UNIVERSITÀ DEGLI STUDI DELLA Tuscia

GIS
Geographic Information System

Andrea Petroselli

Tuscia University, Italy
petro@unitus.it

SUMMARY

- ✓ Part 1: What is a GIS?
- ✓ Part 2: Why do we need a GIS?
- ✓ Part 3: Which are the possibilities of a GIS?
- ✓ Part 4: Practical example

Part 1: What is a GIS?

There are many definitions for a

Geographic Information System

mainly it is a computer based technology which provides the tools for collecting, editing, storing, retrieving, analyzing and displaying spatial data.

A GIS can create new information from existing data. This thing is what distinguishes a GIS from a CAD (Computer Aided Design).
These characteristics makes a GIS useful for a wide variety of applications.

Which are the Physical Components of a GIS?

1) Computer (hardware): 'one' or 'many': 'one' is a *single* position, 'many' is a *computer network*

1) GIS software (the programs that run within the computer); Environmental Systems Research Institute (ESRI) is the world's largest supplier of GIS software. ESRI produces and sells different GIS sofware, for example: **ArcMap, ArcInfo, ArcView**

3) Georeferenced ***Data*** or information

4) GIS Analyst (people who use GIS)

What is Data?

Data is the core of a GIS, because it is what GIS analyzes

'Data' is a collection of observations and measurements usually related by a common theme (e.g.: Temperatures, Elevations, Land Cover, Population, Prices, and so on…)

Data by itself has no particular value, but when we associate Data with *Geographically Referenced information* (spatial coordinates that describe the geographic location of the data) we obtain the **Spatial Data**, that is the core of GIS.

The main characteristic of these Spatial Data is the fact that they have some **Attributes**

What are Attributes?

Attributes are *characteristics* and *information* about spatial features.

Information, built into a larger context, constructs knowledge.

Examples include geographic names, management information, etc.

Generally stored on the computer in a relational

DataBase Management System (DBMS)

that is part of the GIS software.
So, another difference between GIS and CAD is the existence (and linking) of the Data with its Attributes that form the DataBase

These Data Attributes
often are the ways to describe earth surface features.

There are two descriptor choices:

1. **Features are entities**. Described by their attributes or properties.
Position is mapped via a geographic coordinate system.
E.g.: Land use/cover map, Population density
They are: ***Discrete Data***

2. **Feature is a continuous surface or field**. Variation occurs across time and/or space.
E.g.: Slopes, Lake surface temperatures, Elevations
They are: ***Continuous Data***

The important thing is that all features (discrete or continuous) have a

Spatial Reference System

Spatial reference system is critical for a GIS to be operational and fully functional. Otherwise, the GIS software is nothing else than an expensive map display package.

Spatial reference is a mechanism to situate measurements on a geometric body, such as the earth

The **spatial reference system** (or coordinate system) provides a measure of *absolute location* as opposed to *relative location* (street address).

We can have several types
of Spatial Information and Spatial Data

Point
Line → (Vector Data or *<u>Shapefile</u>*)
Polygon

<u>Point</u>: just a location on the earth's surface. No width, no perimeter, neither area.

E.g: the position on trees within a field

<u>Line</u>: it is a linear feature in which width is not important, length is measured and important.

E.g.: roads and rivers

<u>Polygon</u>: it is a feature in which also area and its associated measurements are important.

E.g.: whole fields, countries

The problem we have with vector formats (points, lines, polygons) is that they represent very well features like points, lines or polygons, but they do lack capability (not for us, but for the computer memory and/or its way to think and compute) to represent *continuous* surfaces o features; so when we have to represent and analyze continuous surfaces by computers we deal mainly with **Raster Data** (or **GRID** data)

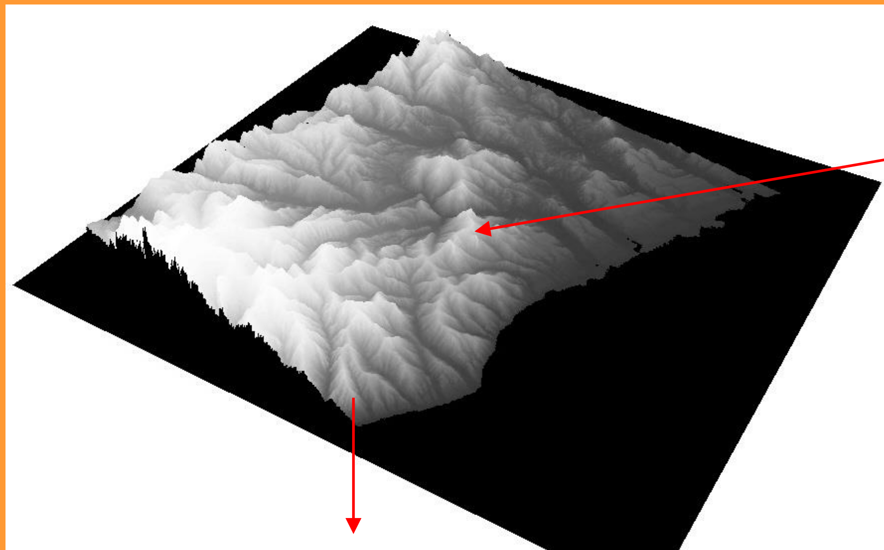
Raster Data is a gridded (simplified) representation of a continuous feature:

The Grid format is made up of individual cells (**pixels**) where area is defined, in opposition with Vector format that is composed of nodes and arcs

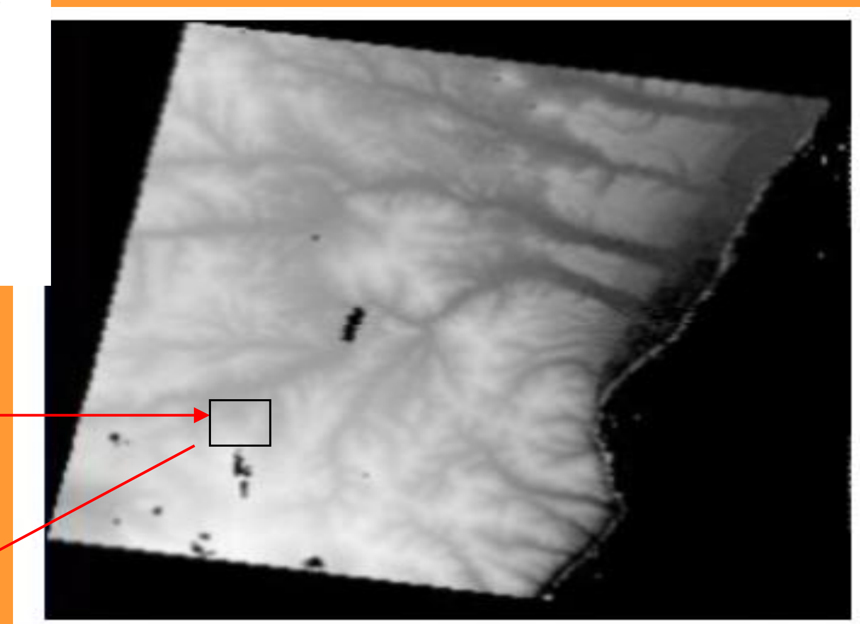
Each pixel in Raster Data or node in Vector Data is associated with a X, Y and possibly a Z coordinate (where Z does not mean always elevation…)

E.g.: Elevation, Slopes, Aspects…

Example of Raster Data:



The tridimensional real feature (elevation)



From the tridimensional
we pass in the bidimensional
representation

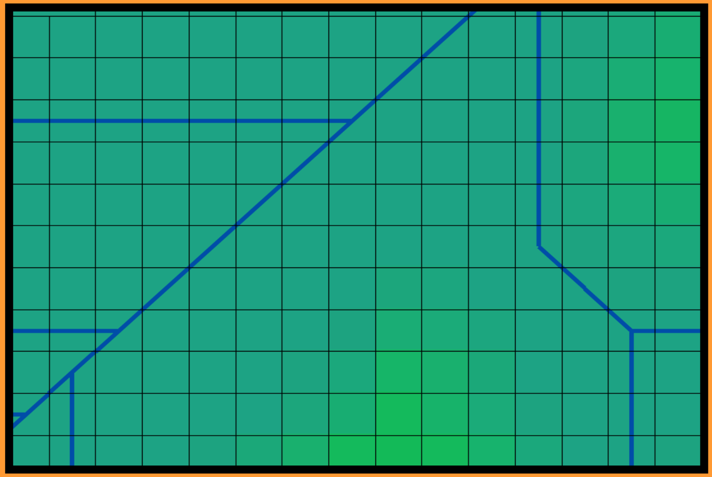
75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	77.25	82.84	86.45
75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	76.17	79.62	82.46
75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.30	76.82	77.74
75.00	75.00	75.00	75.08	77.25	76.85	75.00	75.00	75.00	75.41	75.48
75.00	75.00	75.00	76.03	81.54	80.12	76.36	75.00	75.00	75.00	75.00
75.00	75.00	75.39	78.44	86.47	83.72	78.90	75.11	75.00	75.00	75.00
75.02	75.08	77.07	82.54	89.59	85.79	79.66	75.65	75.00	75.00	75.00
75.73	78.94	83.60	89.43	90.73	89.45	84.54	77.74	75.00	75.00	75.41

The surface seems still continuous, but if we zoom in we see the gridded structure where every pixel (cell) has a X, Y and Z coordinate (referred to the center of the cell)

Example of interaction between Raster and Vector Data:

75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	77.25	82.84	86.45
75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	76.17	79.62	82.46
75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.30	76.82	77.74
75.00	75.00	75.00	75.08	77.25	76.85	75.00	75.00	75.00	75.41	75.48
75.00	75.00	75.00	76.03	81.54	80.12	76.36	75.00	75.00	75.00	75.00
75.00	75.00	75.39	78.44	86.47	83.72	78.90	75.11	75.00	75.00	75.00
75.02	75.08	77.07	82.54	89.59	85.79	79.66	75.65	75.00	75.00	75.00
75.73	78.94	83.60	89.43	90.73	89.45	84.54	77.74	75.00	75.00	75.41

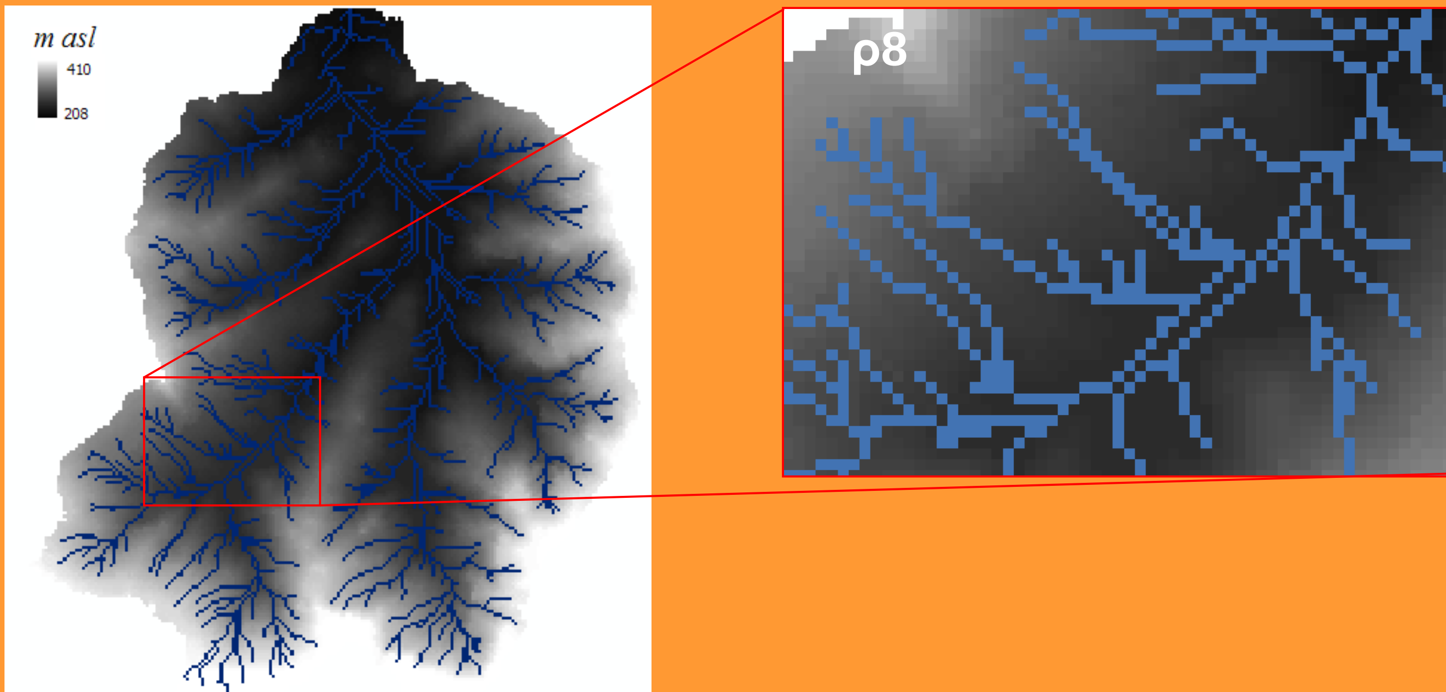
Elevation grid
(Raster Data)



Stream network
(Vector Data)

Using the concept that water follows the steepest descent direction, starting from Grid Elevation, GIS Software can recognize where every cell is headed (comparing for every cell the 8 surrounding slopes and choosing as the flow direction that one with the greatest slope) and reconstruct the whole stream network (represented finally as a vector line)

Why we choose vector data if it is possible



A representation of a linear feature with Raster Data is ugly to see, because if we zoom in we see a gridded scheme that is unrealistic, but in case of large territory is the only way to represent the feature

Part 1 Conclusions:

A GIS is a mix of hardware, software and data that the analyst can use to manipulate territory

The computers are the hardware,
The programs (like the ones developed by ESRI) are the software;
using hardware and software analyst can manage the data (Raster and Vector) and extract information from them.

But what information?!

Part 2: Why do we need a GIS?

As said before in the definition, GIS is only a tool that the analyst can use to:

collecting, **editing**, **storing**, **retrieving**, **analyzing** **and** **displaying** **spatial data**

So, first of all, we must have (or create…) spatial data…

Once we have spatial data, GIS provides analyst whatever operation has to be performed on these data…

So, actually, we need GIS to do many things, but basically to

MANAGE OUR ENVIRONMENT AND TAKE DECISIONS

Why GIS and why not without GIS?

Geography has always been important to humans.

Applied geography, in the form of ***maps*** and spatial information, has served for discovery, planning, commerce, and so on.

Because spatial information is so important, have been developed tools called Geographic Information Systems (GIS) to help us deepen our geographic knowledge.

Because GIS is only a tool, each GIS user decides what features are important, and what is important about them.

For example,

Forests are important to many people.

They protect our water supplies, yield wood, harbor wildlife, and provide space to recreate.

We are concerned about the level of harvest, the adjacent land use, pollution from nearby industries, or when and where forests burn.

Informed management of our forests requires at a minimum knowledge of all these related factors, and perhaps above all the **spatial arrangement** of these factors.

Buffer strips near rivers may protect water supplies, ***clearings*** may prevent the spread of fire, and ***polluters downwind*** may not harm our forests while polluters upwind might.

A GIS aids immensely in the analysis of these spatial relationships and interactions among them.

A GIS is also particularly useful at displaying spatial data and reporting the results of spatial analysis.

In many instances GIS is the only way to solve spatially-related problems.

GIS use has become widespread during the past two decades.

GIS have been used in fields from archeology to zoology, and new applications of GIS are continuously emerging.

GIS are essential tools in business, government, education, and nonprofit organizations.

GIS have been used to fight crime, protect endangered species, reduce pollution, cope with natural disasters, and to improve public health;

In short, GIS have been instrumental in addressing some of our most
pressing societal problems

Concluding, GIS are needed in part because human population and technology have reached levels such that many resources, including

air, land and water (our environment), are placing substantial limits on human action.

Human populations have doubled in the last 50 years, reaching 6 billion, and we will likely add another 5 billion humans in the next 50 years.

The first 100,000 years of human existence caused few impacts on the World's resources, while in the past 300 years humans have permanently altered most of the Earth's surface.

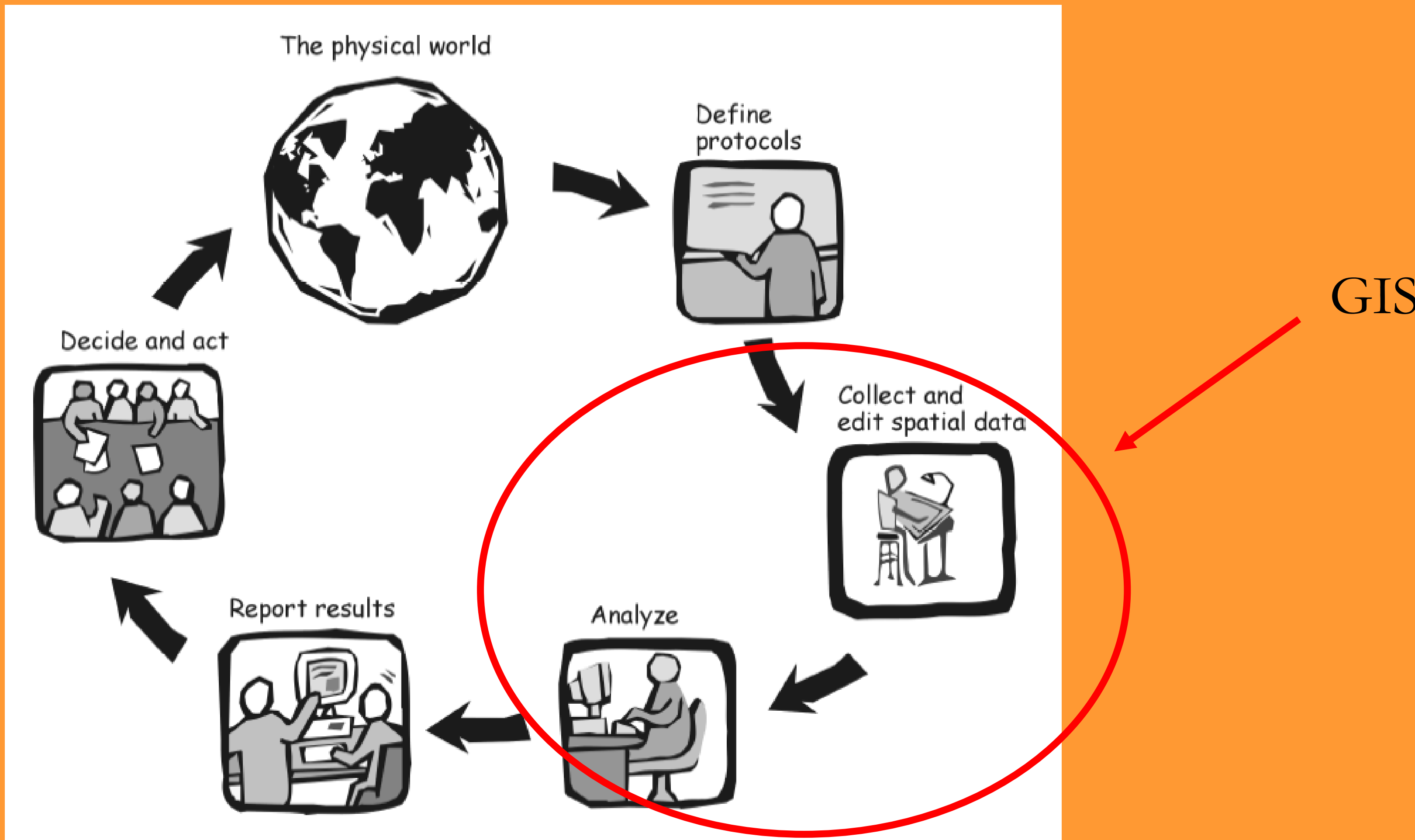
Silt chokes many rivers and there are many localized examples where **pollutants** substantially harm public health.

By the end of the 20th century most suitable lands had been inhabited and only a minority percentage of the terrestrial surface had not been farmed, grazed, cut, built over, drained, flooded, or otherwise altered by humans!

In the end, **spatial data organization and analyses** are widely applied to improve life.

And that is why GIS are widely used in <u>planning and environmental protection</u>.

Part 3: Which are the possibilities of a GIS?



<u>How should be GIS use</u>: effective use of GIS depends on a set of protocols and an integration into the data collection, analysis, decision, and action loop of an organization.

For understanding the GIS possibilities we must go back to the points:

1) each GIS user decides what features are important, and what is important about them

2) a GIS is a mix of hardware, software and data that the analyst can use to manipulate territory and take decisions

So… GIS possibilities are endless, because endless are the fields of interest, and endless are the features to apply GIS to…

For example, these are the functions commonly provided by GIS software.

Data entry
- manual coordinate capture
- attribute capture
- digital coordinate capture
- data import

Editing
- manual point, line and area feature editing
- manual attribute editing
- automated error detection and editing

Data management
- copy, subset, merge data
- versioning
- data registration and projection
- summarization, data reduction
- documentation

Analysis
- spatial query
- attribute query
- interpolation
- connectivity
- proximity and adjacency
- buffering
- terrain analyses
- boundary dissolve
- spatial data overlay
- moving window analyses
- map algebra

Output
- map design and layout
- hardcopy map printing
- digital graphic production
- export format generation
- metadata output
- digital map serving

Because many are the fields of interest, and many are the necessities, we have many kinds of GIS software… from freeware to very expensive…

Each software has its *advantages* and its *drawbacks*

Some software is used at its top to describe Vector Data, others to manipulate and analyze Raster Data…

Some software is used with GPS Survey, other with Remote Satellite Survey,

Some software works better with the production of maps and graphic output and others to make calculation between maps, and so on…

The main GIS software are:

a) **ArcGIS**, developed by ESRI, California.

There are many versions of ArcGIS, from Arcview to ArcInfo:

1) **<u>ArcView</u>:** is an entry-level component of ArcGIS. GIS functions are provided for *basic data entry*, *editing*, and attribute and coordinate manipulation. *Basic spatial data analyses* are supported, and *rapid, easy, basic map layout and printing* capabilities are provided.

2) **<u>ArcInfo</u>** is the comprehensive GIS toolbox from ESRI. It is a large, complex, sophisticated product. It supports multiple data formats, many data types and structures, and literally *thousands* of possible operations that may be applied to spatial data. Substantial training is required to master the full capabilities of ArcInfo

b) MapInfo: it is a comprehensive set of GIS products developed and sold by the Map-Info Corporation, of Troy, New York.
Map-Info products are used in a broad array of endeavors, although use seems to be concentrated in many *business and municipal applications*.

c) Idrisi: it is a GIS system developed by Clark University, in Massachusetts. Idrisi differs from the previously discussed GIS software packages in that it provides both *image processing* (mainly for remotely sensed Digital Elevation Models) and GIS functions.

d) AutoCAD Map: Produced by Autodesk of San Rafael, California, AUTOCAD began as an engineering drawing and printing tool. AUTOCAD MAP adds analytical capability to the already complete set of data input, coordinate manipulation, and data output tools provided by AUTOCAD, but it works better with Vector Data and less well with Raster Data.

e) Manifold: it is a relatively inexpensive GIS package with a surprising number of capabilities. Basic spatial data entry and editing support are provided, as well as basic vector and raster analysis, image display and editing, and output.
The program is extensible through a series of software modules.
Modules are available for surface analysis, business applications, internet map development and serving, database support, and advanced analyses.
Manifold GIS is different from other packages in providing sophisticated image editing capabilities, much like *Adobe Photoshop*,
but in a spatially-reference framework.

By the way, the most used GIS sofware are *Arcview* and *ArcInfo*, and later these two will be explored…

Turning back to possibilities, we have seen that,
due to the large number of GIS software in the world
and due to their use in many fields of interest,
these possibilities are virtually endless…

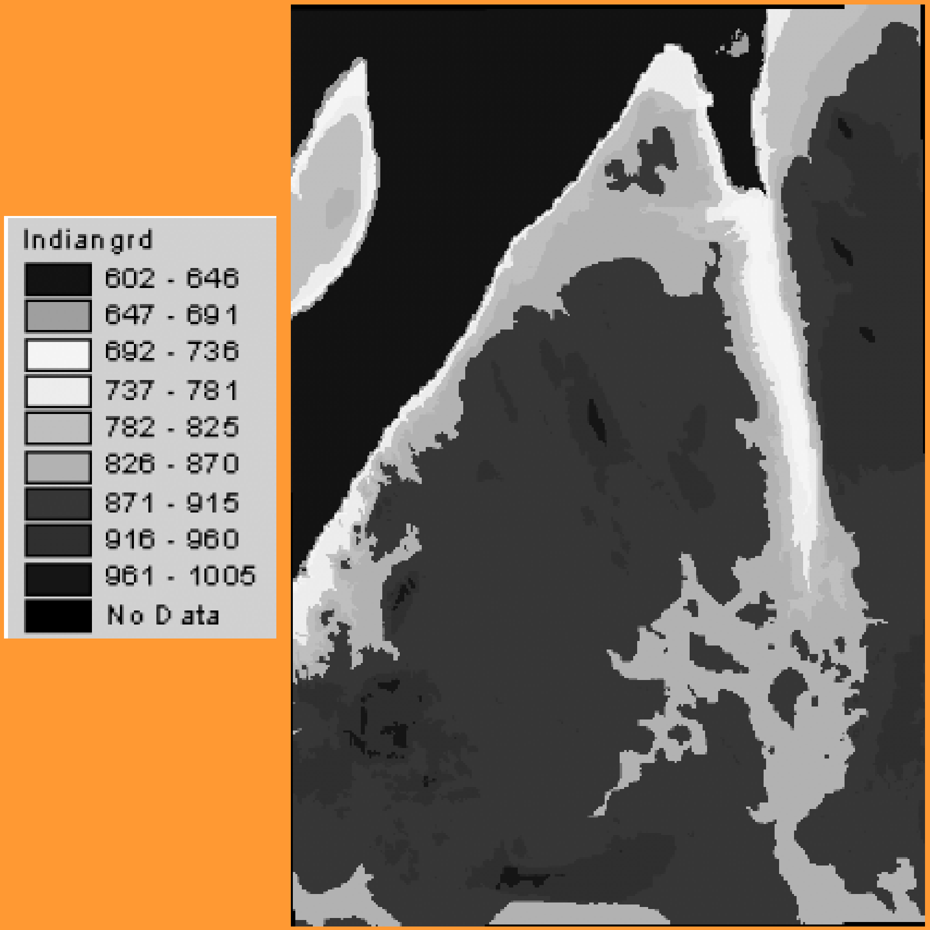
We are interested, by the way, in the ***environment management***,
in particular to what does happen on territory,
and when we talk about land and territory (i.e. a large region),
we are forced to deal with Raster Data.

There is an important (maybe the most important of all) Raster Data
that you can obtain when your attribute is the *Elevation*.
In this case we don't talk about Raster Data but we talk about

DIGITAL ELEVATION MODEL
(DEM)

An example of Digital Elevation Model

Digital Elevation Model Data:
as the others Raster Data, it is a matrix of cells regularly spaced, where the attribute is the elevation datum, usually measured at particular ground positions and later interpolated in order to reconstruct the whole matrix of cells.
This operation of interpolation is done applying several techniques (i.e. Kriging).
DEMs are useful because, from them, we can obtain other interesting data…



Two Types of Data Commonly Derived from DEMs:

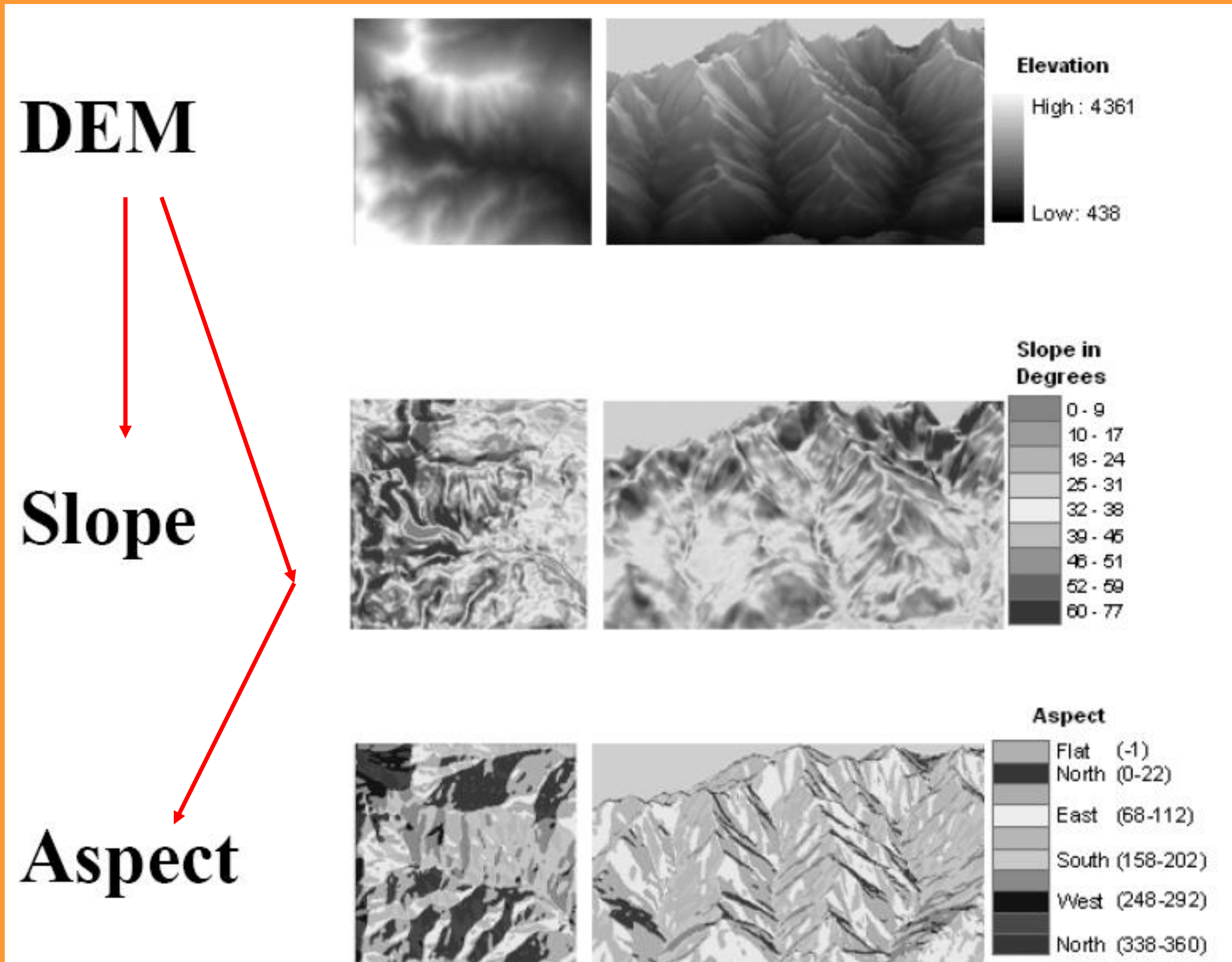
Slope and Aspect

Slope - change in elevation over a predefined distance. With a raster base DEM the predefined distance is usually the pixel size.

Aspect - orientation or direction that a surfaces faces, with respect to the direction of North.

Both maintain the raster format of the DEM. Spatial resolution of the slope and the aspect images cannot be finer than the spatial resolution of the DEM.

Plant Assonometric view



For example, in the case of **Slope**,

ArcGIS calculates slope using a 3 X 3 moving window.
Data is replicated along the edges of the image to create a "pseudo" 3 X 3 window.
The Slope value for the central cell is calculated as follows, starting from the values of the 8 surrounding elevation values:

$$S = \sqrt{[(e_1 + 2e_4 + e_6) - (e_3 + 2e_5 + e_8)]^2 + [(e_6 + 2e_7 + e_8) - (e_1 + 2e_2 + e_3)]^2} / 8d$$

e_1	e_2	e_3
e_4	S	e_5
e_6	e_7	e_8

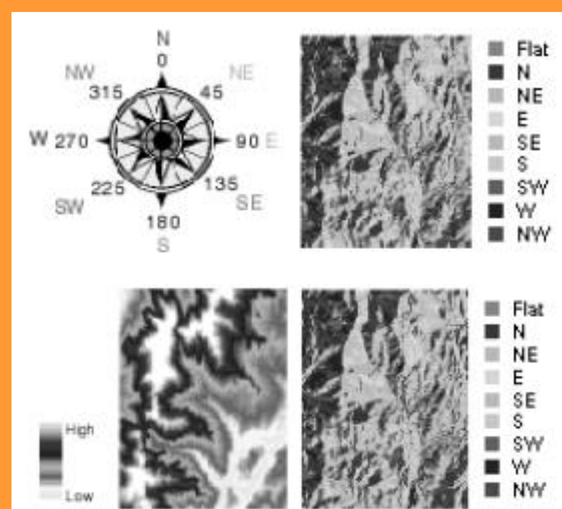
d

instead, in the case of **Aspect**,

The Aspect value for the central cell of the 3 X 3 moving window is calculated as follows, starting always from the values of the 8 surrounding elevation values:

$$D = \arctan([e_6 + 2e_7 + e_8) - (e_1 + 2e_2 + e_3)]/[(e_1 + 2e_4 + e_6) - (e_3 + 2e_5 + e_8)])$$

e_1	e_2	e_3
e_4	D	e_5
e_6	e_7	e_8



But Slope and Aspect are not the only features we can derive from a DEM:

Applications are numerous:

calculate solar insolation;
watershed determination;
viewshed;
construction (roads/harvest operations);
soil erosion (Universal Soil Loss Equation);
contour lines;
hillshade…

Example: Contour Lines

Polylines (Vector Data) that connect points (elevations) of equal value. Their distribution shows how elevations change across a surface and it is an example of interaction between Raster and Vector Data.

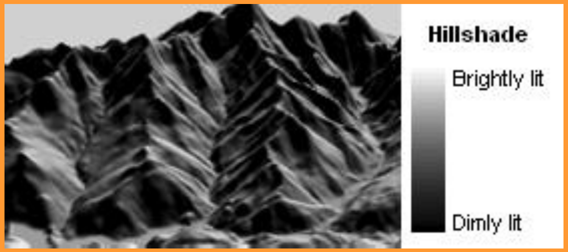
Close- much change; Far apart- little change

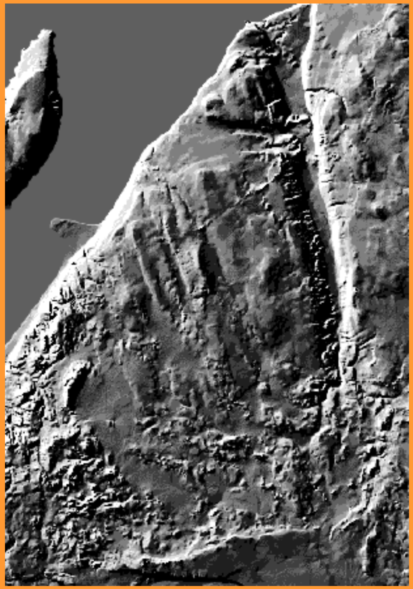
Example: Hillshade

Shows the hypothetical illumination of a surface by determining illumination values for each cell in a raster which is the same file size and spatial resolution of the DEM.

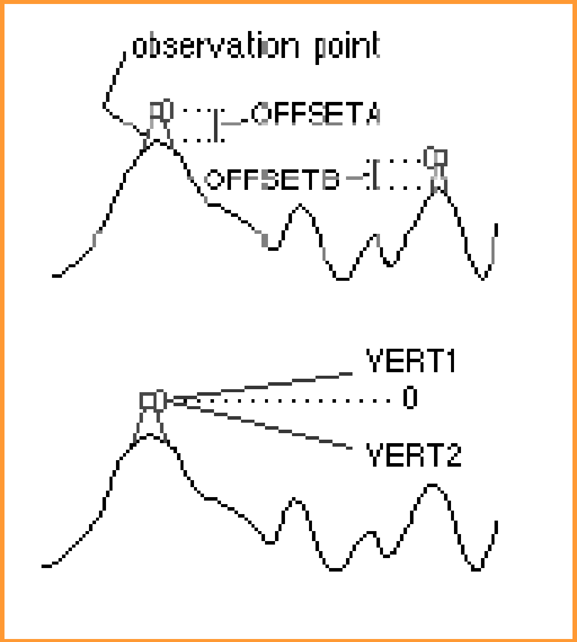
A position is determined for the *light source* and illumination values are calculated for each cell in relation to neighboring cells.

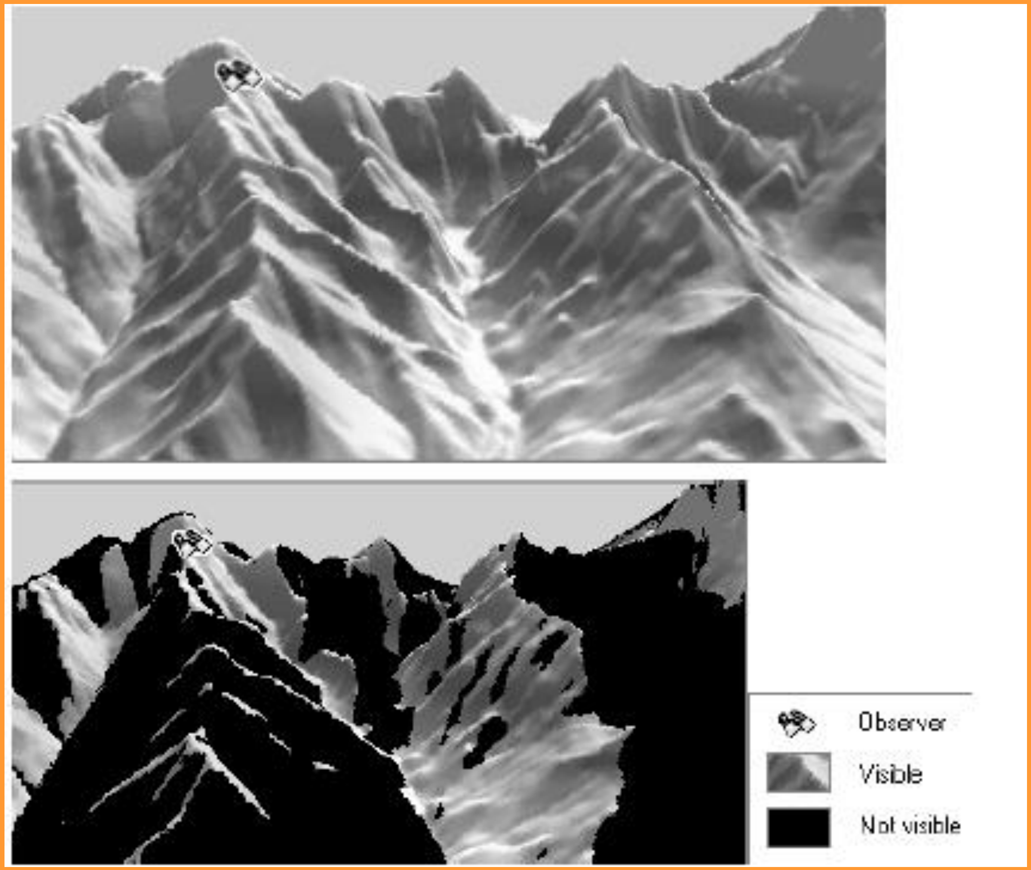
For graphical purposes – shows terrain relief in relation to other Layers; For analytical purposes – show illumination across the landscape by adjusting the sun angle for each hour of the day.





Example: Viewshed





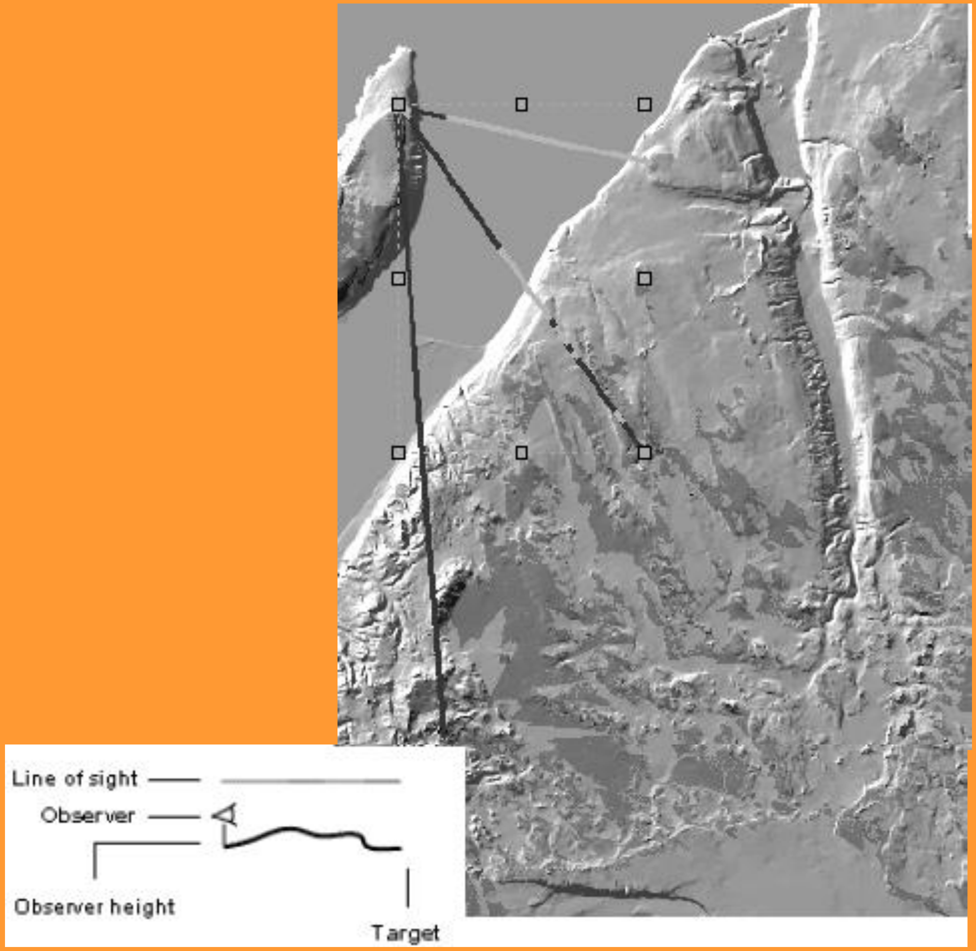
What is a good purpose for the viewshed tool?

The answer is the

Line of Sight Analysis

A line of sight is a line between 2 points that show the parts of the surface along the line that are visible to or hidden from the observer.

Creating a line of sight lets you determine whether or not a given point is visible from another point.



Part 4: Practical example

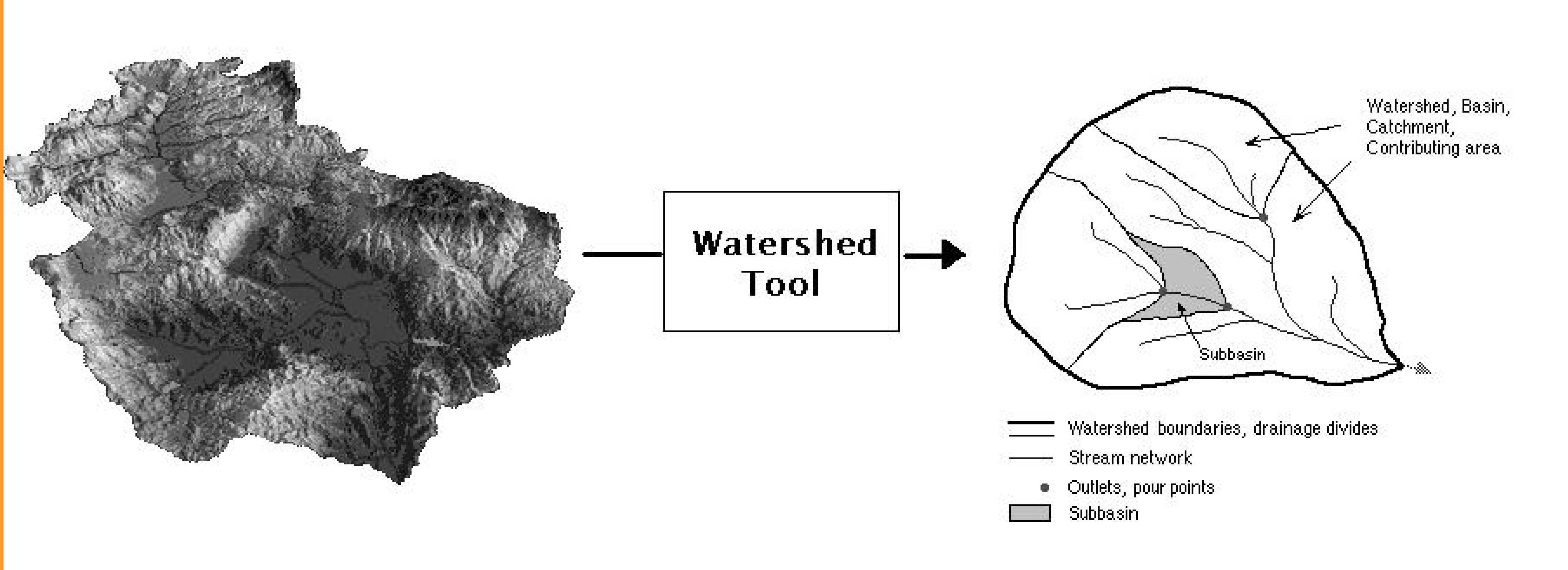
(Using ArcInfo)

Watershed Delineation

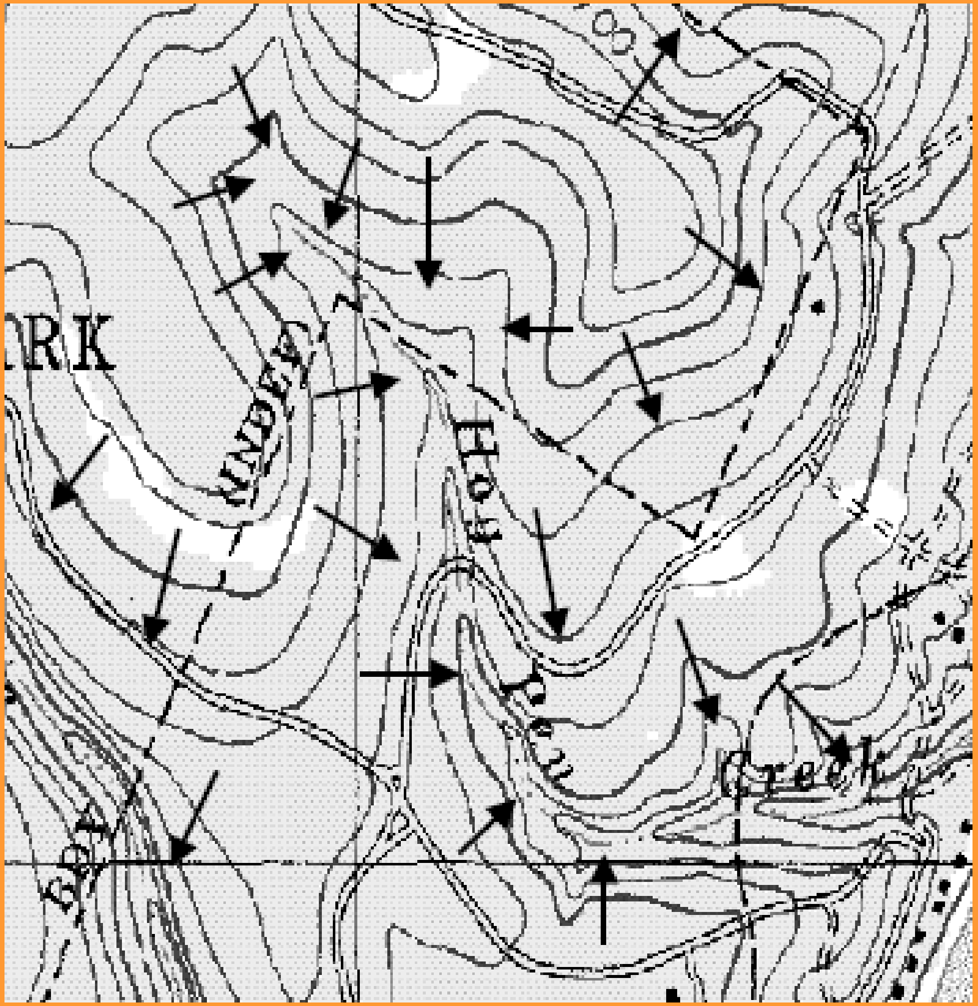
A watershed is an area that drains water to a common outlet as a concentrated drainage. This area is normally defined as the total area flowing to a given *outlet*, or *pour point*.

Watershed delineation is the process of identifying the drainage area of a point or set of points.

For years, have been determined using paper maps



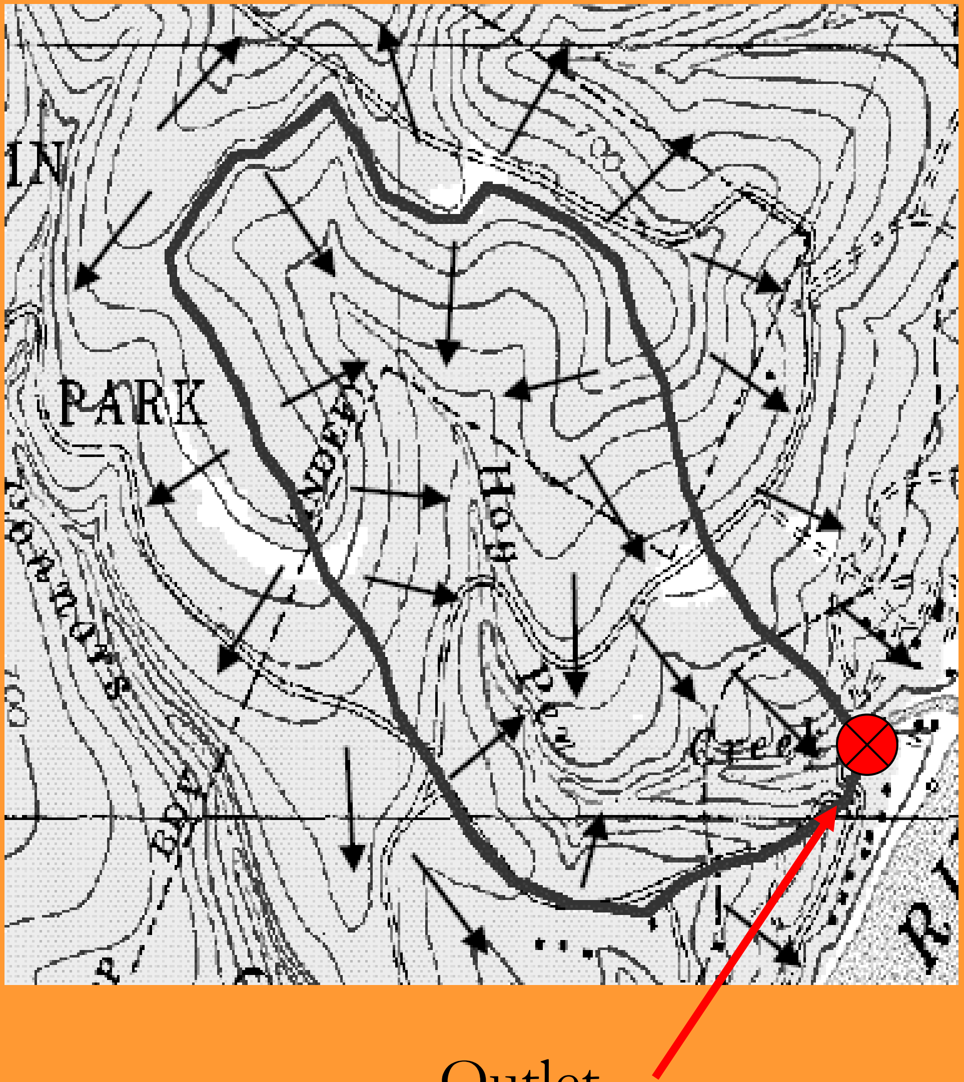
How is it possible identify Watershed from DEM? We know that water flows in the direction of the terrain's steepest downhill slope and is orthogonal to the elevation contour lines.



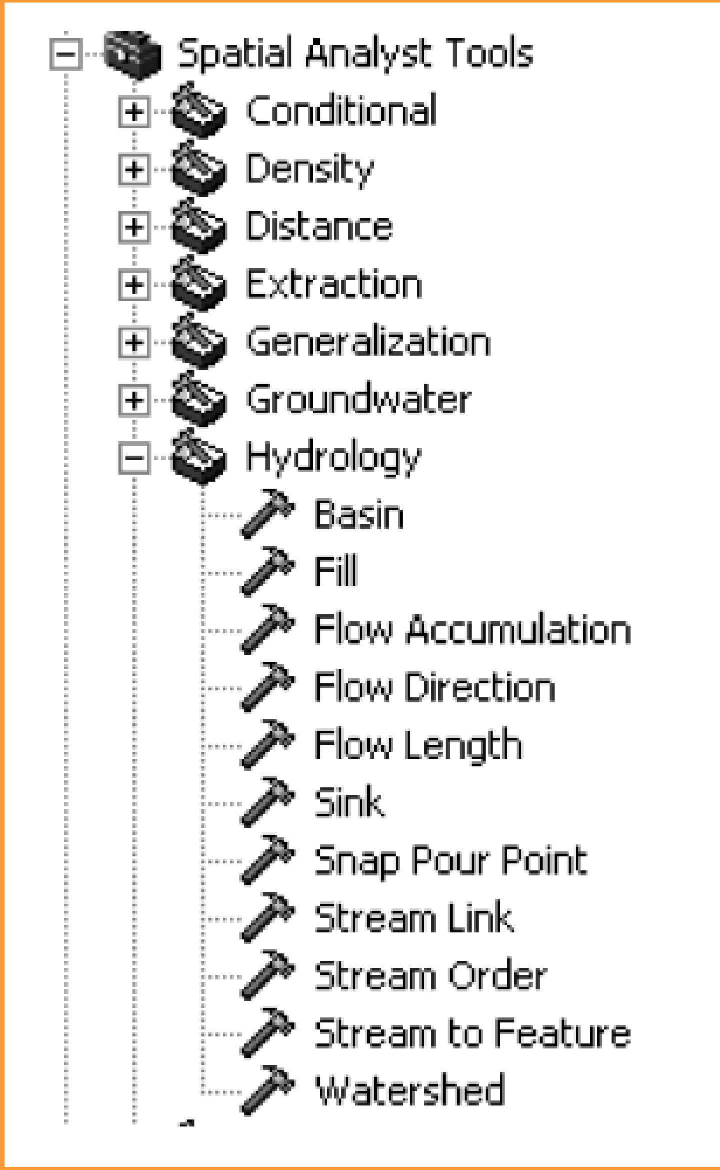
Streamlines do not flow towards drainage divides, and do not intersect them.

Drainage divides are found along the highest points of the terrain.

DEMs store the same type of information as contour lines but with a different data structure (Raster)

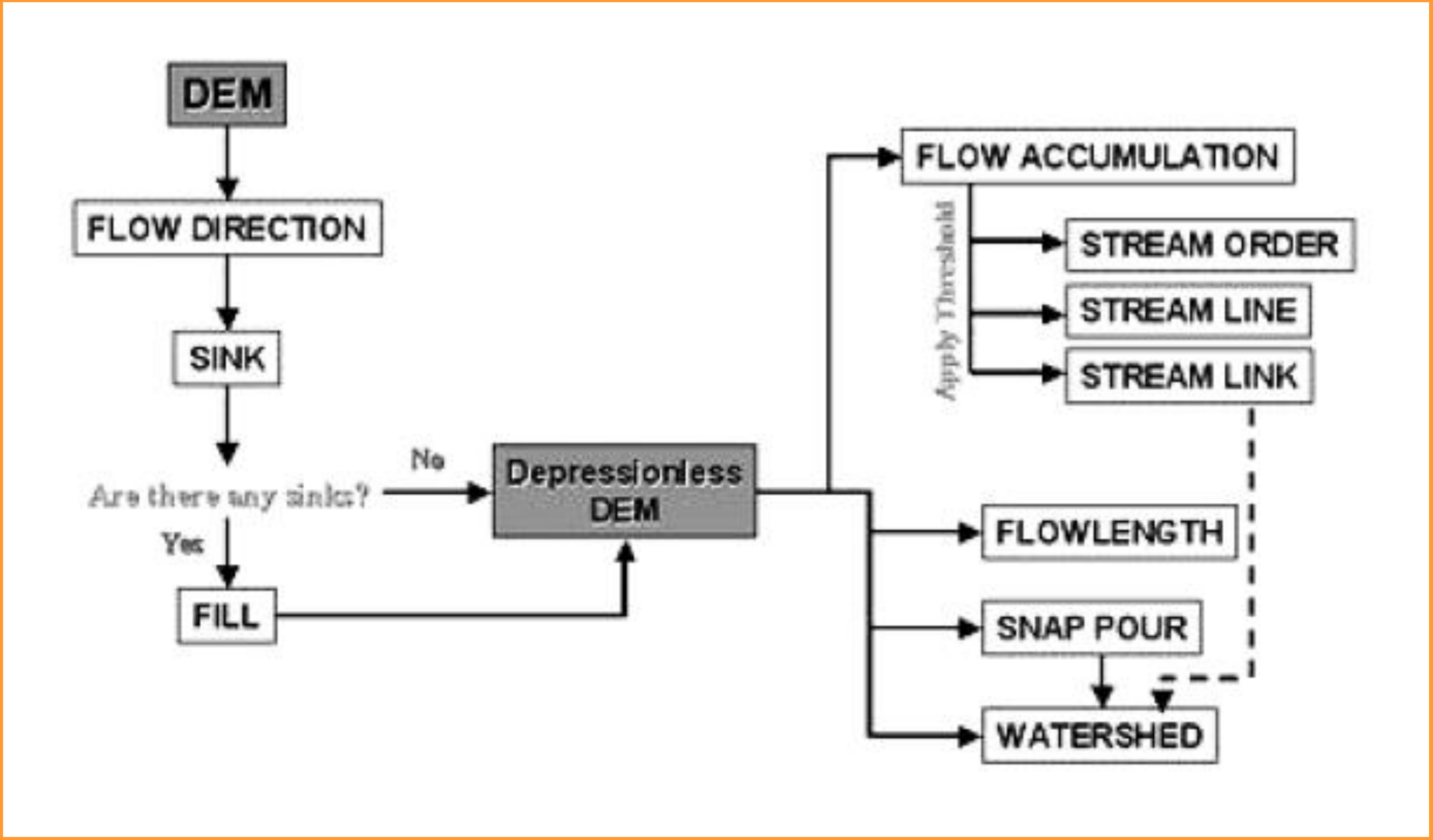


Outlet



ArcInfo is organized in modules, everyone of it doing specific operations: in particular ArcInfo has a **hydrology** package within the toolbox under the **Spatial Analyst tools**.

This is the scheme by which we can derive watershed; saddly, it is not a simple operation (as usual!) because it is the result of several steps one after the other:



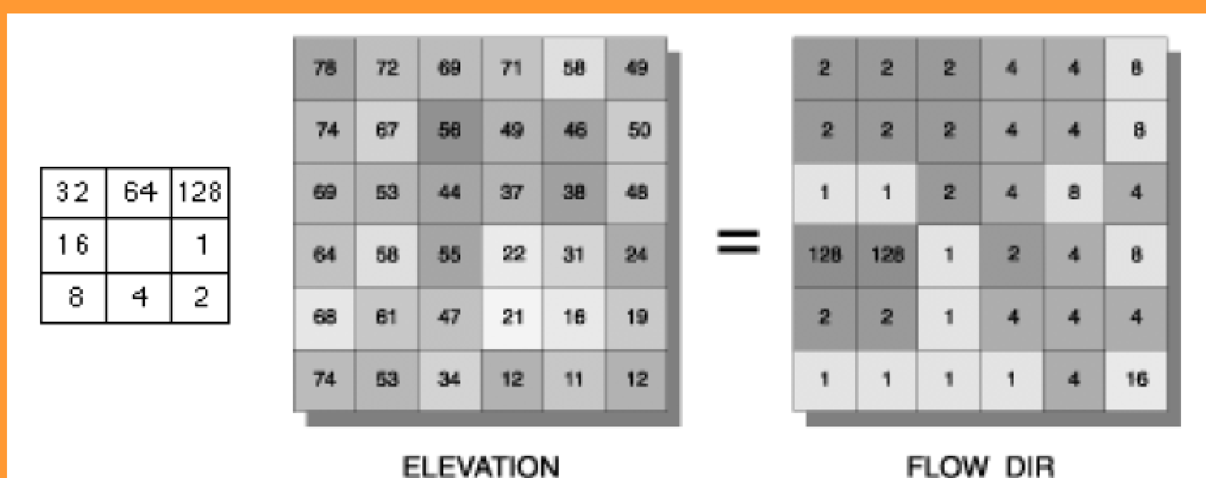
a) Calculating Flow Direction:

Water flows to one of its neighbor cells according to the direction of the steepest descent.

Flow direction takes one out of eight possible values, codified in this way. →

32	64	128
16		1
8	4	2

The eight-direction pour point algorithm assigns a flow direction code to each cell, based on the steepest downhill slope as defined by the DEM. The flow direction code indicates the cell – out of the eight neighbor cells – towards which the water flows.

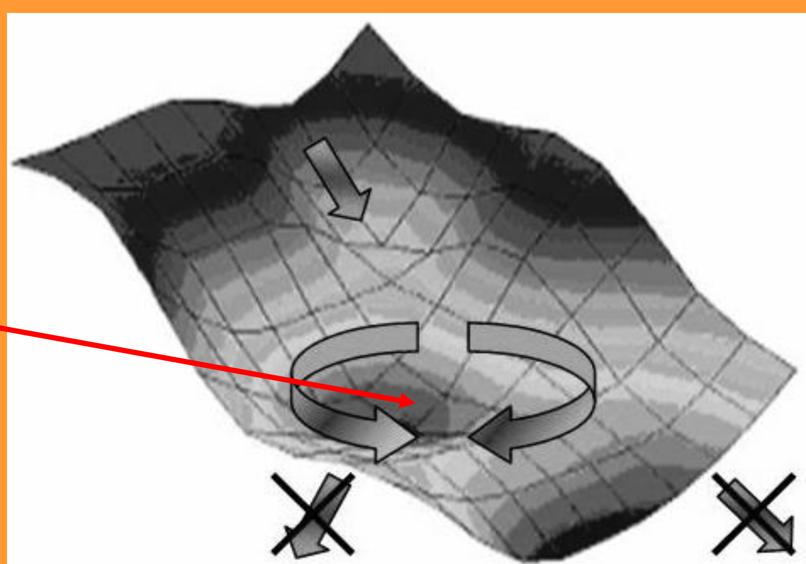


b) Sinks Removal

The problem rises because in every DEM we have errors related to uncertainties in elevation data acquisition methods (GPS, SRTM, digitalization topographic maps, laser scanning, etc.), and interpolation methods used to create regularly spaced.

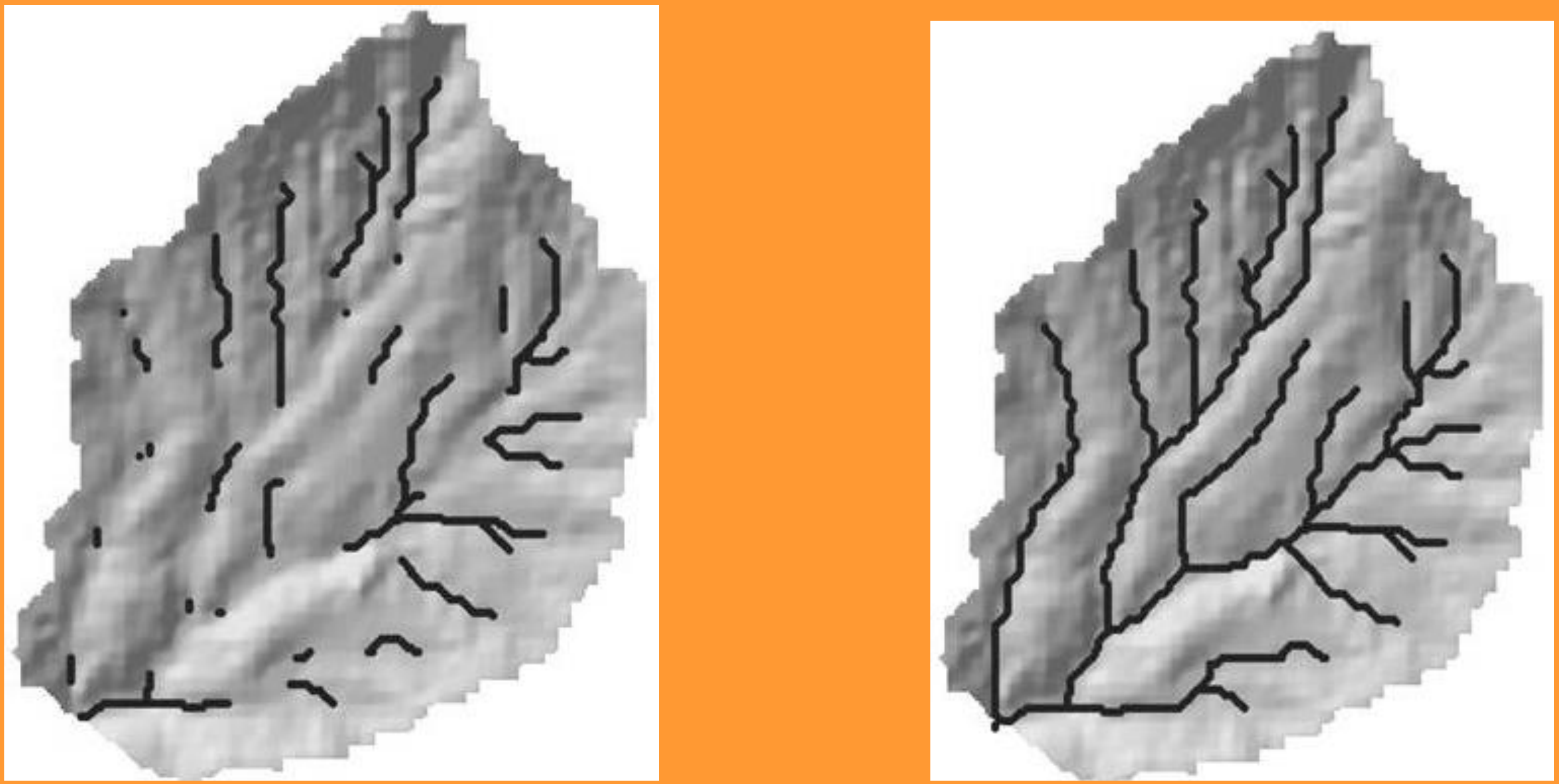
The final effect on DEM is the presence of *sinks* or *pits*:

Pit



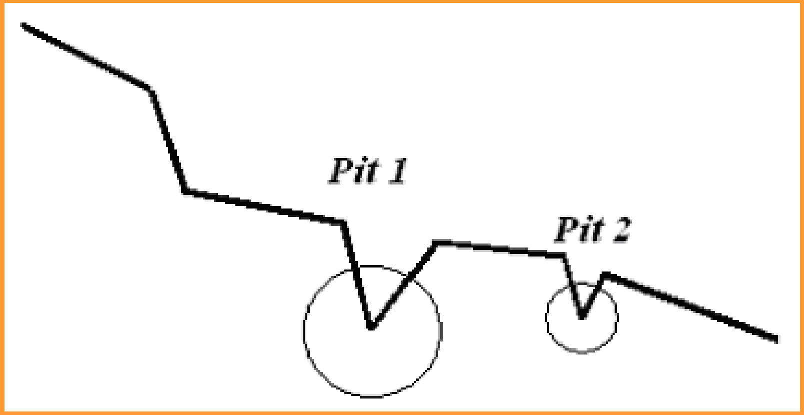
478.00	471.00	458.00	450.00	455.00
475.00	460.00	450.00	450.00	461.00
476.00	450.00	433.00	451.00	470.00
459.00	450.00	441.00	454.00	475.00
475.00	449.00	436.00	460.00	475.00

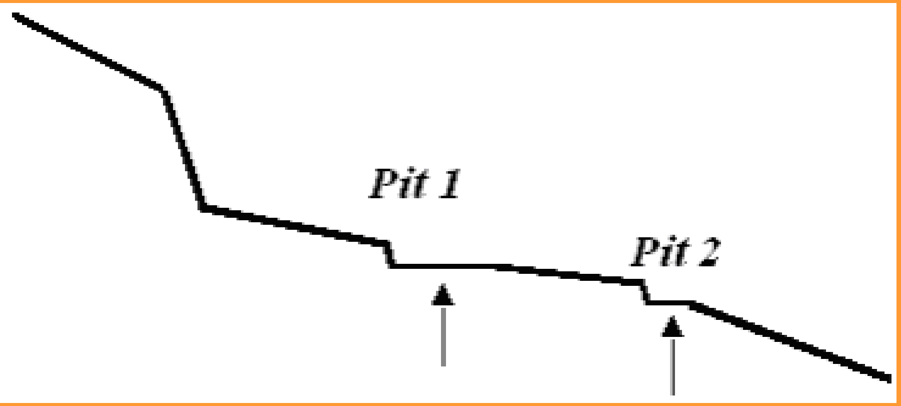
These pits must be removed before calculating the real Flow Direction (and then Watershed) Map, because otherwise we wold see these effect of Pits on drainage network:



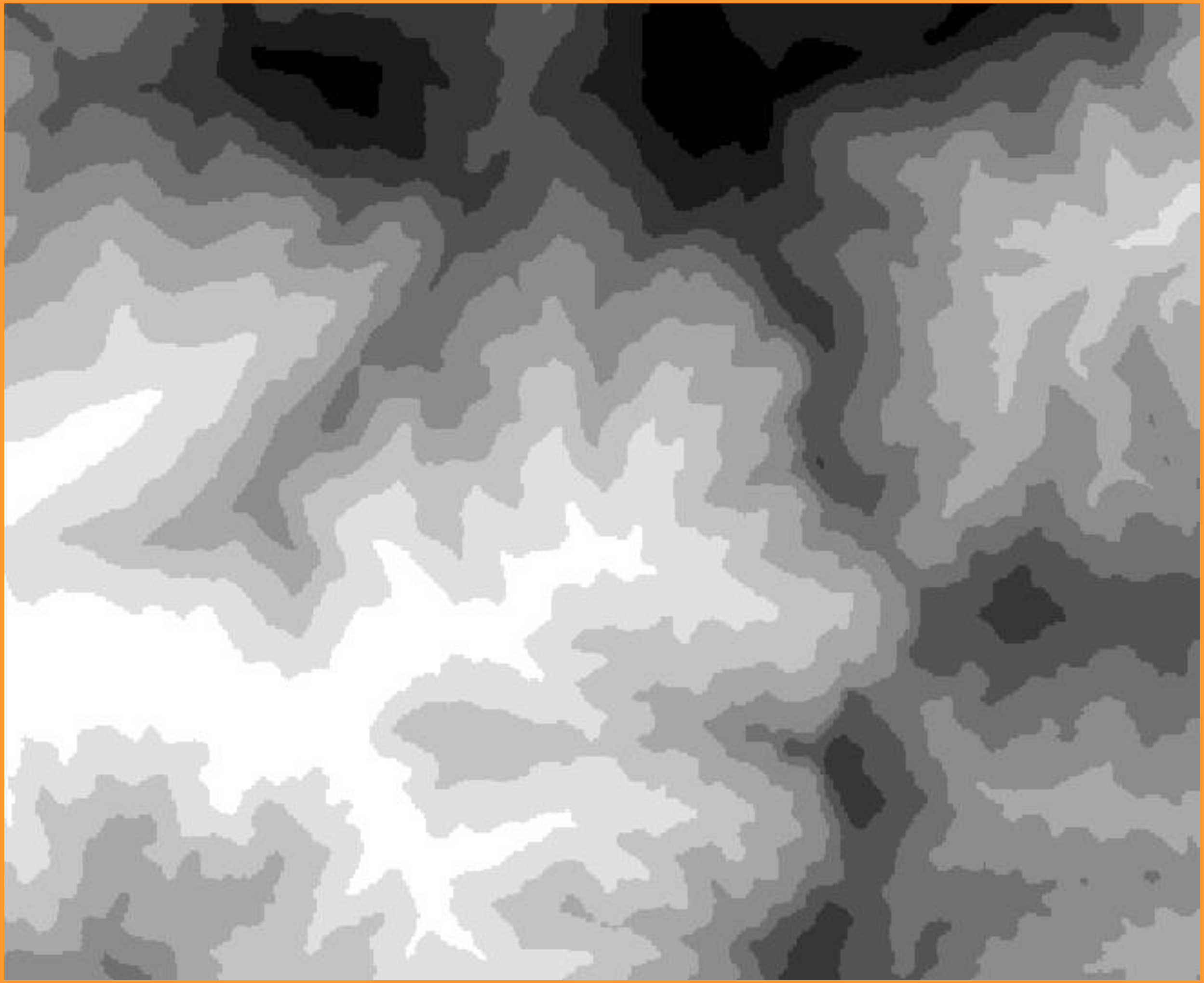
GEMINI Department
Water Engineering Section
Andrea Petroselli
petro@unitus.it

The most used approach (ESRI ArcInfo) to remove pits is the Filling method (FILL command) that arises the elevation of the pit until the lower elevation between the 8 surroundig cells:



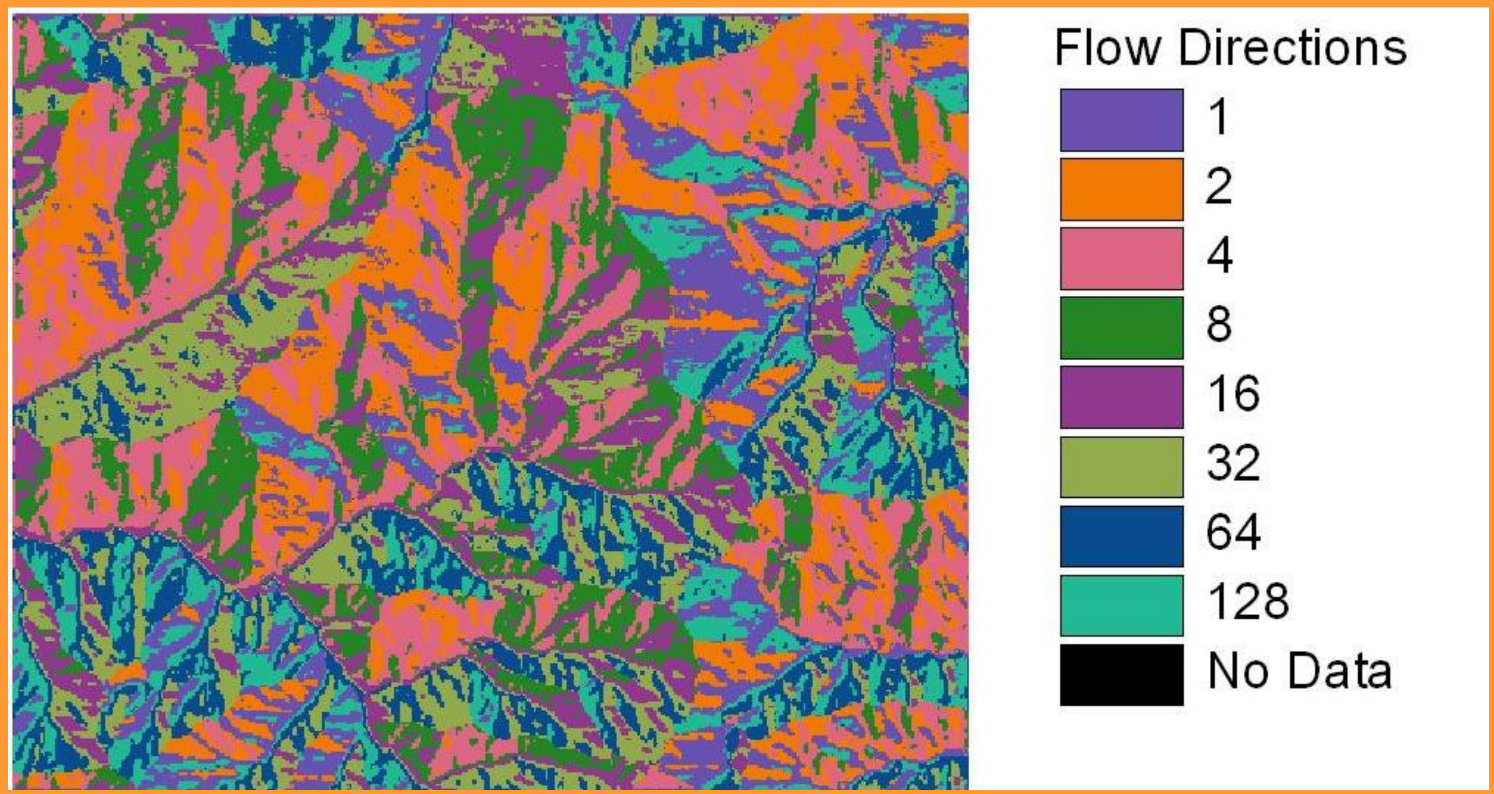


This DEM is the result of a Filling operation (application on FILL command on ArcInfo)



On the previous DEM it is possible now
to calculate the exact Flow direction
(application on FLOWDIRECTION command on ArcInfo)

32	64	128
16		1
8	4	2



Starting now from the Flow direction Raster it is possible to determine the **Flow accumulation Map** (application on FLOWACCUMULATION command on ArcInfo)

Flow accumulation is a measure of the drainage area in units of grid cells, and it is much similar to the river network…

Infact output cells with a high flow accumulation are areas of concentrated flow and may be used to identify stream channels.

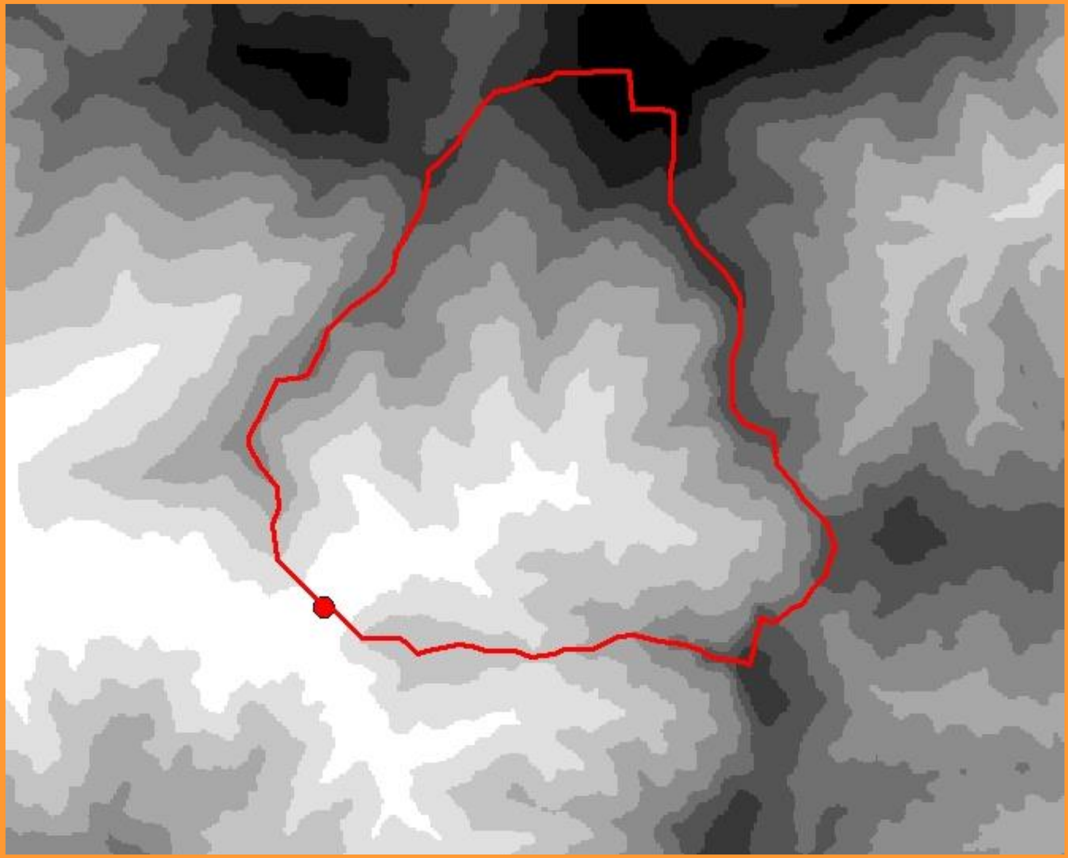
Output cells with a flow accumulation of zero are local topographic highs and may be used to identify ridges.

On the basis of the Flow accumulation Map we can choose our outlet (it must stay on the river network…)

OUtlet

And the ArcInfo command WATERSHED will extract, combining the information of the outlet with that of the flow direction, our watershed.

In this way we are able, with every DEM and choosing every possible outlet, to know how much area has the watershed, which is, which rivers insist within the watershed, and so on…



Thank you for your attention…