	Noeud de départ	Noeud de destination.	Zone gauche	Zone droite
C1	3	N1	N2	0	Z1
C2	0	N2	N3	Z2	Z1
C3	2	N3	N1	Z3	Z1
C4	2	N2	N4	0	Z2
C5	3	N4	N5	Z4	Z2
C6	1	N5	N3	Z3	Z2
C7	0	N5	N6	Z4	Z3
C8	0	N6	N1	0	Z3
C9	2	N4	N6	0	Z4

FIG. 2.13 – Chain file of POLYVRT structure for map M_1.

(b) number of intermediate points,
(c) pointer to the start node,
(d) pointer to the destination node,
(e) pointer to the left zone (or upper zone),
(f) pointer to the right zone (or lower zone).

4. Zone file : the zone label followed by the ordered list of chains forming the zones is stored.

Figures 2.11, 2.12, and 2.13 illustrate the files used to represent map M_1 using the POLYVRT model. The number of nodes in the map is 6, chains are 9, intermediate points are 13, and zones are 4. The same labels used in the intermediate point file to identify a chain are used as pointers in the zone and chain files. Similarly, node labels in the node file are used as pointers in the chain file.

Two remarks are important in the POLYVERT structure. Nodes are stored in a separate file to avoid redundancy. Intermediate point coordinates are listed directly in the intermediate point file because each intermediate point belongs to only one chain. However, a zone file is still maintained even though zones can be derived from the chain file. The advantage of explicitly storing topological relationships is improved system performance and reduced development cost.

2.5 Structure specific to elevation information

Elevation information concerns data related to the height of spatial units. To model relief such as mountains, plateaus, or ocean depths, specific structures are designed such as the irregular triangulated network and contour lines.

2.5.1 Irregular triangulated network structure : RTI (Triangulated Irregular Network : TIN)

This structure was first proposed in 1970 [15]. After acquiring control points, a selection process identifies the most significant points according to elevation variation. Points are connected to form triangles. Three points A, B, C are connected to form a triangle if a circle « C »passing through these points does not contain any other point inside it. Figure 2.14 shows the triangle selection process.

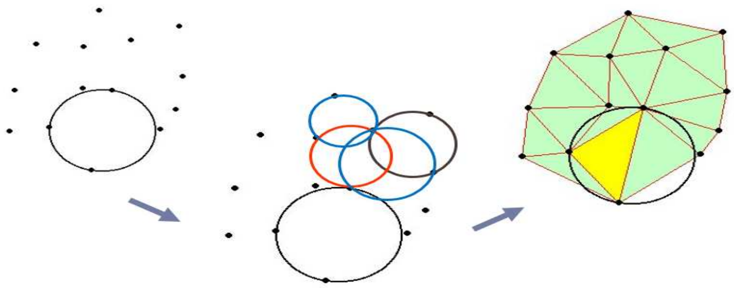


FIG. 2.14 – Construction of triangles from points.

Two methods can be used to encode triangles : by triangle or by vertex.

1. RTI representation by triangle : the database consists of a single file. Each record contains :
   (a) triangle identifier,
   (b) coordinates of the three vertices,
   (c) identifiers of neighboring triangles.
2. RTI representation by vertex : records contain :
   (a) vertex identifier,
   (b) vertex coordinates,
   (c) pointers to neighboring vertices.

Figures 2.15, 2.16, and 2.17 illustrate these representations.

For visualization, a projection onto the (x, y) plane is performed. A color gradient is used to improve visual inspection (see Fig. 2.18).

2.5.2 Topological structure for contour lines

Elevation data structures are based on contour lines to support complex processing needs. The basic idea is to identify neighboring points with the same altitude on the same line, ensuring efficient organization of base data. Topological relationships between neighboring points are ensured by ordering in the point file. Files for upper and lower neighbors allow direct access to adjacent contour levels, modeling line-line relationships efficiently. Figure 2.19 shows an example of contour lines.

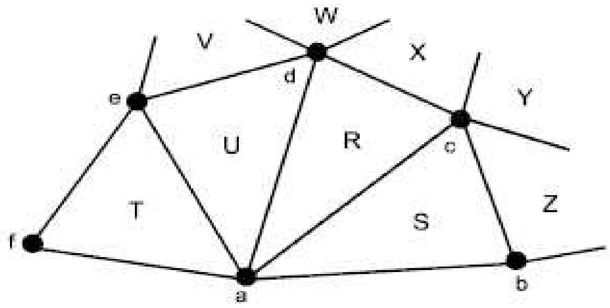


FIG. 2.15 – Irregular triangulated network structure.

Triangle	Sommets	Triangles adjacents
R	a(x,y,z), c(x,y,z), d(x,y,z)	S, X, U
S	a(x,y,z), c(x,y,z), b(x,y,z)	Z, R
etc.		

FIG. 2.16 – Triangle-based data storage.

Sommet	Coordonnées	Sommets connectés
a	x, y, z	f, e, d, c, b, ...
b	x, y, z	a, c, ...
c	x, y, z	a, d, b, ...
etc.		

FIG. 2.17 – Vertex-based data storage.

This structure is developed to handle complex elevation data processing. The topological contour structure is implemented using four files :

1. entity file (line or contour),
2. coordinate file of base points,
3. lower neighbors file,
4. upper neighbors file.

Entity file

This is the main file of the structure. Each contour is defined by an identifier, altitude, shape, and pointers to other files. Each record includes :

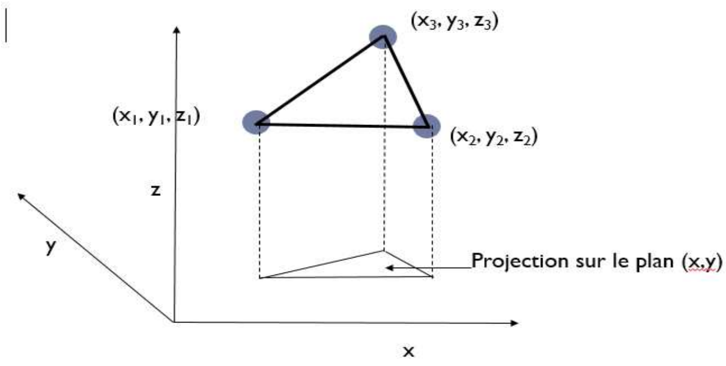


FIG. 2.18 – Projection of points onto the (XY) plane.

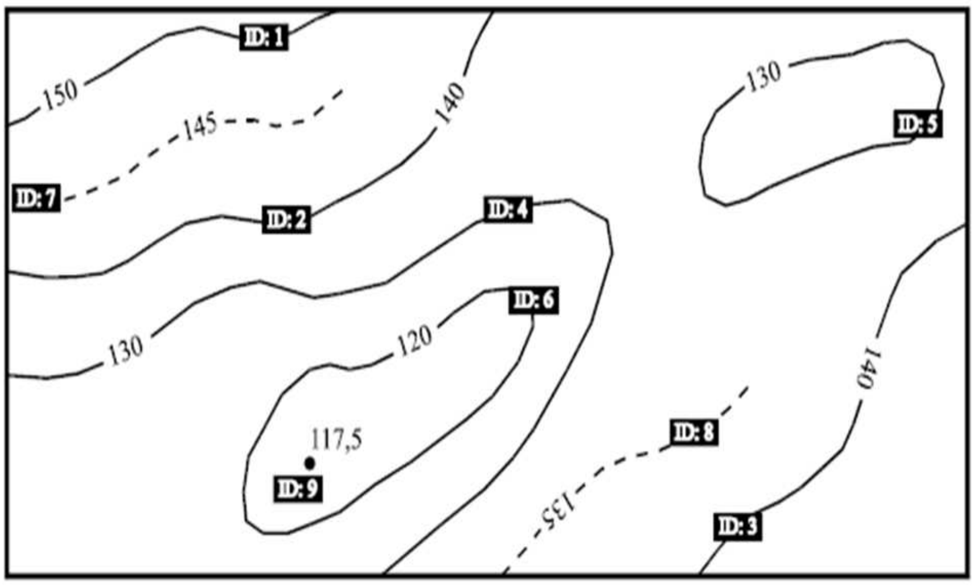


FIG. 2.19 – Contour lines.

1. ID : unique identifier,
2. ALT : altitude,
3. NP : number of points,
4. APD : address of starting point,
5. FOR : shape (F closed, O open, P point),
6. CON : continuity,
7. LON : length,
8. AVS : pointer to upper neighbors,
9. NVS : number of upper neighbors,
10. AVI : pointer to lower neighbors,

11. NVI : number of lower neighbors.

Figure 2.20 illustrates the entity file.

ID	Alt	NP	APD	LON	FOR	CON	NVI	AVI	NVS	AVS
1	150.0	8	1	2749.3	O	C	2	1	0	0
2	140.0	15	9	4754.0	O	C	3	2	3	1
3	140.0	9	24	2862.0	O	C	3	3	1	2
4	130.0	21	33	6893.6	O	C	1	4	4	3
5	130.0	12	54	3505.2	F	C	0	0	4	4
6	120.0	13	66	4852.2	F	C	1	5	1	5
7	145.0	7	79	2563.4	O	I	1	6	1	6
8	135.0	7	86	2571.6	O	I	2	7	2	7
9	117.5	1	93	0.0	P	P	0	0	1	8

FIG. 2.20 – Entity file of the previous map.

Coordinate file

The coordinate file stores all points of contour lines and spot heights. Point order is important since successive pairs define line segments. The file uses APD and NP to identify each contour. If the contour is closed, a segment connects the first and last point. Data are non-redundant since each point belongs to a single contour.

Figure 2.21 shows the coordinate file.

Lower neighborhood file

This file stores lists of identifiers of lower neighboring contours or points. These represent entities with lower altitude. If two entities share the same list, duplication is allowed to ensure extensibility. Figure 2.22 illustrates this file.

Upper neighborhood file

This file stores higher-altitude neighbors. It mirrors the lower neighborhood file but for higher elevation contours. Figure 2.23 shows this structure.

Adresse	X	Y	Adresse	X	Y
1	$x1_1$	$y1_1$	–	–	–
2	$x1_2$	$y1_2$	–	–	–
3	$x1_3$	$y1_3$	88	$x8_3$	$y8_3$
4	$x1_4$	$y1_4$	89	$x8_4$	$y8_4$
5	$x1_5$	$y1_5$	90	$x8_5$	$y8_5$
6	$x1_6$	$y1_6$	91	$x8_6$	$y8_6$
–	–	–	92	$x8_7$	$y8_7$
–	–	–	93	x9	y9

FIG. 2.21 – Coordinate file.

Adresse	ID des Voisins inférieurs
1	7, 2
2	4, 5, 8
3	4, 5, 8
4	6
5	9
6	2
7	4, 5

FIG. 2.22 – Lower neighborhood file.

2.6 Raster mode

The raster mode consists of subdividing the study area using a regular grid. The basic unit is a surface element. Three types of grids are commonly used :

1. square grid,

Adresse	ID des Voisins supérieur
1	1 , 7 , 3
2	2
3	2 , 5 , 3, 8
4	2 , 3, 4, 8
5	4
6	1
7	2 , 3
8	6

FIG. 2.23 – Upper neighborhood file.

2. triangular grid,
3. hexagonal grid.

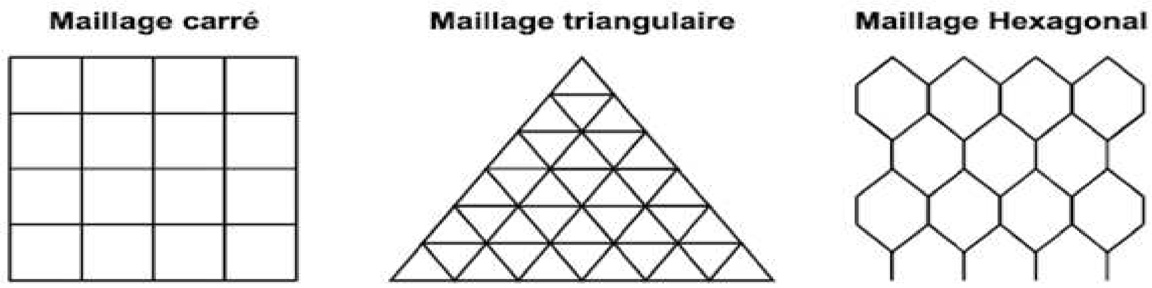


FIG. 2.24 – Three commonly used grid types : square, triangular, and hexagonal grids.

Figure 2.24 shows the three commonly used grids. The square grid is the most used for two reasons :

1. easy use by computers,
2. compatibility with spatial data acquisition and display devices.

The study area is divided into a grid where each square corresponds to a pixel. Each pixel is referenced by row and column indices. The grid is represented as a matrix. Each pixel has three attributes : row index, column index, and a pointer to a table storing grayscale values. Direct storage of the matrix requires large memory space. Compression techniques are used to reduce storage requirements.

In raster structures :

Ligne										
0	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	3	3	3	3	3	3
3	2	2	2	2	3	3	3	3	3	3
4	2	2	2	2	3	3	3	3	3	3
5	2	2	2	2	3	3	3	3	3	3
6	2	2	2	3	3	4	4	4	4	4
7	2	2	2	3	4	4	4	4	4	4
8	2	2	2	4	4	4	4	4	4	4
9	2	2	2	4	4	4	4	4	4	4

Ligne	Valeur	Long.
0	1	10
1	1	10
2	1	4
2	3	6
3	2	4
3	3	6
4	2	4
4	3	6
5	2	4
5	3	6
6	2	3
6	3	2
6	4	5
7	2	3
7	3	1
7	4	6
8	2	3
8	4	7
9	2	3
9	4	7

FIG. 2.25 – RLC method (Run Length Coding). Left : bitmap image. Right : RLC format.

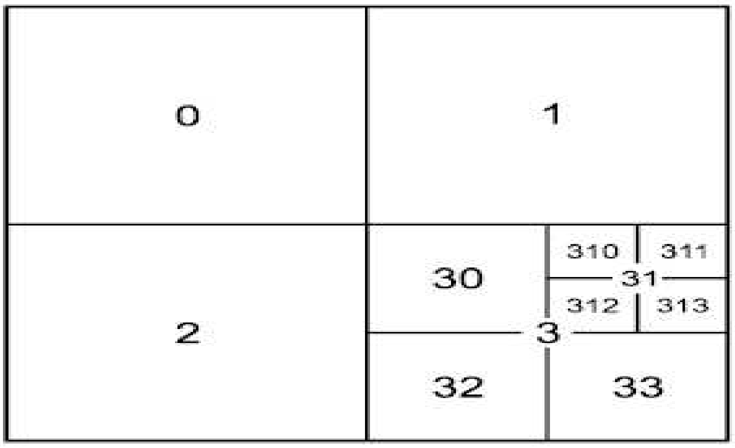


FIG. 2.26 – Square block method applied to image matrix.

1. a point is represented by a pixel,

2. a line is represented by a set of adjacent pixels,
3. a zone is represented by an aggregation of adjacent lines.

A geographic zone is considered a planar surface, while a pixel represents a square terrain portion. Resolution is defined by the relation between pixel size in the database and real-world size. Each pixel has a single value, so a three-layer structure is used for color representation.

Common storage methods include :

1. Bitmap,
2. RLC,
3. Quad-tree square blocks.

Bitmap stores image line by line. The size is $L \times C \times T$ where L is number of rows, C columns, and T bytes per pixel.

Run Length Coding (RLC) compresses repeated pixel values by storing value and repetition count per line. Compression is effective when repetitions exceed 2.

Figure 2.25 shows an example of RLC encoding.

Square blocks generalize RLC in two dimensions. Blocks are defined by :

1. coordinates of top-left pixel,
2. block length,
3. uniform pixel value.

Quad-trees are hierarchical tree structures dividing images recursively into four parts until uniform regions are obtained. A block is defined by :

1. branch code,
2. pixel value.

Raster structures are used when detailed spatial distribution is needed, while vector structures are used for precise localization (cadastre, mapping). Today, both are often combined in multi-layer systems.

2.7 Conclusion

To meet multidisciplinary needs of spatial systems, several approaches from different fields are used such as image processing, graph theory, geometry, etc. Data acquisition, storage, archiving, analysis, and visualization are designed to preserve modularity. Most GIS designers also reuse existing systems to reduce development cost, while research on specialized spatial techniques continues today.

Exercise 1

Let map C be the following :

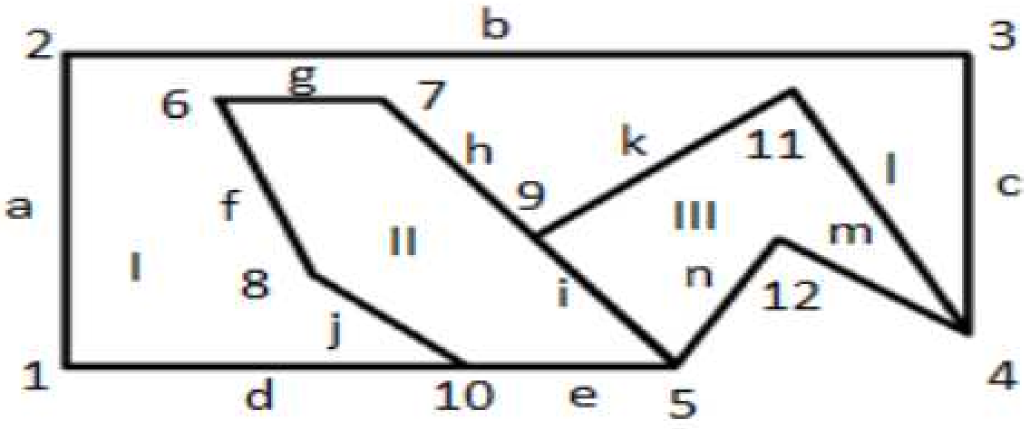


FIG. 2.27 – Carte C

- Give the hierarchical model of map C.
- Give the network model of map C.
- Give the relational model of map C.

Assuming each point has 2D coordinates, modify the relational model by integrating coordinates for each point.

Exercise 2

Let the map C of Exercise 1 be.

- Give the simple sequential structure of this map.
- What is the main drawback of this approach ?
- Provide a solution to overcome this drawback.

Solution

Exercise 1

The hierarchical model of map C is shown in Figure 2.28.

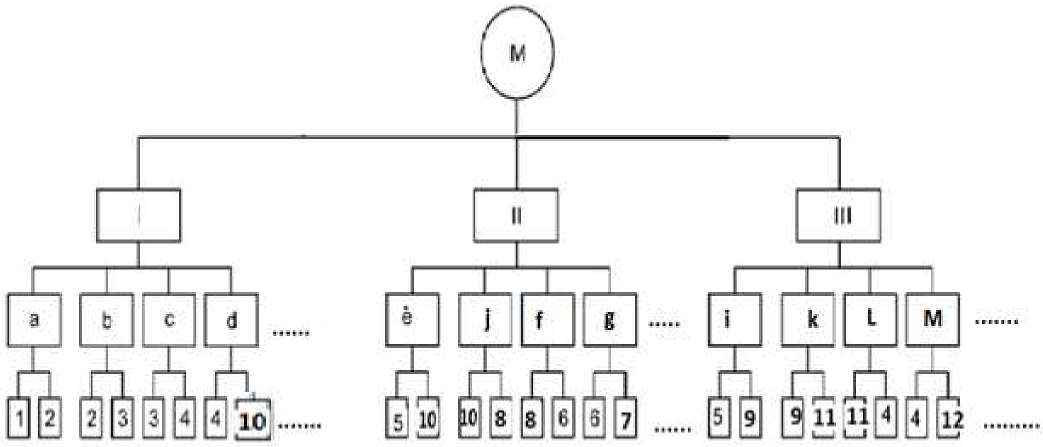


FIG. 2.28 – Hierarchical model of map C

The network model of map C is shown in Figure 2.29.

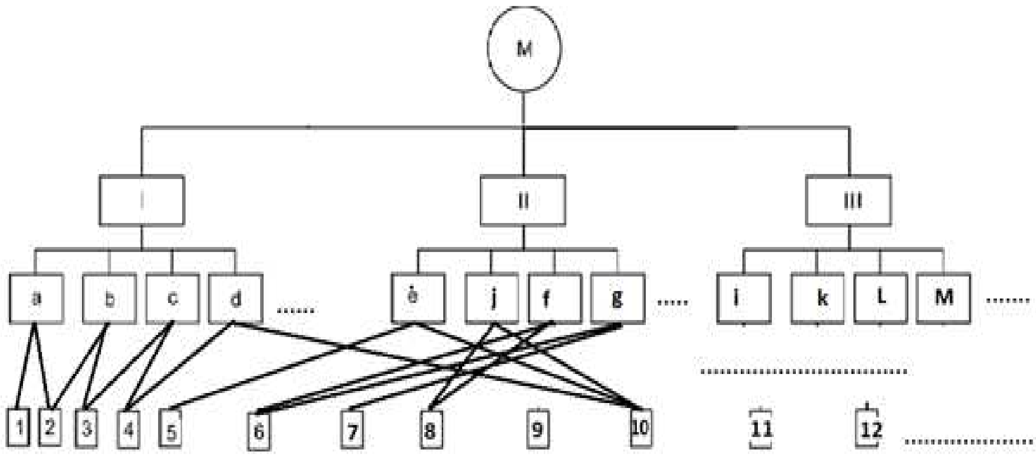


FIG. 2.29 – Network model of map C

The relational model consists of three tables :

1. table-zone(id-zone, id-line) ;
2. table-line(id-line, point-1, point-2) ;
3. table-point(id-point, X, Y) ;

Exercise 2

The simple sequential structure of the map is shown below.

X	Y
X1	Y1
X2	Y2
X3	Y3
X4	Y4
X11	Y11
X9	Y9
X7	Y7
X8	Y8
X10	Y10
X1	Y1
-1	-1
X5	Y5

X9	Y9
X7	Y7
X6	Y6
X8	Y8
X10	Y10
X5	Y5
-1	-1
X5	Y5
X12	Y12
X4	Y4
X11	Y11
X9	Y9
X5	Y5
-1	-1

FIG. 2.30 – Simple sequential structure of map C

The main drawback is data redundancy. The solution is to use topological structures such as POLYVRT.

Exercise 3

Let the vector map C of the city of Chlef be decomposed into 05 zones.

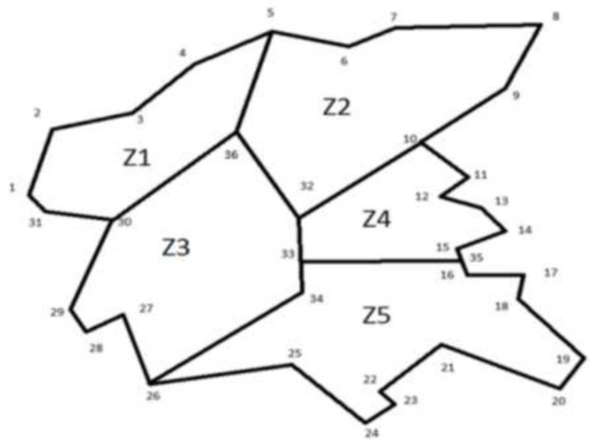


FIG. 2.31 – Map C

1. Give the DIME structure of map C including all required files.
2. Give the POLYVRT structure of map C including all required files.

Exercise 4

1. Give triangle-based and vertex-based description of map **??**.
2. How are triangles constructed ?

Solution of Exercise 3

1. DIME files of map C.

Etiquète Nœud	X	Y
N1	X1	Y1
N2	X2	Y2
N3	X3	Y3
…………….	……………………	…………..
…………..	…………………..	…………..
N36	X 36	Y36

Fig. 2.32 – Node file of map C

Etiquète arrête	Neoud de départ	Nœud de distinati	Zone gauche	Zone droite
N(1-2)	N1	N2	-	Z1
N(2-3)	N2	N3	-	Z1
N(3-4)	N3	N4	-	Z1
……….	………….	…………..	……………	…….
N(36-30)	N30	N36	Z1	Z3
………….	…………….	…………….	……………….	……………
N(26-34)	N26	N34	Z3	Z5

Fig. 2.33 – Edge file of map C

2. POLYVRT files of map C.

Etiquète_zone	Liste des pointeurs de chaines
Z1	C(30 ,5)C(5,36)C(36,30)
Z2	C(5,10)C(10,32)C(32,36)
Z3	C(36,30) C(32,36) C(32,33) C(33,26) C(26,30)
Z4	C(32,33) C(33,35) C(35,26) C(35,10) C(35,10)
Z5	C(35,26) C(33,26) C(33,35)

Fig. 2.34 – Zone file of map C

Etiquète_Nœud	X	Y
N30	X30	Y30
N5	X5	Y5
N36	X36	Y36
N10	X10	Y10
N32	X32	Y32
N33	X33	Y33
N26	X26	Y26
N35	X35	Y35

Fig. 2.35 – Node file of map C

Etiquète chaine	Points intermédiaires
C(30,5)	X31,Y31,X1,Y1,X2,Y2,X3,Y3,X4,Y4
C(5,10)	X6,Y6,X7,Y7,X8,Y8,X9,Y9
C(10,35)	X11,Y11,X12,Y12,X13,Y13,X14,Y14,X15,Y15
C(35,26)	X16,Y16,X17,Y17,X18,Y18,X19,Y19,X20,Y20, X21,Y21,X22,Y22,X23,Y23,X24,Y24,X225,Y25
C(33,26)	X34,Y34
C(26,30)	X27,Y27,X28,Y28,X29,Y29

Fig. 2.36 – Intermediate point file of map C

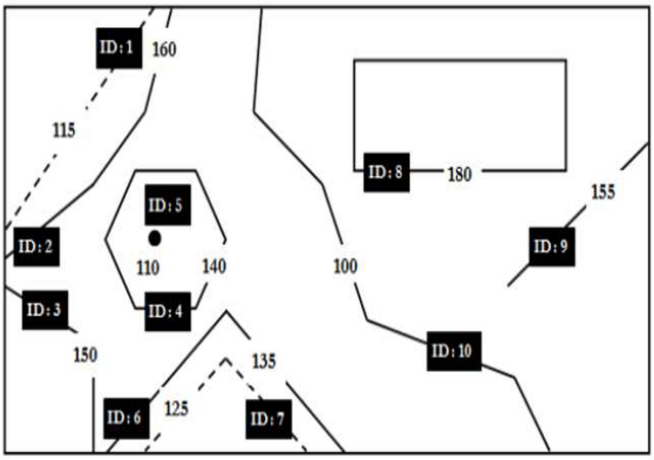
Chaine	Nbre_point_int	Nœud_départ	Nœud_arrivé	Zone_gauche	Zone_droite
C(30,5)	5	N30	N5	-	Z1
C(5,36)	0	N5	N36	Z1	Z2
C(36,30)	0	N36	N30	Z3	Z2
C(5,10)	4	N5	N10	Z2	-
C(10,32)	0	N10	N32	Z3	Z4
C(32,36)	0	N32	N36	Z3	Z2
C(32,33)	0	N32	N33	Z3	Z4
C(33,26)	1	N33	N26	Z3	Z5
C(26,30)	3	N26	N30	-	Z3
C(10,35)	5	N10	N35	Z4	-
C(33,35)	0	N33	N35	Z4	Z5
C(35,26)	10	N35	N26	Z5	-

Fig. 2.37 – Chain file of map C

Elevation structures exercise

Exercise 4

Let the following map be in vector mode.
- Give a complete description of the topological structure files for contour lines.



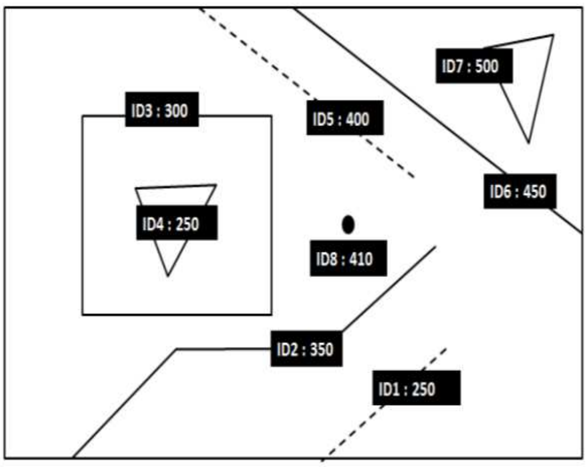
ID	Longueur (m)
1	3780
2	4521
3	2805
4	4320
5	0
6	5106
7	3860
8	9870
9	2087
10	10322

-Extrait d'une carte topographique 1/50.000 (équidistance : 10 m)

FIG. 2.38 – Contour map C-3

Exercise 5

Let the following map be in vector mode.



ID	Longueur (m)
1	1520
2	4570
3	6280
4	1930
5	1870
6	2310
7	1930
8	0

- Extrait d'une carte topographique 1/50.000 (équidistance : 20 m) –

FIG. 2.39 – Contour map C-4

- Give a complete description of the topological structure files for contour lines.

Solution of Exercise 4

The POLYVRT structure files of map C-3 are as follows.

ID	ALT	NP	APD	LON	FOR	CON	NVI	AVI	NVS	AVS
1	115	2	1	3580	O	I	----	-----	1	1
2	160	4	3	4521	O	C	5	1	--	------
3	150	3	7	2805	O	C	3	2	1	1
4	140	6	10	4320	F	C	3	3	2	2
5	110	1	16	0	P	P	----	-----	1	3
6	135	3	17	5106	O	C	2	4	3	4
7	125	3	20	3860	0	I	-----	--------	1	5
8	180	4	23	9870	F	C	2	5	-------	-----
9	155	2	27	2087	O	C	1	6	1	6
10	100	6	29	10321	O	C	--------	--------	6	7

FIG. 2.40 – Contour map C-2

Adresse	Liste
1	1, 3, 4, 6, 10
2	4, 6, 10
3	6, 10, 5
4	7, 10
5	10, 9
6	10

(a)

Adresse	Liste
1	2
2	2, 3
3	4
4	3, 4, 2
5	6
6	8
7	8, 9, 2, 3, 4, 6

(b)

Adresse	X	Y
1	X1	Y1
.	.	.
.	.	.
.	.	.
34	X34	Y34

(c)

FIG. 2.41 – Three files of the contour structure. (a) Lower neighbors file. (b) Upper neighbors file. (c) Point file.

The same steps are used to solve Exercise 5.

2.8 The Route Table

The route table stores the roads of the country. Indeed, all Algerian roads are stored in this table. This table consists of 5 attributes.

1. Id : this is the road identifier. National roads start with "N", regional roads with "W", and communal roads with "C".
2. Long : this is the length of the road in kilometers.
3. Type : there are three types of roads. National roads cross at least three wilayas. Wilaya roads cross at least three communes within a wilaya, and communal roads are located within the same commune.
4. Traffic : it represents the average number of vehicles circulating per unit distance per unit time.
5. Obj : a road is geometrically represented as a polyline. Therefore, the Obj attribute points to a polyline structure.

2.9 The Building Table

The Building table contains information about buildings. In this example, all buildings in the country are stored in this table. The attributes of the Building table are as follows :

1. Id : corresponds to the building identifier.
2. Type : represents the type of building (school, parking, etc.).
3. Area : the surface area of the building.
4. Obj : pointer to the topological structure of a building. A building can correspond either to a zone or a point.

2.10 The Wilaya Table

The Wilaya table represents all wilayas of the country.

1. Id : the wilaya identifier.
2. Name : the name of the wilaya.
3. Area : corresponds to the surface area of the wilaya.
4. Density : number of inhabitants per square kilometer.
5. Number_of_Daira : number of districts (daïras) in the wilaya.
6. Obj : pointer to the zone corresponding to the wilaya. In this case, the zone represents a polygon (aggregation of the polygons of communes belonging to the wilaya).

2.11 Attribute Queries

Attribute queries correspond to standard DBMS queries. Semantic relationships between tables and standard operations are used to filter accurate responses to queries. For example :

1. What is the length of National Road N4 ?
   Query : SELECT Long FROM routes WHERE id='N14' ;
2. What is the average traffic of national roads ?
   Query : SELECT AVG(traffic) FROM Routes WHERE type = "national" ;
3. What is the most populated commune in the Wilaya of Chlef ?
   Query : SELECT Name FROM communes WHERE Popu IN (SELECT MAX(Pop) FROM communes WHERE wilaya = '02') AND wilaya = '02' ;
4. What is the average population of communes in each wilaya of Algeria ?
   SELECT Wilaya, AVG(Pop) FROM Communes GROUP BY Wilaya ;

2.12 Spatial Queries

When there is no relational link between entities, spatial functions are used to answer queries. These are called spatial queries. Spatial queries are often more computationally expensive. Therefore, it is sometimes preferable to store certain spatial information instead of recomputing it each time. For example, storing the surface area of zones is better than recalculating it repeatedly.

Examples of spatial queries include :

1. Which roads intersect with National Road "N4" ?
   Query : SELECT Id FROM Route1 AS R1, Route2 AS R2 WHERE R1.Id = "N4" AND R2.Id <> "N4" AND IntersectLineLine(R1.Obj, R2.Obj) ;
2. Which commune contains a bank ?
   Query : SELECT Commune.Name FROM Commune, Building WHERE Contains(Commune.O Building.Obj) AND Building.Type = "bank" ;
3. Which communes border the Wilaya of Chlef ?
   Query : SELECT Commune FROM Commune, Wilaya1 AS W1, Wilaya2 AS W2 WHERE W1.Name = "Chlef" AND ZoneInZone(W1.Obj, Commune.Obj) AND Adjacent(W2.Obj, Commune.Obj) AND W1.Id <> W2.Id ;
4. Which universities have both a library and a parking area, where the distance between them is less than 100 meters ?
   Query : SELECT B1.Id FROM Building B1, Building B2, Building B3 WHERE B1.Type = "library" AND B2.Type = "parking" AND B3.Type = "university" AND ZoneInZone(B1.Obj, B3.Obj) AND ZoneInZone(B2.Obj, B3.Obj) AND Distance(B1.Obj, B2.Obj) < 100 ;

2.13 Conclusion

Spatial systems involve two modes : raster mode and vector mode. Vector data modeling is based on graph theory and geometry, whereas raster mode is entirely based on image processing. In this chapter, an introduction to the coupling between attribute data and spatial data has been presented. Although the third development methodology of geographic systems is based on relatively new concepts, the use of geometric and image processing principles remains essential.

Exercise

Given the following SIG database tables :

id	Type	Area	Obj
17	University	800	*
18	Library	400	*

FIG. 2.42 – Building Table

id	Length	Type	Traffic	Obj
N14	180	National	4500	*

FIG. 2.43 – Route Table

id	Name	Population	Obj
6	Ouled Fares	500000	*
7	Aougrout	100000	*

FIG. 2.44 – District Table

Name	Area	Density	Number_of_Daira	Obj
Chlef	4791	209	13	*
Laghouat	25057	18	10	*

FIG. 2.45 – Wilaya Table

1. What is the purpose of the **Obj** column ?
2. Write the following queries (extended SQL) and classify them as attribute or spatial queries :
   - Which districts belong to the Wilaya of Chlef ?
   - What is the total population of the Wilaya of Chlef ?
   - How many districts are there in Algeria ?
   - How many national roads cross the Wilaya of Chlef ?
   - What is the length of National Road 14 ?
   - What is the population of the Wilaya of Chlef ?
   - Which wilayas have a university ?
   - Which districts belong to the Wilaya of Chlef ?

Solution

1. The Obj column is a pointer to functions executed by the spatial part of the system, which cannot be handled by a traditional DBMS.
2. SQL3 queries :

   (a) Districts of Chlef (spatial query) :

   SELECT District.Name

   FROM District, Wilaya

   WHERE Wilaya.Name = "Chlef" AND ZoneInZone(Wilaya.Obj, District.Obj) ;

   (b) Total population of Chlef (spatial query) :

   SELECT SUM(District.Population)

   FROM District, Wilaya

   WHERE Wilaya.Name = "Chlef" AND ZoneInZone(Wilaya.Obj, District.Obj) ;

(c) Total number of districts in Algeria (attribute query) :

SELECT SUM(Wilaya.Number_of_Daira)

FROM Wilaya ;

(d) National roads crossing Chlef (spatial query) :

SELECT COUNT(Route.Id)

FROM Route, Wilaya

WHERE Route.Type = "National" AND Wilaya.Name = "CHLEF" AND IntersectLineZone(Route.Obj, Wilaya.Obj) ;

(e) Length of National Road 14 (attribute query) :

SELECT Route.Length

FROM Route

WHERE Route.Id = "N14" ;

(f) Population of Chlef (spatial query) :

SELECT SUM(District.Population)

FROM Wilaya, District

WHERE Wilaya.Name = "Chlef" AND ZoneInZone(District.Obj, Wilaya.Obj) ;

(g) Wilayas with a university (spatial query) :

SELECT Wilaya.Name

FROM Wilaya, Building

WHERE Building.Type = "university" AND ZoneInZone(Wilaya.Obj, Building.Obj) ;

La table route stores the roads of the country. Indeed, all the roads of Algeria are stored in this table. This table is composed of 5 attributes.

1. Id : this is the road identifier. The identifier of national roads starts with « N », for wilaya roads with « W »and communal roads with « C ».
2. Long : this is the length of the road in kilometers.
3. Type : there are three types of roads. National roads that cross at least three wilayas. Wilaya roads that cross at least three communes of the wilaya and communal roads that are located within the same commune. Trafic : it corresponds to the average number of vehicles circulating per unit of distance per unit of time.
4. Obj : a road is geometrically represented as a polyline. Therefore, the obj attribute points to a polyline.

The Road table.

The Building table is associated with buildings. In this example, all buildings of the country are stored in this table. The attributes composing the Building table are as follows.

1. Id : it corresponds to the building identifier.
2. Type : it represents the type of building (school, parking, etc.).

3. Superficie : this is the surface area of the building.
4. Obj : this is the pointer to the topological structure of a building. A building can correspond to a zone or a point.

The Building table.

The Wilaya table represents all the wilayas of the country.

1. Id : this is the wilaya identifier.
2. Nom : this is the name of the wilaya.
3. Superficie : corresponds to the area of the wilaya.
4. Densité : the number of inhabitants per km^2.
5. Nombre-daira : the number of dairas of the wilaya.
6. Obj : the pointer to the zone corresponding to the wilaya. In this case, the zone represents a polygon (aggregation of polygons corresponding to the communes of the wilaya).

The Wilaya table.

2.13.1 Attribute queries

Attribute queries concern standard DBMS queries. Semantic relationships between tables and standard operations are used to filter exact query results. For example

1. What is the length of the national road N4 ?
   Query : SELECT Long FROM routes WHERE id='N14' ;

2. What is the average traffic of national roads ? Query : SELECT AVG(trafic) FROM Routes WHERE type = " national".
3. What is the most populated commune in the Wilaya of Chlef ?
   Query : SELECT Nom FROM communes WHERE Popu IN (SELECT Max(Pop) FROM communes WHERE ville = '02') AND ville= '02' ;
4. What is the average population of communes in each Wilaya of Algeria ?
   SELECT Ville, AVG(Pop) FROM Communes WHERE GROUP BY ville ;

2.13.2 Spatial queries

When there is no relational link between entities, spatial functions are invoked to answer queries. In this case, these are called spatial queries. Often, spatial queries are more expensive in terms of computation time. Therefore, it is sometimes preferable to store certain spatial information instead of computing it (via software processing). For example, it is preferable to store zone areas instead of recalculating them each time. The following queries are examples of spatial queries.

1. Which roads intersect the national road « N4 » ?
   Query : SELECT Id FROM Route1 AS Route, Route2 As Route WHERE Route1.Id= "N4" AND Route2.Id <> "N4" AND IntersectLineLine (Route1.obj , Route2.obj)

If the boolean intersection function does not include the line with itself, we can remove the condition « Route2 <> "N4" ».

2. Which commune has a bank ?
Query : SELECT Commune.Nom FROM Commune , Building WHERE Contains (Commune.obj , Building.obj) and Building.type = "bank"

3. Which communes are on the border of the Wilaya of Chlef ?
Query : SELECT Commune FROM Commune, Wilaya1 AS Wilaya, Wilaya2 AS Wilaya WHERE Wilayas1.nom = "Chlef" AND ZoneInZone (Wilaya1.obj , commune.obj) AND Adjacent(wilaya2.Obj , Commune.Obj) AND Wilaya1.Id <> Wilaya2.Id

4. Which universities have a library and a parking such that the parking is within 100 meters of the library ?
Query : SELECT Building.Id FROM Building1 AS Building, Building2 AS Building, Building3 AS Building WHERE Building1.type = "library" AND Building2.type = "parking" AND Building3.type = "university" AND ZoneInZone(Building1.Obj , Building3.Obj) and ZoneInZone(Building2.Obj , Building3.Obj) and Distance(Building1.Obj , Building2.Obj) < 100.

2.14 Conclusion

Spatial systems involve both modes, raster mode and vector mode. The modeling of geographic data in vector mode is based on graph theory and geometry, while raster mode is entirely based on image processing. In this chapter, an introduction to the coupling of attribute and spatial data is presented. Although the third category of geographic system development methodology is based on concepts considered new, the use of geometry and image-related concepts remains essential.

Exercise

Given the following tables of a GIS database.

id	Type	Superficie	Obj
17	University	800	*
18	Library	400	*

FIG. 2.46 – Building Table

id	Long	Type	Trafic	Obj
N14	180	National	4500	*

FIG. 2.47 – Road Table

id	Name	Popu	Obj
6	Ouled fares	500 000	*
7	Aougrout	100 00A0	*

FIG. 2.48 – Daira Table

Nom	Superficie	Densite	Nombre_Daira	Obj
Chlef	4 791	209	13	*
Laghouat	25 057	18	10	*

FIG. 2.49 – Wilaya Table

1. What is the purpose of the **Obj** column ?
2. Write the following queries (Extended SQL) and state whether they are attribute or spatial queries :
   - What are the Dairas of the Wilaya of Chlef ?
   - What is the total population of the Wilaya of Chlef ?
   - How many Dairas are there in Algeria ?
   - How many national roads cross the Wilaya of Chlef ?
   - What is the length of National Road 14 ?
   - What is the population of the Wilaya of Chlef ?
   - Which Wilayas have a university ?
   - What are the Dairas of the Wilaya of Chlef ?

Solution

1. The object column is a pointer to functions executable by the spatial part of the system and cannot be executed by a traditional DBMS.
2. SQL3 queries
   (a) The Dairas of the Wilaya of Chlef is a spatial query.
   SELECT Daira.Nom
   FROM Daira, Wilaya
   WHERE Wilaya.Nom = "Chlef" AND ZoneInZone(Wilaya.Obj, Daira.Obj) ;
   (b) Population of the Wilaya of Chlef is a spatial query.
   SELECT SUM (Daira.Population)
   FROM Daira, Wilaya
   WHERE Wilaya.Nom = "Chlef" AND ZoneInZone(Wilaya.Obj, Daira.Obj) ;
   (c) Total number of Dairas in Algeria is an attribute query.
   SELECT SUM (Wilaya.Nombre_Daira)
   FROM Wilaya
   (d) National roads crossing the Wilaya of Chlef is an attribute query.
   SELECT COUNT(route.id)
   FROM Route, Wilaya
   WHERE Route.type = "National" AND Wilaya.Nom = "CHLEF" AND IntersectLineZone(Route.Obj, Wilaya.obj) ;
   (e) Length of National Road 14 is an attribute query.
   SELECT Route.Long
   FROM Route
   WHERE Route.id = "N14" ;
   (f) Population of the Wilaya of Chlef is a spatial query.
   SELECT SUM(Daira.Popu)
   FROM Wilaya, Daira
   WHERE Wilaya.Nom = "Chlef" AND ZoneInZone(Daira.obj, Wilaya.Obj) ;
   (g) Wilayas having a university is a spatial query.
   SELECT Wilaya.Nom
   FROM Wilaya, Building
   WHERE Building.Type = "university" AND ZoneINzone(Wilaya.Obj, Building.Obj) ;

General Conclusion

Geographic information systems play a very important role in the development of companies and in the daily life of individuals. Therefore, they are increasingly used by the general public. To meet the growing needs of users, GIS systems continue to evolve. Consequently, developers are faced with new challenges. Among these challenges, we mention :

1. the heterogeneity of human and material resources : GIS are used by land-use engineers as well as by freight transporters in their vehicles. On the other hand, a GIS or an access terminal to geographic information is currently installed on mobile devices, tablets, vehicles, airplanes, etc.
2. the objectives for which GIS are designed : as indicated in the first chapter, the objectives of a GIS are numerous and varied. A GIS can be used to derive statistics on urban and economic development of a territory, as well as to simulate the location of a road interchange. In other words, several disciplines are involved in the development of GIS.
3. different models of storage, presentation, analysis, and implementation : GIS are considered advanced systems. Therefore, the design, implementation, and processing models of geographic information are diverse. This diversity requires additional costs and efforts.

Consequently, in this document we have focused on the structuring of geographic information and on the vector representation mode, while also providing overviews of the raster mode and the use of other disciplines in GIS.

Bibliographie

[1] S. Aronoff. Ottawa, WDL Publications, 1989.

[2] R. Brunet, R. Ferras, and H. Théry. *Les mots de la géographie : dictionnaire critique. Troisième édition.* Reclus, 2009.

[3] L. M. Bugayevskiy and J. P. Snyder. *Map projections : a reference manual.* London : Taylor Francis, 1995.

[4] P. A. Burrough and R. A. Mcdonnell. *Principles of Geographical Information Systems.* New York : Oxford University Press, 1998.

[5] P. Champoux and Y. Bédard. Notions fondamentales danalyse spatiale et dopérateurs spatiaux. *Revue des sciences de linformation géographique et de lanalyse spatiale*, (2) :187–208, 1992.

[6] J. Charre. *Statistique et territoire.* Reclus, 1995.

[7] R. J. Clark. Habiline handexes and paranthropine pedigree at sterkfontein. *World archiology*, 1988.

[8] M. Didier. *Utilité et valeur de linformation géographique.* Economica. Paris, 1990.

[9] ESRI. Tout savoir sur les systèmes dinformation géographique. Technical report, Éditeur de logiciel de système d'information géographique (SIG), https ://www.esrifrance.fr/ (Accessed : 5 Mars 2018), 2018.

[10] B. Huang. *Comprehensive Geographic Information Systems.* Elsivier, 2018.

[11] G. Konecny. *Geoinformation : remote sensing, photogrammetry, and geographic information systems. Deuxième édition.* Taylor Francis, 2014.

[12] K. Merry, P. Bettinger, M. Crosby, and K. Boston. *Geographic Information System Skills for Foresters and Natural Resource Managers.* Elsivier, 2023.

[13] D. Pumain and T. Saint-Julien. *Analyse spatiale : Les localisations. Deuxième édition.* Armand Colin, 2010.

[14] Y. Yamagata and H. Seya. *Spatial Analysis Using Big Data Methods and Urban Applications.* Elsivier, 2020.

[15] Q. Zhou, B. Less, and G. Tang. *Advances in Digital Terrain Analysis.* Springer, 08.