THE BASIS OF AN EU CO-FINANCED PROJECT "QUALIFICATION IMPROVEMENT OF THE LITHUANIAN GEOGRAPHICAL INFORMATION MANAGERS" (NO BPD2004-ESF-2.2.0.-02-05/0143)

Civil servants learning programme

DISTANCE LEARNING OF GEOGRAPHIC INFORMATION INFRASTRUCTURE

Training Material

GEODESY AND CARTOGRAPHY FOR NEEDS OF GEOGRAPHIC INFORMATION INFRASTRUCTURE

GII-06

Vilnius, 2008

Training material „Geodesy and Cartography for needs of Geographic information infrastructure“ (GII-06)

Author

Michael Govorov

Reviewed and edited by

doc. dr. Žilvinas Stankevičius (Vilnius Gediminas Technical University)

Reviewed by

doc. dr. Aida Mačerinskienė (Vilnius Gediminas Technical University)

From English translated and edited by

Astraneta UAB

Study material for civil servants distance learning programme “GII Distance Learning”, which has been approved by a law “RE Approval of Study Programmes for Civil Servants” (No 27V-72; Apr 17, 2007) by the State Service Department under the Ministry of the Interior has been prepared on the basis of an EU co-financed project “Qualification Improvement of the Lithuanian Geographical Information Managers” (No BPD2004-ESF-2.2.0.-02-05/0143)

Beneficiary of the project National Land Service under the Ministry of Agriculture.

Contractor of the project HNIT-BALTIC UAB, Vilnius Gediminas Technical University

Table of Contents

1 An Overview of Geodesy and Cartography

This module outlines the place and significance of cartography and geodesy when developing a Spatial Data Infrastructure (SDI) framework. This introductory module defines cartography and geodesy and examines the characteristics of maps, the classification of maps, and the cartography of atlases. Similarities and differences between traditional and digital cartography are described in this module. The role of geodesy, in a SDI framework, is examined.

<u>Module Outline</u>

- 1: Role of Cartography and Geodesy in Spatial Data Infrastructure
- 2 What is Cartography?
- 3: Role of Geodesy

1.1 Role of Cartography and Geodesy in SDI

1.1.1 Structure of A National Spatial Data Infrastructure

The term "Spatial Data Infrastructure" is used to describe the collection of technologies, policies and institutional arrangements that provides a basis for spatial data discovery by users and providers at all levels of government, the commercial sector, the non-profit sector, academia and individuals [1].

The major objective of a national Spatial Data Infrastructure is to support access to geographic or spatial information. This objective is achievable through the development, maintenance and availability of interoperable digital geographic data and the use of up-to-date digital technologies. Such information is highly valuable to decision-makers at national and local government levels.

The national SDI is composed of the following technologies and information sources:

- Geospatial data
- Clearinghouse or warehouse
- Metadata
- Standards
- Communication network for data access, visualization and analysis (e.g. a Portal)
- Data access policies
- Data protection
- Partnerships
- Leadership

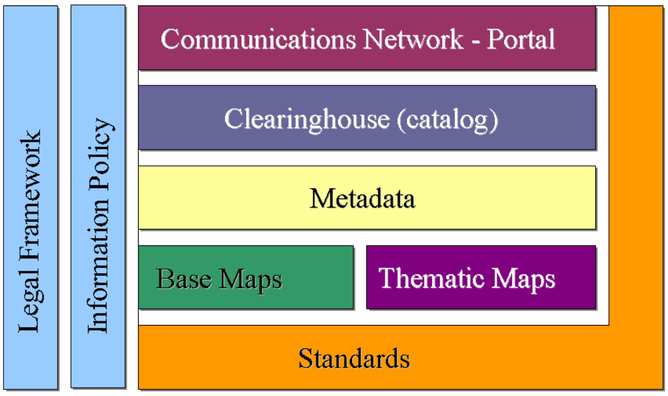


Figure: The components of a national SDI

You will learn about these components in courses delivered in this training program. In this course, you will learn how cartography and geodesy are used to build geospatial data content in a SDI. You will also learn how to visualize and prepare this information for spatial data viewing, presentation and exploration.

1.1.2 Geospatial data and maps for a Spatial Data Infrastructure

The roles of cartography and geodesy for a SDI:

- Cartographic and geodesic information are primary sources of data for a Spatial Data Infrastructure. Standard and non-standard spatial data themes, or layers of basic cartographic content, are known as framework, fundamental, foundation, or core data. These are essential ingredients in the construction of national and global SDIs [1].

Topographic reference maps provide a **framework** for SDI – a base on which other themes of data can be compiled and a foundation on which spatial information and attributes can be added.

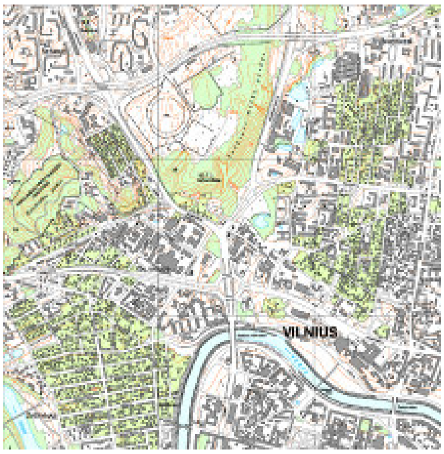


Figure: A sample of Lithuanian reference map at the 1:10,000 scale.

Framework data can be classified by level of usage and topographic themes:

Levels of usage:

- Federal
- Provincial
- Local
- Private

Core topographic themes:

- Geodetic control points
- Entities of settlements (e.g. buildings, blocks)
- Entities of movement or transportation (e.g. roads, water mains, pipelines)
- Cadastral (e.g. land parcels)
- Land use / Land cover
- Hydrography
- Topography and bathometry (e.g. contour lines, spot points, digital terrain models)

- Political and administrative boundaries
- Ortho-photographs
- 

Figure: A sample of ari-orthophoto

Thematic data can be compiled over a topographical framework. This spatial data can be used by government officials, policy-makers and decision-makers. There is a need for this information to be integrated into a common information infrastructure.

The thematic themes or layers can include:

- Thematic geo-referenced data for developing environmental policies on human health, climate, and biodiversity and water, air and soil quality. Displaying this data in a spatial framework is extremely informative.
- Population distributions and structure (age, gender, employment, etc.) generated from census data
- Urban and rural infrastructure data (telecommunication networks, water and sewerage systems, etc.) – within the public and private sectors.

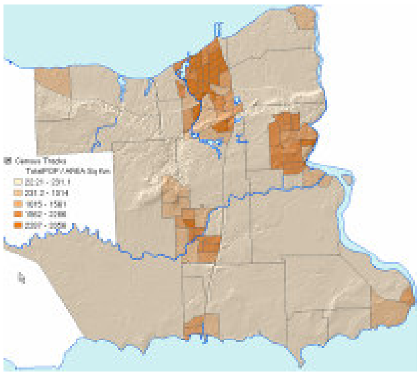


Figure: Sample of a census thematic map

- Knowledge of our world often usually begins by looking at atlases and maps. Maps are a primary source of geographic information. The cartographic visualization of SDI content via any interface (e.g. printed map or the internet map server) can provide a basis for spatial data analysis and spatial pattern exploration.

- Analyzing SDI spatial and non-spatial data through cartographic techniques, such as graduated point symbolization, dots, diagrams, choropleth and dasymetric presentations, and multivariate and isoline mapping techniques are being refined all the time.
- Digital maps and atlases are like geographic navigation switchboards – the spatial data and metadata from a SDI is readily accessible. It is easy for anyone to secure reference maps and spatial data through a multi-criteria search on the Internet.

1.2 *What is Cartography?*

In this course, we will be looking at cartography from the traditional point of view, but keeping in mind that nowadays most maps can be created by using new tools such as GIS, multimedia, image design and web mapping software.

1.2.1 A Map and Cartography

The production of a **map** involves a mathematical transformation of some part of the real world into a generalized, representative, and pictorial format. This representation can be done in analog (hard copy map) or/and digital (computer map). The products from both involve a cartographic visualization process. A map can be produced in three distinct formats: hardcopy (paper map, plastic prints), digital geo-dataset (spatial or/and cartographic) and cartographic visualization of data from geo-database (multimedia map presentation)

The International Cartographic Association defines **cartography** as "The art, science, and technology of making maps, together with their study as scientific documents and works of art. In this context may be regarded as including all types of maps, plans, charts, and sections, three-dimensional models and globes representing the Earth or any celestial body at any scale." [3] This definition can be successfully applied to digital and computer cartography as well, only the word "study" has deeper meaning: the multimedia map has depth or the ability to change the visual appearance and ability to mine for new information from linked databases.

Modern cartography is still pursuing the following two main goals:

1. **Map making** as a process of data collection, design, compilation, and production of maps and atlases.
2. Cartography as **a science** or the study of the artistic and scientific foundations of the rules for map making and use:

   - Cartography is an important branch of graphics (including machine-based graphic production).
   - Cartography is as an efficient way of manipulating, analyzing and displaying spatial objects, attributes, and relationships that occur in two and three-dimensional space.

Cartographic theory studies cartography as a science within distinct conceptual schemes [Cartwright et al, 1999]. There are few theoretical approaches or conceptual models, that attempt to explain cartography as a subject of map making and its scientific base from different points of view. The following are some conceptual cartographic schemes:

- Geometric schema - this concept is based on the need for metrical accuracy and precise geometric measurements when representing spatial information. Maps must provide accurate and precise information for analytical purposes (e.g., engineering, transportation, navigation, cadastral activities).
- Technologic schema – map making involves a series of processes associated with data collection, map design, production and reproduction.
- Presentation and artistic schemas – map design is a core element of this concept. It focuses upon visual forms – symbology, color and pattern, typography. The artistic

approach often exceeds the formal principles of map design; it employs an understanding of graphic characteristics to create forms and associations that invoke appropriate impressions and sensations.

- Communication schema – a concept that involves a series of information transformations and communication between cartographers and map users. Making and using a map are treated equally.
- Cognitive schema – an important concept that involves the role of the cartographer's and reader's perception of reality and cognitive processing of the flow and transformation of spatial information within communication schema.

There is some overlap to the above schemas and they often supplement one another. Thus, the following four categories of processes in cartography can be viewed as a combination of technologic and communication schemas – map making and map reading [3]:

- Collecting the data for mapping
- Designing and compiling the map
- Reading or viewing the map
- Responding to or interpreting the data

The geometric model represents stages of data collection and map compilation; the presentation and artistic schema deals with map design and user response processes.

Cartography can be subdivided into sub-disciplines. These divisions reflect the technological steps of map production as well as the classification of maps and technology used for map production and presentation.

- Mathematical cartography – studies cartographical projections
- Map production – processes of map creation includes planning, generalization and drafting, editing and revision
- Topographic cartography – defines the technology and standards for topographic map production
- Thematic cartography – studies the techniques for representation of qualitative and quantitatvie characteristics of point, line, polygonal and surfaces on a map (e.g., graduate symbol, choropleth, dot, or isopleth mapping techniques)
- Map design – concerns the graphic presentation of all information that is contained on a given map
- Map reproduction – making printed maps
- Multimedia cartography - studies scientific visualization and analysis of digital spatial data via audiovisual means
- Automated or computer cartography – process of map production with the aid of computer and other automated devices
- Geovisualization (including web cartography) – refers to techniques and tools designed to interactively "visualize" spatial phenomena

1.2.2 Cartography Now

The long history of cartography has been driven by human needs and the evolution of new technologies. The significant achievements in the history of map making are:

- 30,000 BC - figurative mapping (cave painting, bamboo stick maps)
- AD 200 - astronomical mapping
- AD 500 - nautical navigation charting
- AD 1700 - topographical mapping
- AD 1800 - thematic mapping
- AD 1900 - image mapping

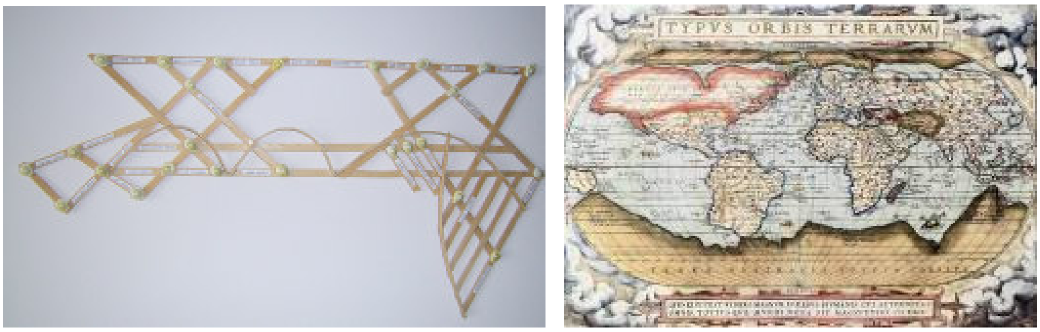


Figure: A bamboo stick map and Ortelius world map 1570

Technological revolutions have had significant impacts on the production of maps. Mapping in the Western world has undergone seven major technological revolutions:

- Manual technology - creation of tangible cartographic products on a paper
- Magnetic technology - use of a compass for surveying and mapping
- Mechanical technology - introduction of engraving and reproduction machines
- Optical technology - use of optics for precise measurement and cartographic image transfer
- Photo-chemical technology - development of photography and lithography were used for data collection (remote sensing), map compilation and printing
- Electronic technology - computer-assisted mapping systems used digital data, software and computer hardware
- Information age technology - use of GIS, spatial analysis, databases, multimedia software and networks

Modern technological tools make new revolution in cartography. These tools not only give the possibilities to produce better-looking maps at a faster rate, but also introduced completely new cartographic products and ways of map reading and usage.

Figure: A sample of modern interactive web map (from http://www.planet9.com/)

Cartographic production today is undertaken predominantly on computers and most often by using Geographical Information Systems (GIS). Traditional cartography is different from GIS in many ways, not only in the use of different tools but also in a different philosophy. However modern information cartography uses the same cartographic principles that have developed over centuries and GIS tools quit often simply replicate traditional cartographic techniques for map symbolization, lettering, layout, etc.

One of significant benefits of using GIS or computer mapping programs for map production is an iterative mapping improvement. A user can apply multiple design rules and handle findings by trial and error. Another huge difference between traditional cartography and modern digital cartography is that a map is only a cartographic representation data available in an underlying geo-database. A cartographic representation has great potential for interactive visual analysis, making symbolization on different attributes of spatial entities a possibility. Different symbolization and media techniques can be used and the results can be distributed over a network using different software.

Computer and web technologies bring the revolution in cartography in the form of:

- Digital map production
- Distributed spatial and cartographic databases
- Multimedia computer graphic for visualization
- GIS analysis
- Internet mapping

1.2.3 Characteristic of Maps

All maps are concerned with two fundamental elements of reality from which many spatial relationships can be formed. The basic element is spatial location that can be represented in two-, three-, or four-dimensional space and any number of attributes can be attached to that spatial location.

Traditional hardcopy maps depict positions in one predefined projection and coordinate system. The symbolization techniques are used to show attributes via the visual variables (e.g. size, color, pattern, etc). Digital mapping can use any available projection and coordinate system for cartographic visualization and can retrieve available attributes at location from a linked database and iteratively use the multiple visualization techniques. Digital cartography has more possibilities for scientific visualization, which incorporates not only the presentation of graphic information, but also its effect on the viewer.

Both traditional and digital maps inherit three main properties:

- All maps are mathematical transformation of reality. Most of maps have mathematical bases:
    - Scale – defined dimensional relationship between reality and the map
    - Map projection – a transformation of the spherical surface to a flat surface. With the exception of topographic plans, that do not use projections, they are scaled planimetrical representations of reality
    - Ellipsoid – model of the spherical surface
    - Systems of coordinate – a starting point and system of measurements
    - Datum – a frame of reference for measuring locations on the surface of the earth
    - Net of graticule or/and metric grid

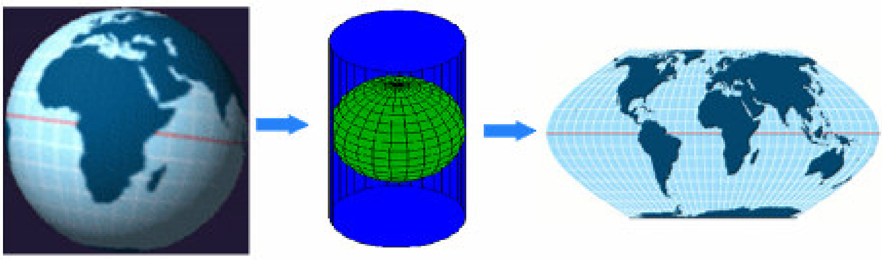


Figure: A map is mathematical transformation of reality

- All maps use symbolization for representations of reality. All maps use symbols (signs) or/and text to represent real world entities:
    - Various types of symbols are often used, such as lines, dots, colors, tones, patterns, etc.
    - Elements of map design based on a set of principles
    - Thematic cartography uses symbolization techniques (methods) for presentation of points, lines, polygons and surfaces

Figure: Maps use symbolization for representations of reality (from Web of Sony in Japanese)

- All maps are abstractions of reality. Word images undergo a verity of generalization operations, such as selection, simplification, classification, aggregation etc. Types of generalization:

- Model-based generalization – geometric transformations of real entities for the purpose of their representation in digital databases
- Object-based generalization - geometric transformations of digital databases objects for the purpose of their representation in different scales
- Cartographic generalization – graphic transformations for visualization purposes and map production

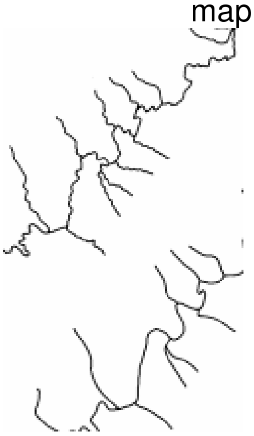


Figure: A map is generalization of reality

Digital maps inherit the same characteristics but with some peculiarities. Indeed digital maps have two representations, the first one is database, and the second is cartographic or visual. The database representation is a digital model of reality; it can be, for example, a geo-relational model of vector and/or raster representations.

Digital cartographic representation involves graphically modeling reality. This represents the production of graphical objects from database. Thus, two steps of abstraction have to be applied in order to produce digital cartographic representations. These steps are the model-based generalization – abstraction of reality in order to create its database model - and cartographic generalizations that convey a map's clarity.

Digital cartography introduced new symbology, representation techniques and visual variables (e.g. frequency, transparency, etc) for cartographic visualization. However, fundamentally any map remains the symbolic representation of reality – whether displayed in hard copy or via multimedia.

1.2.4 Classification of Maps

Like people, cartographers group and categorize subject matter in order to understand them better. Thus, maps also can be categorized for practical (e.g. for cataloguing in a library or database) and scientific reasons (e.g. for understanding division between base and thematic contents). Maps are categorized on the basis of their intended use. Maps can be classified from three points of view: subject matter; scale; and, area coverage.

- **Subject matter**:
  There are three base classes of maps:
  - General reference maps – another name commonly applied to the general reference map is "general-purpose map"
  - Thematic maps – sometimes called "special-purpose" or "statistical" maps
  - Special maps – charts, navigation maps, etc.

You will learn about a topographic map (a type of general reference map) and thematic maps in the following modules.

General reference maps can be further classified into:

- Topographic plans – large-scale, no projection applied
- Topographic maps – medium-scale, usually Transverse Mercator projection
- General topographic maps – small-scale map (up to 1:1,000,000), projection applied

Figure: A general reference and thematic maps

Thematic maps can be classified on the basis of any number of themes, including:

- Maps of physical world
    - the atmosphere
    - the oceans
    - geomagnetism
    - geology
    - the land surface
    - vegetation
    - animal life
- Maps of peoples and their activities
    - population
    - characteristic of peoples
    - cadastral
    - economic activities
    - movements of goods
- Maps of social environment
    - crime statistics
    - medical maps
    - living condition
    - ecological maps

▪ Classed by **scale**:

For general reference maps:

- topographic plan (scale <= 1:5000), projection is not applied

    - topographic maps (1:5000 < scale <= 1:200,000)
    - general topographic maps (1:200,000< scale <= 1:1,000,000)
    - general maps (scale > 1:1,000,000)

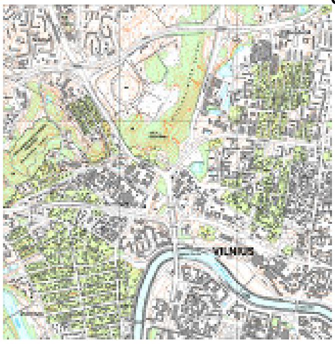


Figure: A topographic map (1:10,000)

For thematic maps, classification may depend upon the content or subject matter of the map:

- large-scale maps
- medium-scale maps
- small-scale maps

- Classed by **territory** coverage:
    - the World
    - continents
    - countries
    - economic regions
    - states / provinces
    - cities

1.2.5 Atlas Cartography

An atlas is a collection of maps that display spatial information in an organized and coherent manner. The types of geographic information displayed in any atlas can be visually interrelated by the cartographer and map reader. Atlases can be focused on specific themes (e.g., topography, national wealth, human health, etc.).

Atlases are classified on the bases of content, functionality and production format:

- Content:
    - General reference atlas - collection of general reference maps. Nowadays these atlases may include ortho-photo maps - standard topographic maps with satellite images and areal photographs on the background
    - Physical world atlas - collection of physical maps displaying landforms, lakes, etc.
    - Social-economic atlas - collection of social-economic maps showing population characteristics, GDP, etc.
    - Complex regional atlas - territorial collection of maps that includes the maps of different physical and social-economic themes for appropriate area. Such atlases combine detailed topographic maps at scales that are proportional to the size of the region and an abundance of thematic maps that normally cover the whole region but often also include maps of regional districts, cities and areas of special interest.

- Function:
    - General reference atlas – designed for specialists and general public
    - Educational atlas – designed specially for particular teaching subjects in schools and universities

- Production Format:
    - Traditional printed atlas
    - Electronic atlas – a collection of maps and associated databases that are available in digital, multimedia and/or net environments. Visualization, animation, spatial analysis and mapping tools can be part of an electronic atlas

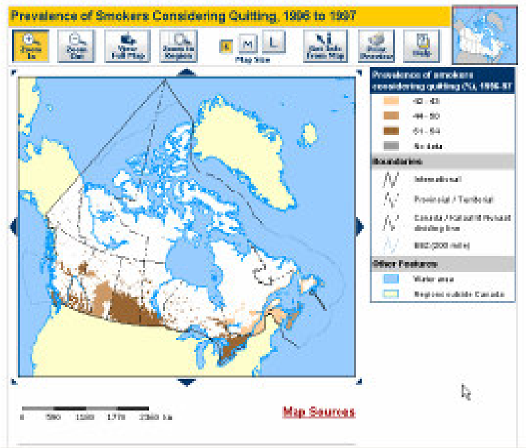


Figure: Web Atlas of Canada (from http://atlas.nrcan.gc.ca/)

Today it makes much more sense to produce atlases in digitally and make them available on CDs or websites.

1.3 Tasks of Geodesy

Geodesy is the science of the measurement and mapping of the earth's surface. There are a few major tasks that geodesy addresses:

- Determination of the shape or figure of the Earth's surface and the external gravity field of the Earth and other cosmic bodies as functions of time
- Determination of the mean of the Earth's ellipsoid from parameters observed on and beyond the Earth's surface
- Determination of reference systems, coordinate systems and datums

In general, these tasks lead to the creation of *physical* and *mathematical* models of the Earth.

1.3.1 Global Geodesy

Geodesy, as like other professions, spans activities that range from the purely theoretical to the very practical. Most of the theoretical activities are in the field of global geodesy. Global geodesy determines the figure of the Earth including what includes: size and shape and external gravity field.

Global geodesy defines two models of the Earth:

- Mathematical - **Spheroid**
  - Definition of mathematical models of Earth
  - Mathematical computations on mathematical model of Earth
- Physical - **Geoid**
  - Determination of the Earth's external gravity field
  - Determination of physical size and shape of Earth

A reference **spheroid** is a 3-dimensional circle and ellipse flattened at the poles that mathematically models the physical surface of the Earth (geoid) and provides a basis for the accurate measurement of the location (horizontal) and elevation (vertical) of places on the Earth's surface.

J. Richer in 1672 discovered that the length of a regulated pendulum had to be changed has he was travelling from Europe to Cayenne, Guiana. This implied a change in the gravitational field on the surface of the Earth, which suggested a flattening of the area near the Earth's poles. Newton (1687) obtained a rotational ellipsoid as an equilibrium figure for a homogeneous, fluid, rotating Earth based on the *law of universal gravitation*. The flattening was equal to $f = (a - b)/a$, where a and b are respectively the major and minor equatorial radii. A current estimate for the equatorial radii - a = 6378137 and b = 6,356,752.3 meters. A reference spheroid should be used for medium and larger scale maps (i.e., 1:25,000 or greater).

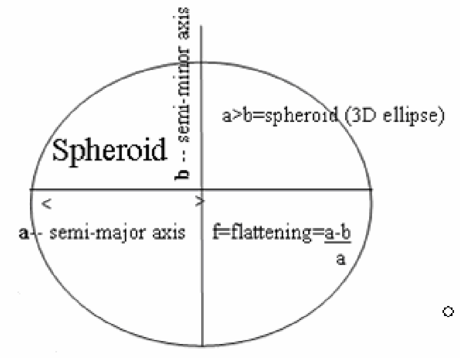


Figure: A spheroid is a 3-dimensional circle and ellipse flattened at the poles

Sometimes term reference spheroids are referred to in the literature as "reference ellipsoids". More strictly speaking spheroids are biaxial body of revolution, and ellipsoids are triaxial body.

By the 19th century, the assumption of an ellipsoidal Earth was recognized as false. Adjustment of several arcs for determination of *a* and *f* resulted in discrepancies that were larger than could be accounted for by measurement error. Friedrich Robert Helmert developed a model of the Earth (i.e., geoid) that accounted for the deflections of the vertical.

Geoid is a model of the physical surface of the Earth. It is the border between the solid and fluid masses of the Earth and the atmosphere. The geoid surface is an equipotential surface of the Earth's gravity field. The material that makes up the Earth complicates the study of its shape. The Earth is not a mass of solid rock. This leads to the fact that the Earth's gravity field isn't the same everywhere. Although these differences are very small, measuring these differences make it possible to help figure out the Earth's shape.

Geoid approximates the mean sea level in open oceans without disturbances that would occur close to the ocean's surface, such as tides, winds, waves, temperature, pressure, and salinity differences, etc.

The horizontal and vertical spatial reference systems are the common reference systems geodesists and surveyors use to uniquely determine positions anywhere in the Earth. These spatial reference systems are defined by horizontal and vertical **datums** that are the primary practical product of global geodesy which we have to consider for mapping. These are used to define parameters in GIS and GPS.

The question becomes: What spheroid should be used for a particular country or region and how does one fix it in three-dimensional space? The reference spheroids or ellipsoids are the horizontal surfaces to which the geodetic latitude and longitude are referred. But to serve in this role, a spheroid (together with the associated Cartesian coordinate system) must be fixed with respect to the Earth. Such a fixed spheroid is often called a **horizontal datums**.

Vertical datums define a reference from which vertical positions are measured. Traditionally the starting points for measuring elevations or heights are mean sea level points or **geoid** surface.

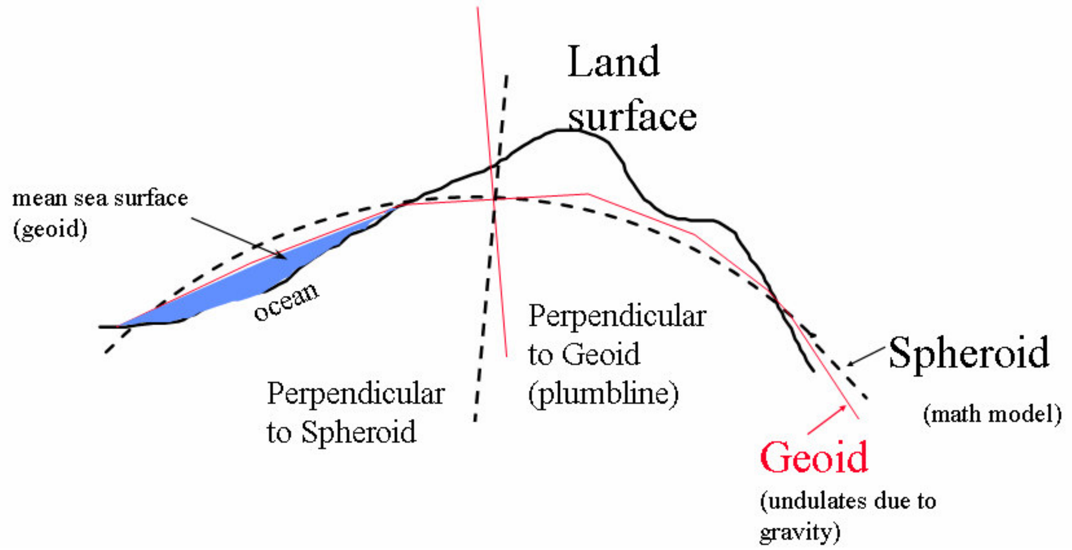


Figure: Surfaces of the Earth

We will discuss notions of **spheroid** and **geoid**, and especially **datum** (horizontal and vertical) and their practical use in surveying and cartography in the following modules.

1.3.2 Surveying

A vital practical part of geodesy is surveying. **Surveying** is the art and science of measuring the surface of the Earth and its features. Our physical environment needs accurate surveys and maps to build living infrastructure, in addition to locating natural resources and boundaries. Surveying determines the spatial location of points or establishes pre-determined points on or near the Earth's surface. Surveying is one primary way of collecting accurate data for SDI databases and mapping.

There are many different technologies used in surveying and many different types of surveys produced. Generally speaking, surveys will either take into account the true shape of the Earth (geodetic surveys) or treat the Earth as a flat surface (plane surveys). Each of these types of surveying can be carried out for the purpose of positioning features on the ground (horizontal surveys), determining the elevation or heights of features (vertical surveys) or a combination of both.

- Geodetic surveying defines the surface of a country by:
    - Coordinates of a sufficiently large number of control points (a control **geodetic networks**)
    - Must consider overall curvature of Earth
    - Most work in this area being done by GPS today

Figure: A handheld computer with GPS

- Plane surveying
    - Includes topographic surveying, cadastral surveying, and engineering surveying
    - Does not consider the overall curvature of the Earth
    - Often uses control points established by geodetic surveys
    - Work in this area is being done using traditional and modern surveying techniques, including GPS

Traditionally, the classical geodetic way of positioning points has been in the mode of **geodetic networks**, where a whole set of points is treated simultaneously. After the observations have been made in the field, the positions of network points are estimated using optimal estimation techniques that minimize the quadratic norm of observation residuals, from systems often considering hundreds of thousands of equations. A whole body of mathematical and statistical techniques dealing with network design and position estimation (network adjustment) has been developed.

The horizontal and vertical geodetic networks for a nation's extent are usually called **national control networks.** A control network is a series of well-spaced and interconnected markers in the ground, which have accurately determined positions, or coordinates, and elevations. Control networks provide fixed or anchor points on which to base or "reference" surveys and have been the main tool for positioning needed when mapping boundary demarcations and other spatial infrastructure applications.

The surveyor is the guardian of accuracy. Accuracy levels of control networks can be classified hierarchically: primary control, secondary control, and local control.

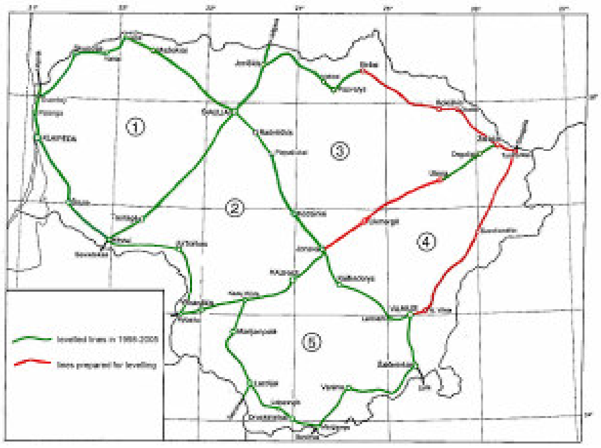


Figure: The Lithuanian National Geodetic Vertical Network of the primary level

Horizontal control can be established using differing methods such as trilateration, triangulation, theodolite, Doppler, inertial, GPS, and photogrammetry. Vertical control may also be established by many different methods, the primary vertical control network is established using differential leveling only.

You will learn the basics of traditional surveying and principles of GPS measurements in the next modules.

1.3.3 Conclusion

Spatial Data Infrastructure is an umbrella of policies, standards and procedures under which organisations and technologies interact to foster more efficient use, management and production of spatial data. The main tools of collection, representation and analysis of spatial data are geodesy and cartography. In this course you will acquire knowledge of traditional and modem cartographic and geodetic techniques which can be required to build, support and use Spatial Data Infrastructure recourse.

Module self study questions:

- What are the three main characteristic of maps and how do they define the nature of a map?
- Outline at least three principles differences between traditional and digital cartography.
- What is the principle difference between geodetic surveying and plane surveying?
- Define and explain the term "datums."

Required Readings:

- **Chapter 5: Geospatial Data Visualisation: Online Mapping,** Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI, Version 2.0 25 January 2004, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf
- **Cartographic Communication,** The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/cartocom/cartocom_f.html
- **Chapter 2: How maps inform,** Modeling Our World, Zeiler, M., ESRI Digital Library, 1999.
- **All sections from "What is Geodesy?" Chapter,** Canadian Spatial Reference System on-line manual, http://www.geod.nrcan.gc.ca/geodesy/whatis/index_e.php

ESRI Virtual Campus Course:

- **Module 1: Basics of Data and Information**, Turning Data into Information Using ArcGIS 9

Assignment:

- **Assignment 1:** Starting with Map Design in GIS

References

- [1] Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI, Version 2.0 25 January 2004.
- [2] Thematic Cartography and Geographic Visualization, Terry A. Slocum, Robert B McMaster, Fritz C. Kessler, Hugh H. Howard, 2nd ed., 2004.
- [3] Elements of Cartography, Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, 6th ed., NY: John Wiley & Sons Inc., 1995.
- [4] Modeling Our World, Zeiler, M., ESRI Digital Library, 1999.
- [5] The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/notes.html

Terms used

- SDI
- Cartography
- General-purpose map
- Thematic map
- Cartographic symbolization
- Cartographic abstraction
- Atlas
- Geodesy
- Surveying
- Spheroid
- Geoid
- Geodesic Control Network

Positioning: This term is used in geodesy as a synonym for the "determination of positions" of different objects, either stationary or moving.

Coordinate system: In three-dimensional Euclidean space, which we use in geodesy for solving most of the problems, we need either the Cartesian or a curvilinear coordinate system, or both, to be able to work with positions. The Cartesian system is defined by an orthogonal triad of axes coordinate; a curvilinear system is related to its associated generic Cartesian system through some mathematical prescription.

Coordinates: These are the numbers that define positions in a specific coordinate system. For a coordinate system to be usable (to allow the determination of coordinates) in the real (earth) space, its position and the orientation of its Cartesian axes in the real (earth) space must be known.

Ellipsoid/spheroid: Unless specified otherwise, we understand by this term the geometrical body created by revolving an ellipse around it minor axis, consequently known as an ellipsoid of revolution. By spheroid we understand a sphere-like body, which, of course, includes an ellipsoid as well.

GPS: Global Positioning System based on the use of a flock of dedicated satellites.

Datums: A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines.

2 Nature and Sources of Spatial Data

This module outlines the roles of spatial data within a Spatial Data Infrastructure (SDI) framework. The module describes the formats of spatial and cartographic data storages. The various input sources and techniques of spatial and cartographic data collection are discussed. Primary spatial data resources in Lithuania are listed within this module.

<u>Module Outline</u>

- Role of Spatial Data within SDI
- Spatial and Cartographic Data Storage
- Techniques for Spatial and Cartographic Data Collection:
    - Paper map digitizing
    - Manual entry with geocoding
    - Raster data input
    - Sampling and census
    - Ground surveys
    - Attribute entry

2.1 Topic 1: Role of Spatial Data within SDI

Spatial data, especially **core** geo-reference data discussed in the previous lecture, provides a framework for any Spatial Data Infrastructure. Today, spatial data (as a form of geographic information) is a mass-market product on its own and is offered in **hard copy** and **digital** formats. There are many private businesses involved in general-reference data creation and update, especially with reference to digital data. For example, NAVTEQ offers digital maps that provide a highly accurate representation of the detailed road network including up to 204 attributes, such as turn restrictions, physical barriers and gates, one-way streets, restricted access and relative road heights for navigation for 58 countries of Europe, North America, etc.

The main driving force behind these developments is GIS, GPS and Remote Sensing technologies. This is particularly the case in the development of desktop and handheld GIS equipment, free accessibility of satellite images through the Internet, and simplicity of coordinate measurements with handheld code positioning GPS devices.

In the era of traditional ‘map making’, geographic information in the **hard-map** formats was primarily a government **asset.** General-reference maps were generally created and printed by government agencies or under their supervision. In some countries, government’s strictly controlled the distribution of general-reference maps. Therefore, governments determined map specifications (e.g. accuracy, scale ranges, datums and projections, symbolization systems, generalization rules etc.) and determined formats for core cartographic products, such as topographic plans, topographic and general topographic maps, geodetic control points, geographic feature names, ortho-photo imagery and cadastral maps, and sometimes for general purpose maps (scale > 1:1,000,000).

Today, many countries have moved to a digital mapping environment and more and more private organizations are involved in general-reference data collection, mapping and distribution. In the Republic of Lithuania, the GIS Centre, which is owned by the State, is one of the major organizations involved in geo-reference mapping and spatial data production.

Thousands of organizations spend billions of dollars each year producing and using geographic data as part of their activities. Source data may be collected from many different organizations such as federal or local governments, utilities, research organizations, and commercial businesses. Today, this data is often available online.

Data collection for mapping and its digital storage remains the most expensive and time-consuming stage of spatial product creation. This may cost 60-80% of the overall costs of a spatial project. A key initiative of SDI addresses the *leveraging* of individual geographic data efforts so that data can be exchanged at a reasonable cost or free by government, commercial, and nongovernmental contributors.

There are a number of ways data can be acquired. Data can be given freely, exchanged by an agreement, or bought and sold commercially. Data distribution is an on-going issue in the most countries of the world. There are two main approaches and/or arguments as how to distribute spatial data that is produced by state agencies:

- Approach 1 – Freedom of information and access to public resources (e.g. the United States policies)
- Approach 2 – Cost recovery and profit making (e.g. the Canadian and Lithuanian policies)

In this module, we will discuss, in some detail, the main sources of spatial data and identify what techniques are used for spatial data collection for mapping and SDI storage. Maps were, and remain, one of the main forms of spatial data collection, storage, and representation.

2.2 Cartographic and spatial data storage

There are two main media formats for spatial data storage. Hard copies of cartographic products are the traditional way spatial data are presented and stored. Digital spatial and/or cartographic databases, e.g. in form of spatial warehouse, are another and modern way of spatial data collection. The information from such collection requires data to be in digital format and can be distributed over networks.

The sources and techniques of spatial data collection for hard- and digital-forms of mapping have some similarities and differences. Before we discuses the spatial data collection topic, we must understand the main concepts behind data storage formats. In addition, these data collections can be used for map-making and the creation of new data storages.

2.2.1 Hard-copy Maps

Traditionally, the hard-copy map (e.g. paper) has performed a storage function for spatial information. General-reference paper maps usually are stored by government agencies and distributed upon request. Today, many government agencies have on-line services for map preview, order, payment and downloadable in raster printable format. Thematic maps can be found in map collections in libraries and map rooms of research and educational institutes. These can be purchased in the book shops and on-line. In many countries, topographic maps also can be bought in commercial or specialised book or map stores.

The following hard-copy geo-reference printed map series of Lithuania are available from the GIS Center store:

- Printed topographic maps at scales of 1:50,000, 1:250,000, and 1:1,000,000 for the entire country
- Printed topographic maps at a scale of 1:10,000 for major cities and their surroundings

Printed topographic maps at scales of 1:5,000, and 1:1,000 for major cities and their surroundings are stored in municipalities.

2.2.2 Digital Storage for Cartography

There are a few levels of spatial data representations or modeling – conceptual, logical and physical. In general, data models represent a set of guidelines to convert the real world (called entity) to the logically and/or digitally represented spatial objects consisting of the attributes and geometry. These attributes are arranged by thematic or semantic structures while the geometry is represented by geometric or geometric-topological, or raster structures. A data model is a collection of concepts for describing data.

A conceptual data model is a spatial data representation of phenomenon structured by user views. On the conceptual level, a spatial data can be represented digitally by using vector and/or raster modeling or a combination of both (hybrid). Vector models use discrete points, lines and/or areas with attributes to represent discrete objects corresponding to whole or parts of real world entities. Raster models use regularly spaced grid cells or pixels in specific sequence. Attribute can be linked to pixels.

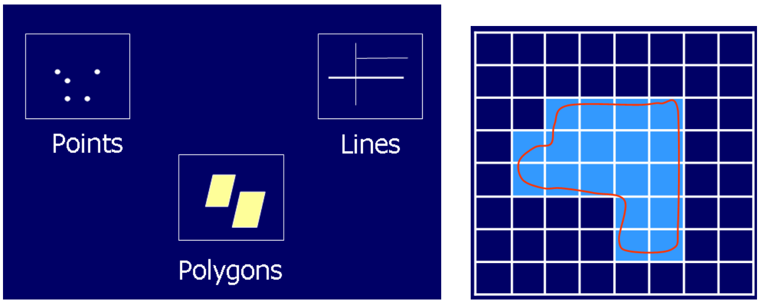


Figure: Vector and raster elements of object representations

Cartographic symbology can be applied to the vector graphic or raster cells in order to distinguish objects by their attributes.

Logical data models are structure of data organization in spatial datasets. Logical data models are used to represent spatial data within software views, for example within DBMS (Database Management System) and GIS. Logical data representations for spatial dataset can support geo-relational, relational, and object-oriented structures.

Typical spatial logical data models use geo-relational data structures. Spatial data and attributes are stored in different structures. Attributes are stored in a relational (tabular) data structure. Geometry of spatial entities are represented by points lines and/or polygons and are stored in large objects or files with a specific structure. Each spatial entity has a *common identifier* which is a *key (pointer)* to a linked *dataset* containing the attributes (tabular data) about the entity. DBMS of ArcGIS, MapInfo, etc. uses such dual architecture. Cartographical information can be represented on this level by correspondence between shapes of digital objects, its attributes and map signs (map palettes), and rules for visualization of an object's shape and symbology.

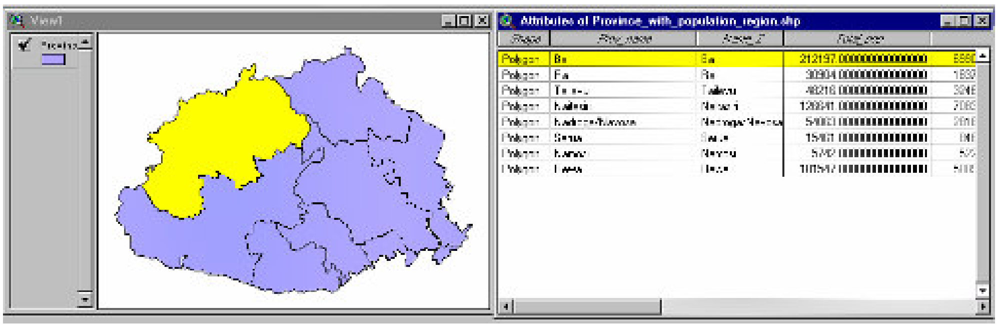


Figure: Geo-relational model – the link between graphic and tabular data

Physical data models describe data structure of records on hardware devices. Spatial data can be stored in different files or database formats that include industrial (SHP, MDB, GeoTIFF, DXF etc.), national (DLG, DEM etc.) or international (GML, ISO STDS). Designated file formats (e.g. MXD, WOR etc) can also store cartographic representations of spatial objects.

Spatial data sets can be stored in spatial databases that are structured digital collections of spatially referenced data that models the elements in reality – entities. An entity is represented in the database as a digital object. An object contains two types of information – *geometry* (a shape with location in the form of point, line, area, pixel, grid cell, or Triangulated Irregular Network) and *attributes* (qualitative and quantitative characteristics or attributes of entity).

In a database, spatial objects are usually grouped into *layers*, also named coverages or themes. One layer may represent a single entity type or a group of conceptually related entity types. For example, a layer may have only one stream or many streams, lakes, coastlines and other line elements of hydrography.

Cartographic databases include collections of spatial objects plus cartographic representations of the objects (*map symbology* and/or *rules for visualization* of objects' shapes and symbology).

More user-oriented and structured collections of spatial data are often called spatial data *warehouses.* A spatial data *warehouse* stores all types of geo-data directly into the repository at different levels of resolution, allowing end-users to query vast amounts of disparate geo-data with a very fast response time and visually recovering (or navigation) geo-data relating to a specific subject or spatial component.

Another organization for spatial data collection is a *clearinghouse*. A clearinghouse is a repository structure that collects, stores, and disseminates information, data, and metadata and provides widespread access to information from distributed warehouses.

You will learn more about spatial data storage and formats in GII-01 and GII-05 courses.

2.2.3 Digital Core Datasets in Lithuania

Key digital data sets, acting as the **core** or framework of the national GI infrastructure, are already in place or are being developed. These datasets include:

- Cartographic and geo-reference databases KDB10LT and GDB10LT at 1:10,000 scale
- Geodetic points database PPDB
- Digital database LTDBK50000-V at scale 1:50,000
- Digital topographic maps at scale 1:50,000 prepared by NATO - standards
- Digital ortho-photos at scale 1:10 000 are now covering the whole country
- Geo-reference database GDB200 at scale 1:200,000.
- Geo-reference database MapBSR at scale 1:1,000,000 integrated into the geo-reference Database of the Baltic Sea region
- Database of administrative units, settlements, streets, and addresses are under production
- CORINE land cover database, where land cover objects are classified into 44 classes

Geo-reference base dataset of the Republic of Lithuania is GDB10LT at a scale of 1:10000. It covers the entire country and is represented in 2782 map sheets (2229 map sheets are for general usage and access to 553 map sheets is restricted). GDB10LT contains the following layers: map sheet division scheme for Lithuania, points of geodetic network, geographic name annotations, state borders, rivers, streams and canals, roads, railroads, hydrographic objects.

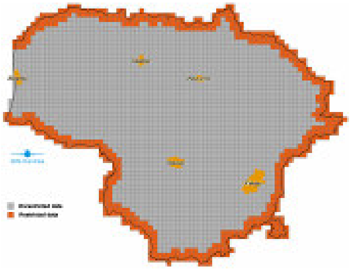


Figure: Area coverage of the state by GDB10LT geo-database map sheets (From http://www.gis-centras.lt/)

Cartographic data base of the Lithuanian topographic map KDB10LT at scale of 1:10,000 covers Kaišiadorys, Elektrėnai, Trakai and Vilnius city municipality, as well as Marijampolė city. This database consists of 279 map sheets.

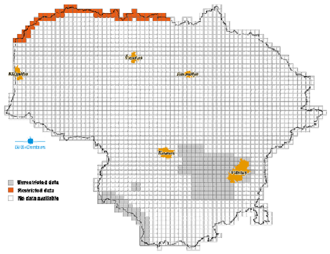
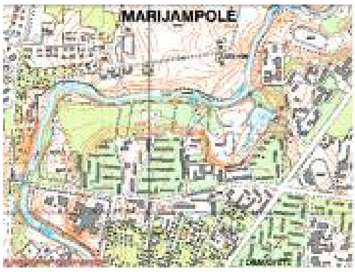


Figure: Sample of a map of KDB10LT and area coverage of the state by KDB10LT map sheets (http://www.gis-centras.lt/)

Raster TOP10LT-SR database for the Lithuanian topographic maps at scale of 1:10,000 is being prepared for major cities. TOP10LT-SR is being created using KDB10LT database TOP10LT-SR include the following objects: roads, railroads, streets, buildings, blocks of flats, individual houses and industrial territories, thematic areal objects (forests, parks, orchards, fields, water bodies, cemeteries, stadiums, etc.).



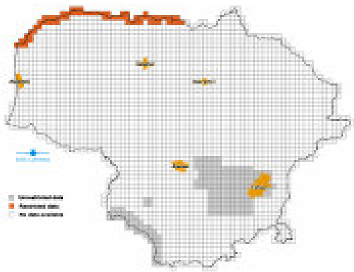


Figure: Sample of a map of TOP10LT-SR and area coverage of the state by TOP10LT-SR map sheets (http://www.gis-centras.lt/)

Lithuanian geo-information GDB200 database at scale of 1:200,000 can be effectively used as a background for thematic mapping and includes the following layers: relief, hydrology, land use, settlements, administrative boundary, transportation, energy, and nature-restricted territories.

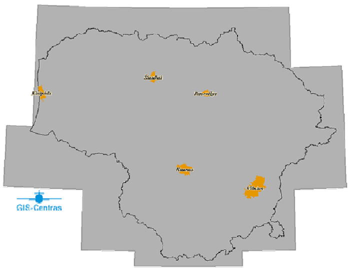
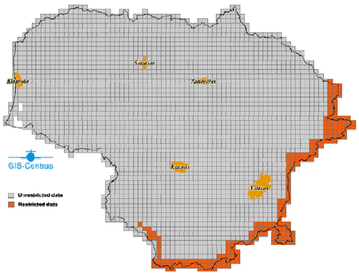


Figure: Sample of a map of GDB200 and area coverage of the state by GDB200 map sheets (http://www.gis-centras.lt/)

Digital ortho-photographic ORT10LT maps of Lithuania at a scale of 1:10,000 have spatial (ground) data resolution of 0.5 m.



2.3 Techniques for Spatial and Cartographic Data Collection and Input

Data collection and input are the operations of spatial data acquisition for mapping or data entry to a database. The creation of accurate databases is a very important part of any spatial project including SDI creation and support. Today, a fundamental issue of setting up any mapping or spatial project is getting spatial data in digital (software) formats. Up to 80% of resources in mapping and spatial projects is spent on acquiring data. This step is fundamental to getting the project working efficiently down the line. Bad data produces bad analysis and bad decisions.

There are a number of issues that arise when acquiring spatial data for projects. The following issues must be considered:

- Format of data storage (vector or raster format). It must be keep in mind that vector data is very easy to convert to raster format, but not the reverse
- Accuracy and level of details of spatial data must correspond to a specific scale of paper map. Spatial resolution (size of a pixel) has to be considered for raster data
- List of thematic layers
- Scales of attribute data collection (nominal, ordinal, interval, or ratio)
- Data preparation techniques or methods
- Hardware
- Software
- Technicians and analysts
- Time
- Money

2.3.1 Getting Spatial Data

Spatial data can come from the following four main sources:

- Existing general-reference and thematic maps (digital or hardcopies)
- Ground survey and positioning
- Remote sensing data collection
- Census and sampling, reports and publications

There are several methods or techniques that can be used for acquiring and entering spatial data into spatial storage and using this data in the latter stages (e.g. for mapping). These techniques include:

- Transfer/import data from **existing digital** sources
- Create **new digital** data from:
    - Manual digitising and scanning of analogue maps
    - Spatial data entry via geocoding

- Image data input and conversion (air and satellite imageries, DEMs, GRIDs)
- Data input from census and sampling surveying
- Data entry from ground surveys, including global positioning systems (GPS)
- Manual entry or transfer of attributive data

Data Source	Method	Equipments	Approx. Accuracy	Approx. Relative Cost of Equipments
Analog Map	Manual tablet digitizing	Digitizer, PC	0.1 mm on a map	cheap
	Semi-automatic or automatic tracing	Scanner	0.1 mm on a map	high
Aerial Photo	Analytical photogrammetry	Analog stereo plotter	10 cm	high
	Digital photogrammetry	Digital photogrammetry workstation	10 cm	very high
Satellite Imagery	Visual interpretation	Computer display, stereoscope	2-50 m	cheap
	Digital image processing	Image processing system	1-30 m	high
Ground Survey	Field measurement	Total station, GPS	1 cm	very high
Reports	Keyboard entry	Keyboard, PC		cheap

Table: Comparison of main spatial data acquiring techniques

2.3.2 Using Existing Digital Data

A large amount of information is now available in digital formats, so there are great opportunities to secure and use existing digital sources. In 1980s and 1990s most of the geo-referenced data was derived from digitising analogue maps or from direct data entry in many countries and worldwide.

To acquire some digital sources, users must contact the producers directly to gain the necessary data in a compatible format but some spatial data can be downloaded online for free. The Internet is being used, more and more, to distribute data by many organizations and agencies.

There are still several constraints when using existing digital data. These include:

- Cost of data may be very expensive
- Incompatible with some user platforms (e.g. not inter-exchangeable data format)
- Incompatible with user required standards (e.g. poor data quality, not appropriate scale, etc.)
- Restrictions for data use and sharing

The data obtained from other sources should contain meta-data and a "data quality report" from the provider. This meta-data report should describe exactly what is in the file, how the information was compiled (and from what sources), level of accuracy, data format, and how the data was checked.

Vector and raster data can be loaded or imported/exported from existing data warehouses. In order to use this data for a particular project, these data may require conversion from data format that was used in the spatial storage system. Such conversion can be requested from the data provider or can be done by the user. Some data providers are offering data in few formats.

There are many different software packages that are used to work with spatial data. Most of these software use proprietary formats. Thus, data is stored in file formats, which can only be read using certain software. For example, ESRI shape files can only be opened directly in ArcGIS and not in MapInfo. To use ArcGIS as the spatial software, all data must be converted to appropriate file formats.

There are two main approaches in the conversion of data:

- Direct translation - that is a software operation that converts data between formats physically (e.g. Universal translator from MapInfo converts data from MapInfo TAB to ESRI SHP and creates new shape files)
- Neutral format - that is proprietary or non-proprietary formats that some spatial software can read directly (e.g. BIL digital image format)

In many cases cartographic information (symbolization palettes etc) are hardly converted.

Themes	The key juridical body	The responsible institution
Reference data themes: ▪ Geodetic control ▪ Elevation data and models ▪ Ortho-imagery ▪ Land information Georeference data sets: ▪ GDB M 1:10 000, 50 000, 200 000, 1 000 000 ▪ Digital maps: ▪ CDB M 1: 10 000 ▪ CDB M 1: 50 000	National Land Service under the Ministry of Agriculture	State Enterprise National Centre of Remote Sensing and Geoinformatics "GIS-Centras"
Reference data themes: ▪ Administrative units ▪ Addresses ▪ Cadastral parcels	Ministry of Justice	State Enterprise Centre of Registers
▪ Toponymy	State Commission for Lithuanian Language at the parliament of LR	Institute of Lithuanian Language
▪ Hydrography ▪ Protected areas ▪ Physical planning ▪ Land cover	Ministry of Environment	Environment Protection Agency / State Service of Protected Areas
▪ Transport network	Ministry of Transport and Communications	Lithuanian Road Administration
▪ Forests	Ministry of Environment	State Forest Survey Service
▪ Geology	Ministry of Environment	Geological Survey of Lithuania
Municipality data on Environmental themes ▪ CDB M 1:500 – M 1:5 000 ▪ Territorial planning II and III level ▪ Utilities themes	Municipalities	Municipal Services

▪ Transportation networks		

Table: Major stakeholders of digital core spatial data in Lithuania

Sometimes datasets may be unavailable in digital format.

2.3.3 Manual Digitising and Scanning of Analogue Maps

Digitising is the transformation of information from analog format, such as a paper map, to digital format, so that it can be stored and displayed with a computer. Digitising can be manual, semi-automated (automatically recorded while manually following a line), or fully automated (line tracing).

Manual digitizing involves manually tracing lines on a map using a digitizing mouse on a **digitizing table** or on a **computer screen**. The first technique is known as "heads-down digitising" and second as "heads-up digitising". This technique is tedious and time consuming but fairly accurate, if careful.

When digitizing a hard copy map, a special table (or tablet) is used for tracing features from the map into digital format. A map is fixed to the table and a special mouse is used to trace over features. The table is connected to a computer running digitizing software, which records the movement of the mouse and creates points, lines, and polygons according to the user's controls. The table and puck acting together with the computer can locate the puck s position relative to reference information provided by the operator.

The major digitising operation steps include:

- Step 1: A map fixed to a digitizing table
- Step 2: Control points or tics at four corners of the map sheet should be digitized by the digitiser and inputted to the computer, along with the coordinates of the map's four corners.

  Geo-referencing of the map needs to be performed for each new digitising session, as well as each time the map's position is changed on the digitiser. Geo-referencing involves the conversing of coordinates of the digitiser into world coordinates. The digitising program will require the coordinates of the control points which will be used. These control points should generally be well spaced, for example near the corners of the map. The locations of control points then need to be digitised.

Once at least 4 pairs of control points are entered, the software calculates the Root Mean Square (RMS) error and shows an error to an operator. If the calculated error is less than a required error limit, a geo-reference function is enabled to register the scanned map.

- Step 3: Map information is digitized according to the map's layers and the map's coding system in either *point* mode or *stream* mode at short time intervals.

In point mode the digitising operator specifically selects and encodes those points deemed "critical" to represent the structure of the line or significant coordinate pairs. In stream mode the digitising device automatically selects points on a distance or time parameter, which can generate an unnecessary high density of coordinate pairs.

- Step 4: Editing errors such as small gaps at line junctions, overshoots, duplicates, etc. should be made for a clean dataset without errors.
- Step 5: Conversion from digitiser coordinates to map coordinates to store in a spatial database.

Major problems of map digitisation are:

- the map will stretch or shrink day by day which makes the newly digitized points slightly off from the previous points
- the map itself has errors
- discrepancies across neighboring map sheets will produce disconnectivity
- operators will make a lot of errors and mistakes while digitising

Vector data input via **scanning** and **vectorization** can be done through the three vectorisation tracing techniques:

- **Manual on-screen** digitising on raster (with snapping and tracing)
- **Semi-automatic** tracing of raster (with manual initial navigation and control, snapping, setting options)
- **Automatic** line tracing or recognition of raster (with setting options)

For **on-screen manual digitising,** a paper map first has to be scanned in a raster format, and opened within the digitising software. The processes of on-screen digitising is similar to table digitising. Rather than using a table digitiser and a cursor, the user creates the map layer up on the screen with the mouse and with referenced information as a background. On-screen digitizing may also be used in an editing session to add new features from a reference image or raster map.

The steps for manual on-screen digitising are:

- Source map or image is scanned

Maps are often scanned in order to use digital image data as a map background and/or to convert scanned data to vector data for use in a vector format. Scanning requires that the map scanned be of high cartographic quality, with clearly defined lines, text and symbols; lines of 0.1mm width or wider should be clean and clear.

- Scanned maps or images are geo-referenced

- Features can be digitized in point mode by entering vertices of a line or features can be manually traced on the screen using a mouse and software controls with automatic snapping to a raster line
- Attribute data can be entered manually

Another approach is to use a scanner to automatically convert the analogue map into computer-readable format. Automatic vectorisation procedures are used for conversion from raster to vector data, which is often called raster-vector conversion. A simple algorithm of vectorisation is explained in the figure below, in which the original image in raster format is converted to vector data through thinning and chain coding.

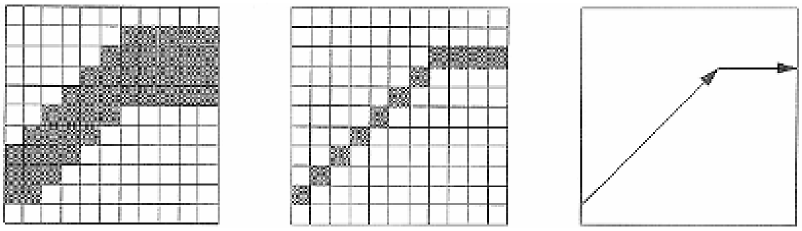


Figure: Original image   Thin ring with chain coding   Vector data

Special image recognition software is use for semi and automatic digitizing or tracing of raster data. The function of image recognition can be included in GIS functionality. For example, the ArcScan extension of ArcGIS can be used for such purposes.

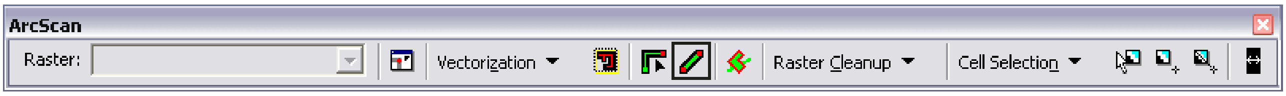


Figure: ArcScan toolbar of ArcMap

Steps of semi- or automatic on-screen digitising are:

- Map is scanned using a scanner
- Map image is converted to binary image to create lines representing features and boundaries
- Binary image is geo-referenced
- Vectorisation setting needs to be defined. These settings, in conjunction with the vectorisation options, determine how raster features are traced or extracted to create vector features (e.g., follow the centerline of raster line, minimum length of created line, minimum distance between the vertices etc.).
- Software traces features by following lines **semi-automatically** (with user interaction and control) or **automatically** (when in batch mode)

Figure: Equipment for automatic vectorisation, raster lines and derived vector lines

Semi-automatic vectorisation is designed for tracing connected raster cells. An operator has to place the tool at an appropriate start location in the raster and point to a destination location of tracing and then click to begin. The tool will then automatically create a vector line following the center line (or outlines) of the raster cells before a nearest intercession. An operator has to define a new destination location for tracing from the intersection.

Automatic vectorisation is a fully automated technique based on image recognition algorithms for converting raster data into vector features. This process relies on user input to control how to perform the vectorisation. Factors such as image resolution, amount of noise in the image, and the actual content of the scanned document all play a role in determining the outcome of the vectorisation.

There are a few issues of automatic digitizing. These include:

- Faster, but relies on human supervision
- Requires intensive manual editing
- Manual entry of annotation data
- Manual entry of attribute data

Digitising errors such as undershoots, overshoots, slivers, line spikes, etc. will always occur. Editing of digitised features may involve error correction, entering missing data, forming topology, etc.

Much of existing Lithuanian digital data was created via vectorisation processes:

- Basic topographic maps were used for creating digital ortho-photographic GDB10LT maps at scale 1:10,000
- LTDBK50000 was compiled by using existing map information and digital data bases together with up-to-date SPOT panchromatic ortho-photo images and geometrically corrected multispectral images - later updated using GDB10LT
- Lithuanian Geo-information Data Base GDB200 is created on the background of topographic maps, using GIS digitizing technologies
- Vilnius city large-scale Cartographic Data Base KDB500V was compiled by using existing large-scale topographic maps with utility information

2.3.4 Manual Data Entry

Spatial data can be entered and geo-positioned manually from printed or hand-written sources, such as reports, tables, books, etc. *Geocoding* procedures need to be applied in order to link the report data to geographic locations.

Geocoding can be done by:

- Simple coordinate pairs of information that describes the location of a feature
- Information that describes the location of a feature, such as address descriptive formats
- Locating point events along a linear network

In a typical situation, geocoding is often error-poisoned and data is often missed.

Most of spatial software (e.g. GIS) can create a geographic feature representing the event that occurs *on* or *along* a reference feature. These events are:

- XY event: X and Y coordinate pair of information that describes the location of a feature
- Address event: Address events are features that can be located based on an address matching with a street network, or other features with an address identifier such as postal code or lot number, country or region names etc.
- Existing data indirectly captured as spatial information: Street addresses, city names, or even telephone numbers. We understand street addresses, computers don't. To use locations in spatial software, operators must create coordinates. Address geocoding or address matching is the process of converting the textual description of a place into a geographic location with computer software.
- Route event: Route events are associated with a base route system. This process is called linear geo-referencing and involves dynamic segmentation of networks. Linear referencing is a measuring system for linear features such as river milage and route mileposts. Linear referencing records locate data by using relative positions along existing linear features.

2.3.5 Raster Data Input

Sources of spatial raster data can come from different sources such as:

- Satellite and air-photo imagery
- Scanned maps, photographs

On the other hand, raster can be derived from vector or other raster data. In GIS, results of rasterisation of vector data are often called GRIDs

One of most cost-effective way to capture up-to-date spatial information is remote sensing. Remote sensing can be also be used as an important data source for the development of thematic models and can be used to validate models.

In remote sensing, electromagnetic energy may be detected by sensors either *photographically* or *electronically*. Imagery refers to any pictorial representation. A photograph refers specifically to images that have been detected as well as recorded on photographic film.

Digital photogrammetic mapping uses air-photo images of a pair of overlapping photographs for compilation of topographic maps. In order to use photographic prints, scanners are used to convert images from analog photographs or maps to digital images in raster format. Digital image data are usually integer-based with one byte gray scale (256 gray tones from 0 to 255) for black and white images, and a set of three gray scales of red (R), green (G) and blue(B) for color images.

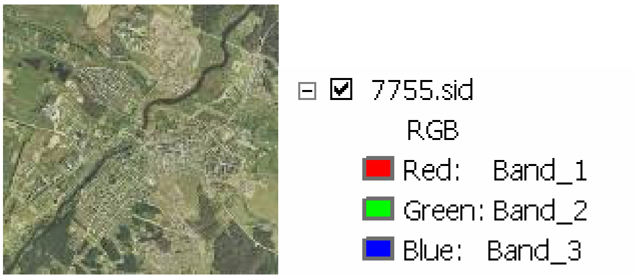


Figure: RGB color image of air-photograph

Modern satellite images are already captured and distributed in digital raster formats. These can be converted from a vendor format to a software format. Many commercial data suppliers, such as EOSAT and SPOT, provide radiometrically corrected data in a customer specified format. There are many different data formats used for storing digital remotely-sensed data. There are three major data remote sensing formats used by government and commercial data suppliers: band interleaved by pixel (BIP), band interleaved by line (BIL), and band sequential (BSQ) format.

The imagery from the following main modern commercial land observation satellites are often used today for mapping and spatial data collection:

- LANDSAT-7 provides ground resolutions of 30 meters for color images, 15 meters panchromatic
- SPOT-5 provides ground resolutions of 10 meters color, 2.5 meters panchromatic
- IKONOS provides ground resolutions of 4 meters color, 1 meters panchromatic
- QuickBird provides ground resolutions of 2.5 meters color, 0.6 meters panchromatic

Still, air photographs can have ground resolution less then 10 cm. Certainly, air photographs have better ground resolution than imagery from commercial satellites.

Figure: Landsat – 30 m ground or spatial resolution

Figure: Ikonos – 4 m ground resolution

Figure: QuickBird – 2.5 m ground resolution

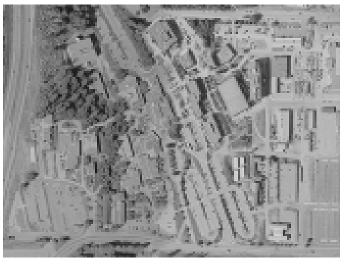


Figure: Air Photo – 20 cm ground resolution

Satellite and air-photograph images have to be corrected (*rectified* and *ortho-rectified*) in order to map the information. Ortho-imagery combines the image characteristics of an air photograph with the geometric qualities of a map. Unlike normal air photographs, distortion and relief displacement is removed so ground features are displayed in their true planimetrical correct position.

Correction processes that bring an image into planimetric view (e.g. eliminate tilt of platform and perspective view) are called *rectifications***.** Correction processes that eliminate distortions due to differences in elevations are called *ortho-rectifications***.**

Finding position of an image in any coordinate system is called *geo-referencing***.** Common methods for geo-referencing are affine transformation, polynomial equations, and rubber sheeting.

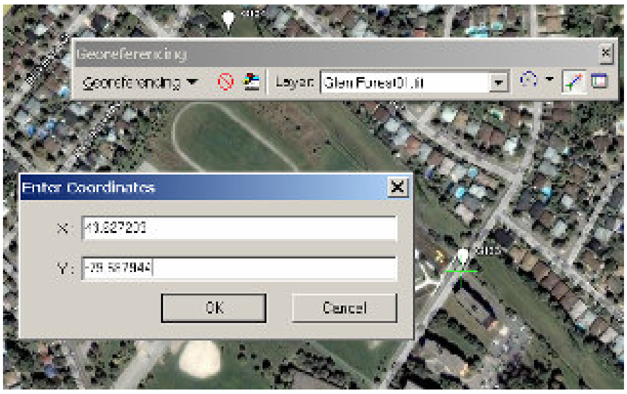
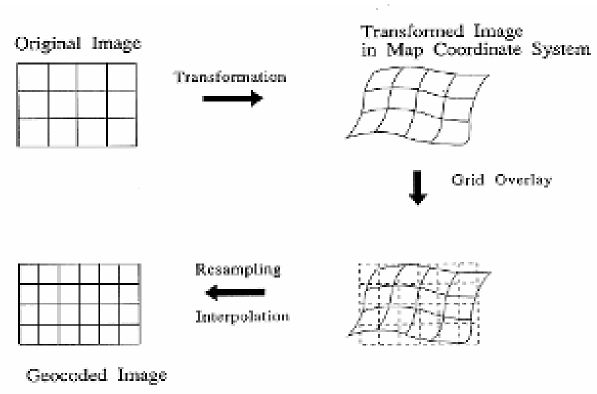


Figure: Image geo-referencing by control points in ArcMap

Any image geometric transformation creates a new grid. Data *resampling* is therefore required to fill each cell of a new grid with the value of the corresponding cell or cells in the original grid.



Another source of raster data is a GRID. Grid data represents continuous variations in value of a single attribute for each raster cell for floating GRID. For example, Digital Elevation Models (DEM). Integer GRID can be used to represent discrete features and a few attributes can represent each raster cell (e.g. land-cover GRID).

Grid data can be derived from some sources that involve computations:

- Regularization of non-uniform GPS or surveying data
- Regularization point data collected from maps
- DEM compilation using stereo images (photogrammetric techniques)
- Interpolation of vector data (contour lines)

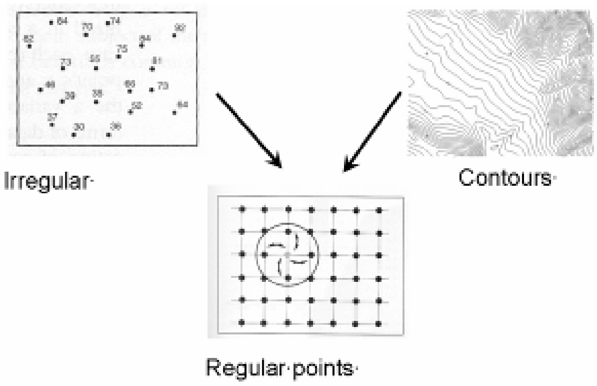


Figure: Deriving a grid

Most often, a GRID application is a Digital Elevation Model (DEM) that represents a measured elevation above or depth below sea level.

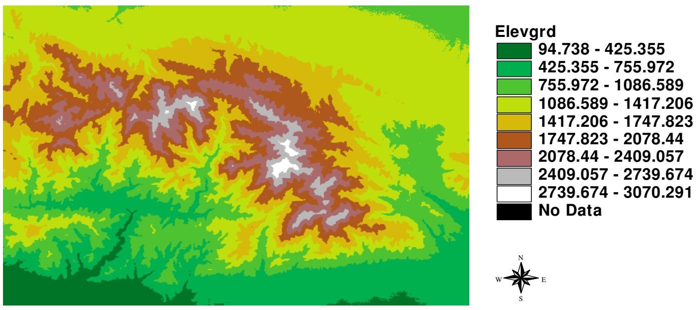


Figure: A sample of DEM

Spatial interpolation methods are used to produce regular grids. Spatial interpolation is the way continuous surfaces are derived from discrete data (digitised contours, field sampling data, etc.). *Interpolation* is the procedure of predicting the value of attributes at unsampled sites from measurements made *within* the sampled area. Predictions of values *outside* the sampled area are called *extrapolation*.

2.3.6 Sampling and Census

Sampling and census techniques can be used for collection of spatial information or information that has spatial components.

Sampling is the process of selecting individual observations intended to yield some population knowledge about objects or events. In general, the sampling process includes five stages:

- Define population
- Specify a set of objects or events for measurement purposes
- Specify sampling method for selecting objects or events from the set
- Determine the sample size
- Implement the sampling plan
- Sample and collect the data

- Review sampling process

Sampling methods of spatial data collection can be:

- Total - obtaining information about every object or event
- Random - all spatial features have the same chance of being selected regardless of location or time
- Systematic - systematic sampling involves choosing data by certain criteria such as every other square km
- Stratified - stratified sampling involves sampling sub-populations so that each sub-population has adequate representation

Ultimately, some spatial data will be lost due to the last three sampling technique used.

In general, a *census* is the process of obtaining information about every member of a population (a total sample). The term is mostly used in connection with national population and housing censuses. The United Nations recommends that national censuses be taken every 10 years.

There are also agriculture censuses (for all agricultural activities) and business censuses (for all enterprises). These censuses can be compared with subsets of a desired population.

Typically, census data can be geocoded (geo-referenced) by using vector (coordinate reference system for address matching) or raster (based on matrix or grid notation). Census information can also be aggregated for different levels of census regions (district > city > municipality > country).

Lithuania's population census answers this guestion: How many people, households, and dwellings are found in the country? The last Lithuanian census was taken between April 6-13, 2001. The results of this census is produced in tables, graphs and maps that show population, ethnicity and religion characteristics of the country. Some of this census data are broken down to the county and municipal levels. The census questionnaire and methodology can be found in the Government Statistical Department web site http://www.stat.gov.lt/ .

The Agricultural Census in Lithuania was conducted in June 2003. It captured information about land, crops, livestock, housing, agricultural equipment, the number of family members working in the agricultural sector, the number of hired employees, their working hours, the farm group by income level, etc.

2.3.7 Data Entry from Ground Surveys

The following types of data can be collected by field survey:

- Points coordinates
- Distances
- Angels
- Elevations
- Attribute data

The data entry of this data includes importation and computation, if original data comes in digital format. Manual data entry is entered using a keyboard. This topic will be discussed in more detail in the next module.

In surveying, measured angles and distances from known points are used to determine the position of other points. Surveying field data are almost always recorded as polar coordinates and transformed into rectangular coordinates.

A Global Positioning System (GPS) is a set of hardware and software designed to determine accurate locations on the Earth using signals received from selected satellites. Location data and associated attribute data can be digitally recorded and transferred for mapping, geo-referencing, and digital storage. An operator can create complex data dictionaries for attribute data collection with GPS.

Figure: GPS device

2.3.8 Attribute Data Input

The storage of spatial information can also contain attributive data in spatial locations. The attributes can be used for symbolization and annotation of maps, and can be acquired from different sources by different entering methods:

- Importing from GIS databases (ArcGIS, MapInfo, etc.)
- Importing from attributive databases (MS Access, Oracle, etc,)
- Manually entering from keyboard
- Deriving from existing data (e.g. from classification, computation, etc.)
- Importing from field observations

2.3.9 Conclusion

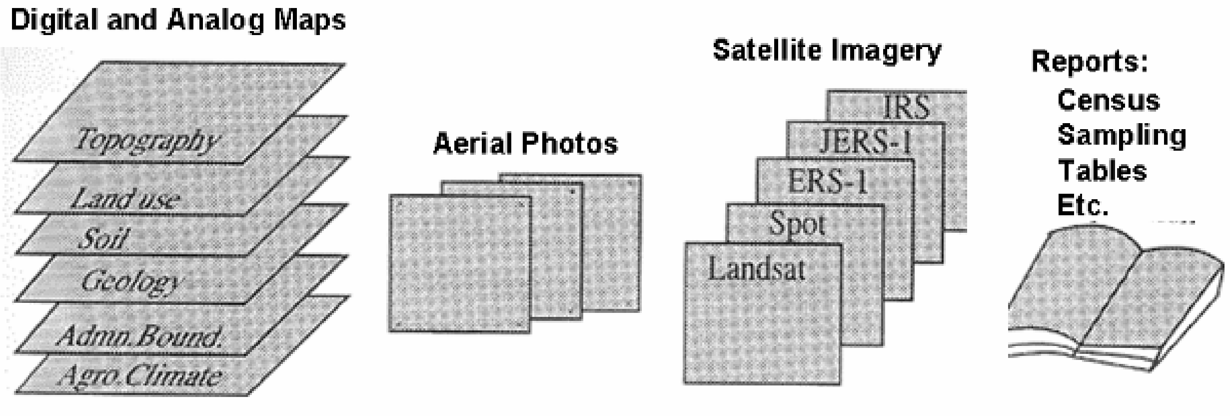


Figure: Main sources of spatial data

Module self-study questions:

- List three other methods besides digitising for spatial data input.
- Describe the role of geometric transformations in raster data input.
- List *two* methods for converting paper maps to digital data.
- Which of the following is an automatic digitising method?
  - digitising using a scanner
  - digitising using a table digitizer
  - on-screen digitising
- One can create a digital map by using a text file with *x*-, *y*-coordinates.
  - true
  - false

Required Readings:

- **Chapter 2: Geospatial Data Development: Building data for multiple uses,** Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI, Version 2.0 25 January 2004, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf
- **Data Sources for GIS,** The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/sources/sources_f.html
- **Collection of Spatial Data Sources, Data Input, Data Conversion**, **Spatial Data Accuracy** sections, Geospatial Data Clearing House Notes site, correspondent links http://corpsgeo1.usace.army.mil/geo/metadata/sec7input/datacoll.html http://corpsgeo1.usace.army.mil/geo/metadata/sec7input/datainput.html http://corpsgeo1.usace.army.mil/geo/metadata/sec7input/dataconv.html http://corpsgeo1.usace.army.mil/geo/metadata/sec7input/dataacc.html
- **Chapter 11: Finding locations,** Modeling Our World, Zeiler, M., ESRI Digital Library, 1999.
- **Satellites and sensors** section and all subsections**,** Canada Centre of Remote Sensing (CCRS), 2004 Fundamentals of Remote Sensing. Section URL: http://ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter2/01_e.php
- **The Concept of Remote Sensing**, **The Nature of Satellite Digital Data**, **Digitization sections,** Short, N., NASA, 2005, The "Short" Tutorial. Website URL: http://www.fas.org/irp/imint/docs/rst/Front/tofc.html Sections' URLs: http://www.fas.org/irp/imint/docs/rst/Intro/Part2_1.html http://www.fas.org/irp/imint/docs/rst/AppB/B2.html http://www.fas.org/irp/imint/docs/rst/AppB/B3.html

ESRI Virtual Campus Course:

- Module 2: Cartography, Map Production, and Geovisualization, Turning Data into Information Using ArcGIS 9

Assignment:

- **Assignment 2:** Geo-referencing of air-photographs

References

- [1] Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI, Version 2.0 25 January 2004.
- [2] Thematic Cartography and Geographic Visualization, Terry A. Slocum, Robert B McMaster, Fritz C. Kessler, Hugh H. Howard, 2nd ed., 2004.
- [3] Elements of Cartography, Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, 6th ed., NY: John Wiley & Sons Inc., 1995.
- [4] Modeling Our World, Zeiler, M., ESRI Digital Library, 1999.
- [5] The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/notes.html
- [6] Fundamentals of Remote Sensing**,** Canada Centre of Remote Sensing (CCRS), 2004, Website URL: http://ccrs.nrcan.gc.ca/

Terms used

- Data layers
- Digitizing
- Batch mode digitizing
- Vectorization
- Rasterization
- Geocoding
- Rectification
- Ortho-rectification
- Georeference
- Address event
- Route event
- GRID
- DEM
- Interpolation
- Sampling
- Stratified sampling
- Census

3 Ground Survey and Positioning

This module introduces elements of geodesy that are necessary to understand if you are working with spatial data for map production within a Spatial Data Infrastructure (SDI) framework. The module introduces you to coordinate systems that are used in Cartography and GIS; the importance of global and geodetic surveys; horizontal and vertical datums; transformation between datums; and the main techniques and methods of horizontal and vertical surveys. Parameters of Lithuania's coordinate systems are identified within this module.

Module Outline

- Introduction
- Coordinate Systems
- Geodesy
- Control Surveys
- Principles of Surveying, Measurement Technology
- Horizontal Surveys
- Vertical Surveys
- Total Station Surveys

3.1 Introduction

Geodesy is the ancient science of the measurement and mapping of the Earth's surface. Geodetic data and measurement can be useful in a SDI to:

- Establish geo-reference base (global coordinate system and datum) for mapping and spatial data referencing. Any spatial system like GIS has options to define a coordinate system for spatial data display and storage. Spatial data that are not spatially geo-positioned or georeferenced can by useless for practical applications
- Acquire spatial data. More precise, but expensive, methods of spatial data acquisition include ground and GPS surveys.

Data that comes from surveying generally has a different format than GIS data. Surveying measurements usually come in the form of distances, angels and height differences. Surveying or GIS software must be used to convert survey measurements into conventional spatial database representations, such as points, lines, polygons or grids. Results of the conversion can be stored within a SDI spatial database.

Geodesy has sub-disciplines based on differing application tasks, such as global geodesy, geodetic surveying and plane surveying.

- Global geodesy determines the form of the Earth that includes the Earth size and shape and external gravity field. Global geodesy measurements are used to define the horizontal and vertical coordinate systems and used for precise horizontal and vertical geo-referencing of spatial data in spatial storage systems.
- Geodetic surveying is used to define the surface of a country by coordinates of a sufficiently large number of control points by taking into consideration the overall curvature of Earth. Geodetic surveying is also used for defining local coordinate systems and establishing control points for practical needs of plane surveying.
- Plane surveying includes topographic surveying, cadastral surveying and engineering surveying. The results of plane surveying are used for topographic and cadastral mapping, engineering work and plans, and a number of other practical applications. In general, surveying is used to establish location, changes in location, relative distance between points, and measuring angles and heights in point phenomena.

Surveys have to be done in a reference **coordinate system** (i.e., geographic, global rectangular or local etc.).

3.2 Coordinate System

Coordinate systems can define points in n- dimensional space. In surveying and cartography, it is two-dimensional or three-dimensional space.

René Descartes introduced systems of coordinates based on orthogonal (right angle) coordinates. These two and three-dimensional systems used in analytic geometry are often referred to as *Cartesian* systems. In this system, a point is referenced by abscissa and ordinate.

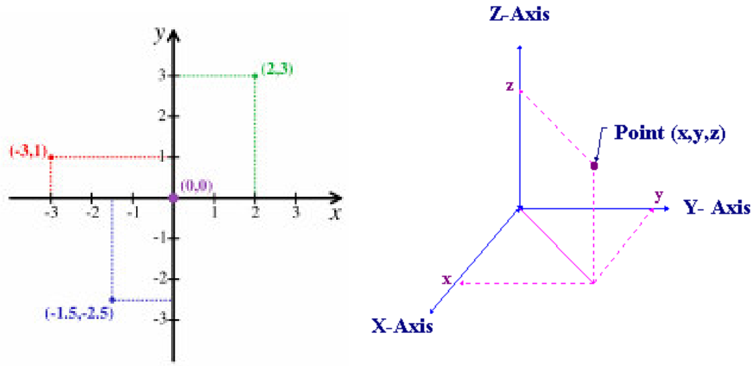


Figure: Two and three-dimensional Cartesian coordinate systems

Systems based on angles from baselines are referred to as *polar* systems. In this system, a point is referenced by radius and angle(s) of a directed line.

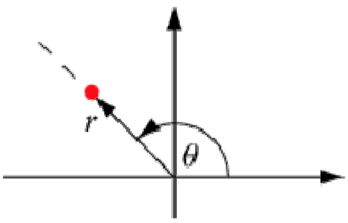


Figure: Polar Coordinates

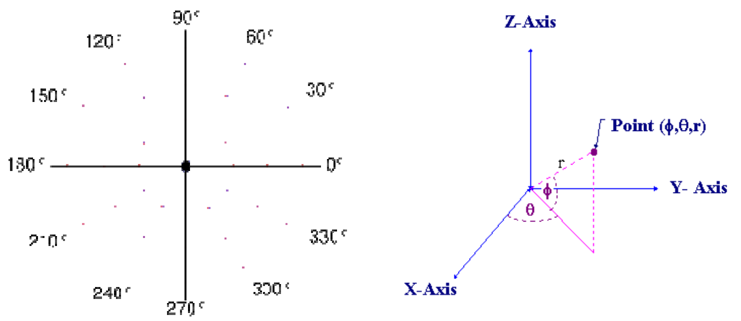


Figure: Two and three dimensional polar coordinate systems

In fields of geodesy, surveying and cartography both these coordinate systems are used.

3.2.1 Coordinate Systems Used in Geodesy, Cartography and GIS

The following coordinate systems are most often used for geo-referencing :

- Global Cartesian coordinates (X, Y, Z): a system for the entire Earth. The starting point of this coordinate system is the center of the Earth. This coordinate system is extremely cumbersome and difficult to relate to other locations when reduced to two dimensions

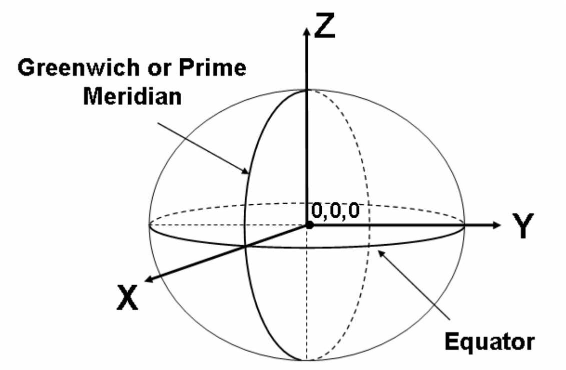


Figure: Global Cartesian coordinate system

- Geographic coordinates (φ, l, h): Uses an angular unit of measure for φ and l. The φ-latitude and l-longitude are related to a particular earth figure, which may be a sphere or an ellipsoid. Parameters of geographic system are the prime meridian and datum. Geographical coordination is generalized concept of geodetic (discussed below) and astronomical coordinate (based on astronomic observations) systems.
  - The astronomical coordinate systems are based at one reference point in space with respect to which the positions are measured, the *origin* of the reference frame (typically, the location of the observer, or the center of Earth, the Sun, or the Milky Way Galaxy). Any location in space is then described by the "radius vector" or "arrow" between the origin and the location, namely by the *distance* (length of the vector) and its *direction*.

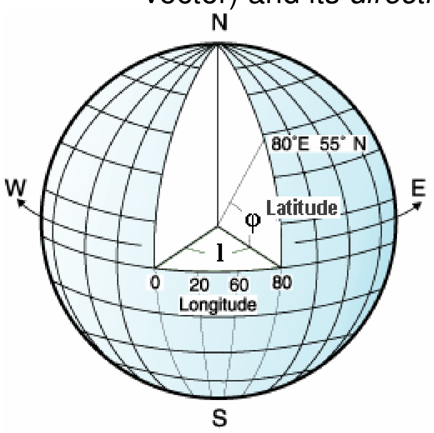


Figure: Geographic coordinate system

- Projected coordinates (x, y, z): An area of the Earth's surface is projected into Cartesian or planar coordinates. We will discuss projected coordinate systems in the following module.

The z-coordinates (heights) in global and projected coordinate systems are defined **geometrically**; in a geographical coordinate system, the z-coordinate is defined **gravitationally.**

The **geographic coordinate system** is based on a set of imaginery lines that intersect one another and encompass the Earth's surface. Determining location and direction using this system is based on two key lines - the prime meridian and the equator.

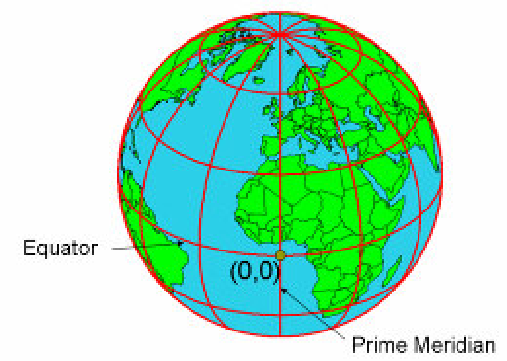


Figure: The meridians and parallels in geographic coordinate system

Geographic coordinates are useful in determining location and orientation and identifying changes in location or relative distance between points on a spheroidal surface. The concept of latitude and longitude (φ, l) and vertical distance the above the geoid (z) are the basis for any expression of geographic coordinates on the Earth's surface.

The basic elements of a geographic coordinate system:

- Equator is formed by the intersection of a plain bisection of the Earth at right angles to the axis of rotation
- Meridians are formed by the intersection of vertical planes passing through the center of the Earth and the sphere or spheroid (ellipsoid)
- Latitude (φ) and longitude (l) are defined using an ellipsoid, an ellipse rotated about an axis
- Elevation (z) defined using geoid, a surface of constant gravitational potential
- Earth datums define standard baseline values of the ellipsoid and geoid (more on this later)
- Latitude is measured in degrees north or south of the equator
- Longitude is measured in degrees west or east of the prime meridian
  - The prime meridian passes through the Royal Observatory at Greenwich, England. This base meridian is now internationally recognized

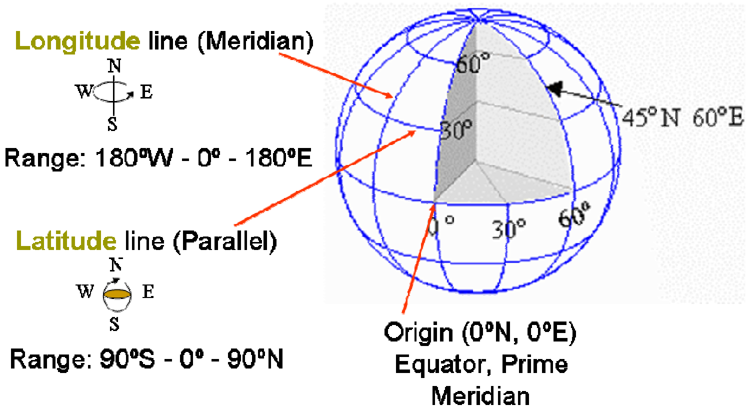


Figure: Elements of the geographic coordinate system

Figure: Greenwich, England

In spatial information and mapping software, like GIS, it is easier to use decimal degrees (DD) than degrees, minutes, and seconds (DMS) to represent geographic coordinates. Letters like "E" and "W" are used define locations east and west of the prime meridian, respectively, in each hemisphere. In the decimal degree system, positive and negative signs are used to identify locations. For example,

the geographic coordinate of Vilnius in degrees minutes and seconds (DMS) is: 25° 16' 00'' W and 54° 41' 00'' N. In decimal degrees (DD) it is: +25.276 and 54.689.

Distance on the Earth's surface can be measured by using spherical trigonometry equations of sine or cosine law:

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}$$

$$\cos a = \cos b \cos c + \sin b \sin c \, Cos \, A$$

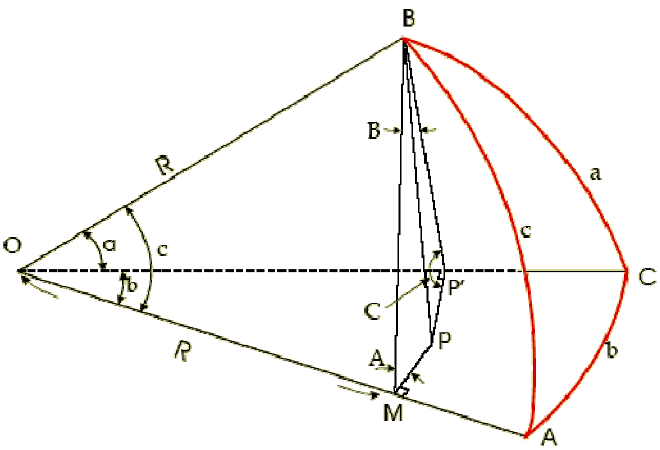


Figure: Measurement of distance on the Earth's surface

3.3 Geodesy

Geodesy is the science that studies the size, shape and gravity field of the Earth based on the fact that the Earth is flattened into an ellipse and is not a true sphere. In geodetic surveying, the computation of the geodetic coordinates of points are performed on an ellipsoid, which closely approximates the size and shape of the Earth in the area of the survey. The ellipsoid is a mathematically defined regular surface with specific dimensions.

3.3.1 Ellipsoid

An ellipsoid (also known as a spheroid) is basically a 3-dimensional sphere that is flattened at the poles. In geodesy, the geographic coordinates are not used to reference the actual shape of the Earth. Geodesy is based on a perfect ellipsoid that approximates the Earth's actual shape.

An ellipsoid is that part of a geographic coordinate system that defines the shape of the Earth's surface and is used for determining latitude and longitude**.** The main parameters of an ellipsoid are:

- a - major axis
- b - minor axis
- f - flattening of the ellipse $f = \frac{a - b}{a}$ $f \approx \frac{1}{300}$
- e - 1st eccentricity of an ellipse $e = \frac{\sqrt{a^2 - b^2}}{a}$

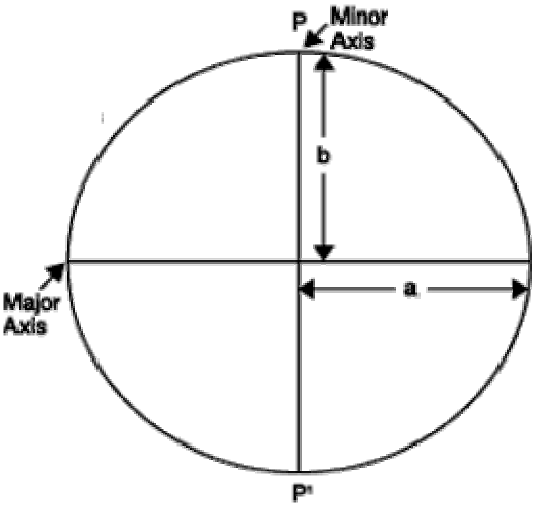


Figure: An ellipsoid

Parameters of the ellipsoid that are used to define geo-referenced data within Lithuania's database system are:

- Name - GRS_1980: international Geodetic Reference System

- *a* semi-major axis - 6378137,0
- *b* semi-minor axis - 6356752,31414035610
- 1/f flattening - 1/298,257222101000020000

3.3.2 Geoid

Another important surface involved in geodetic measurement is the geoid. The **geoid** is an even closer approximation of the shape of the Earth than an ellipsoid. The geoid coincides with that surface to which the oceans would conform over the entire earth if free to adjust to the combined effect of the Earth's gravity and the centrifugal force of the Earth's rotation. This surface has a constant gravity potential to which the direction of gravity is always perpendicular.

As a result of the uneven distribution of the Earth's mass, the geoidal surface is irregular and since the ellipsoid is a regular surface, the two will not coincide. The separation of these two surfaces is referred to as geoid undulations, geoid heights, or geoid separations. The actual measurements made on the surface of the Earth, using certain instruments, defines the geoid.

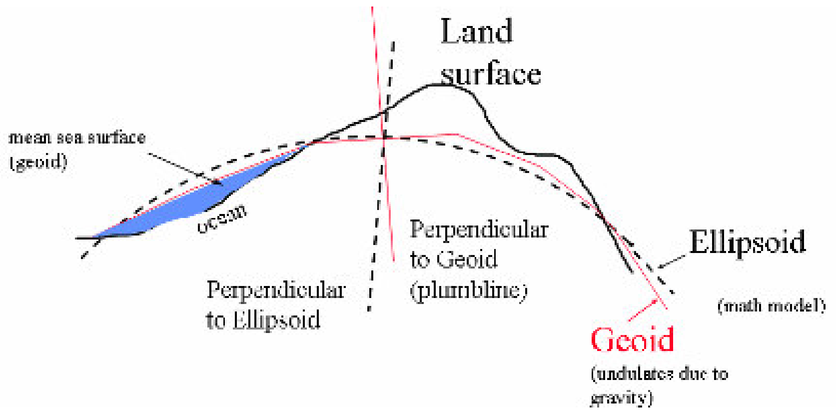


Figure: Land, Ellipsoid and Geoid surfaces

The angle between a plumb line, which is perpendicular to the geoid (called "the vertical") and the perpendicular to the ellipsoid (called "the normal") is defined as the deflection of the vertical. This angle can reach 3'' and more. Accordingly, the position of point that is determined in astronomic and geodesic coordinate systems can differ in vicinity about 30 m and more.

3.3.3 Physical Geodesy

Physical geodesy utilizes measurements and characteristics of the Earth's gravity field to deduce the shape of the geoid and the Earth's size. With sufficient information regarding the Earth's gravity field, it is possible to determine geoid undulations, gravimetric deflections, and the Earth flattening.

Two distinctly different types of gravity measurements are made:

- Absolute gravity - the value of acceleration of gravity can be determined at the point of measurement directly
- Relative gravity - the differences in the value of the acceleration of gravity as measured between two or more points

Physical geodesy techniques are used to define ellipsoid models in order to approximate the geoid and provide the basis for determining accurate location (horizontal) and elevation (vertical). However, different ellipsoids must match the geoid in different geographic regions; therefore, there are a large number of ellipsoids used in different areas.

Name of Ellipsoid	a	f
Geodetic Reference System 1980	6378137.0	298.257222101
International 1924	6378388.0	297.0
Clarke 1866	6378206.4	294.9786982
Krassovskij 1940	6378245.0	298.3
Everest 1830	6377276.345	300.8017
WGS 84	6378137.0	298.257223563
WGS 72	6378135.0	298.26

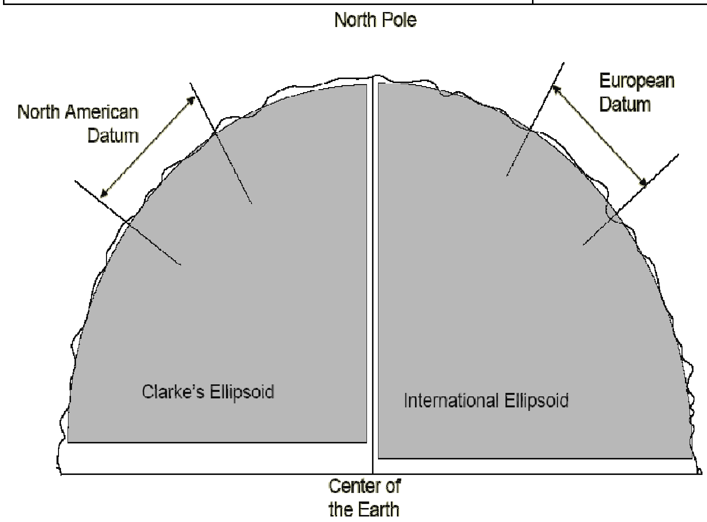


Figure: Ellipsoids used in different areas

3.3.4 Datums

The horizontal and vertical datums are elements of common reference systems used to uniquely determine positions anywhere on Earth

- Horizontal datum is a particular application of an ellipsoid
- Vertical datum is a level from which elevation is measured

Datums provide a frame of reference for measuring locations on the surface of the Earth. They define the origin and orientation of lines of latitude and longitude and determination of elevation.

Horizontal Datum

A horizontal datum defines the position of the approximated spheroid. There are two types of horizontal datums:

- An **earth-centered** or **geocentric** datum - uses the Earth's center of mass as the origin
    - based on satellite orbital data

- A **local** datum - aligns its spheroid to closely fit the Earth's surface in a particular area
    - based on geodetic ground survey measurements

Recently, an **Earth-centered,** or **geocentric datum** (e.g. WGS1984) uses the Earth's center of mass as the origin. Some geocentric datums are global and intend to provide good average accuracy around the world. For example, WGS 1984 serves as the framework for locational measurement worldwide. An Earth-centered, or geocentric global datum (e.g. WGS1984) uses the Earth's center of mass as the origin.

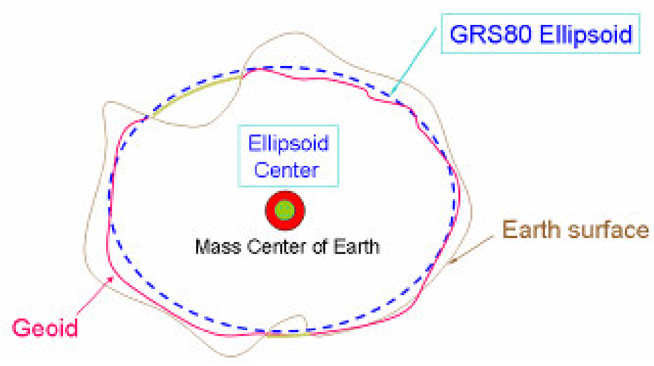


Figure: Origin of global geocentric datum

A geocentric national datum aligns its ellipsoid so that it closely fits the Earth's surface in a particular geographic area (e.g., region, country). For example, NAD83 serves as the framework for locational measurement in the North America; D_Lithuania_1994 is defined as the location and orientation using the GRS ellipsoid for the Republic of Lithuania.

A horizontal Earth datum is defined by a specific ellipse with an axis of rotation (a geocentric datum) or an ellipsoid with a base ground point where the ellipsoid is applied (a local datum). A geocentric datum often describes the position of an ellipsoid by shift parameters dx, dy, dz and rotations (rx, ry, γz) in 3-dimensional space, relative to WGS84 global datum.

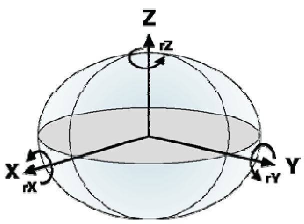


Figure: Rotations rx, ry, γz parameters

A **local datum** (e.g. NAD27, SD42 Pulkowo) aligns its ellipsoid to closely fit the Earth's surface for a particular area. This datum specifies by ellipsoid used and location of an ellipsoid relative to a base ground point where the ellipsoid is applied. The **local origin ground point** of the datum is matched to a particular position on the surface of the Earth and all other points are calculated from it. For example, NAD27 uses the Clarke 1866 Spheroid which minimizes error between the spheroid and the geoid at Meades Ranch, Kansas. This is the geographic center of continental U.S.A.

The coordinate system origin of a local datum is not at the center of the Earth. A local datum (e.g. NAD27, SD42 Pulkowo) aligns its ellipsoid to closely fit the Earth's surface in a particular area.

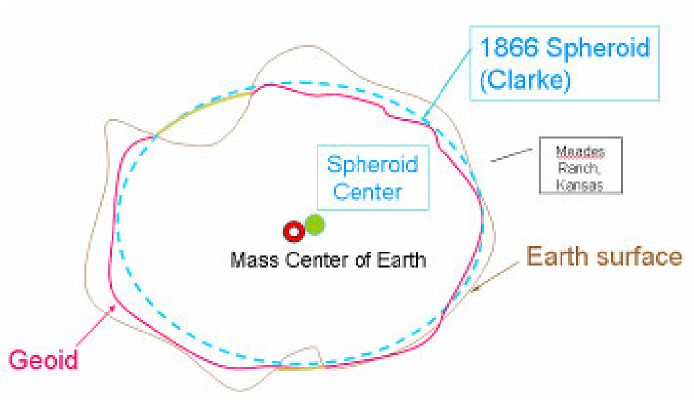


Figure: Origin of local datum

Datums are important because coordinates based on different datums will place features that are coincident on the Earth's surface at different locations on a map. Thus, there are significant differences in the horizontal positions of points with *identical* coordinates based on different datums. These differences can cause as much as 100-200 meter shifts.

Because datums are tied to a specific ellipsoid, it is equally important to know what ellipsoid was used to determine the geographic coordinates of the data you want to use.

The official horizontal data that is used in the Republic of Lithuania is Lithuanian Horizontal Datum - D_Lithuania_1994 with parameters GRS80 for spheroid, and shift parameters dx, dy, dz, rotations rx, ry, γz, and scaling is equal to zero (the center ofGRS80 spheroid or ellipsoid coincides with WGS1984 datum).

A horizontal coordinate **geodetic reference system** is defined by the datum and spheroid. Two of these parameters are required to specify a geodetic coordinate reference system. The units of linear measurement are another parameter that must be specified for any geodetic coordinate system. Many coordinate systems also have additional parameters that must be specified to complete the coordinate system.

The current Lithuanian National Geodetic Coordinate System is Geographic Coordinate System GCS_**LKS**_1994 that is based on common European Coordinate System ETRS 89. Parameters for this system are:

- Angular Unit: Degree (0,0174532925199433299)
- Prime Meridian: Greenwich (0,0)
- Datum: D_Lithuania_1994
- Spheroid: GRS_1980
- Semimajor Axis: 6378137,0
- Semiminor Axis: 6356752,31414035610

- Inverse Flattening: 298,257222101000020

A former geodetic reference system was SD42. It was based on Pulkowo datum and the Krassovskij 1940 ellipsoid. The topographic maps of the Soviet era were created in a SD42 reference system.

Name	Origin point	Spheroid	Parameters						
			dx [m]	dy [m]	dz [m]	rx ["]	ry ["]	rz ["]	ds [ppm]
LKS94 (ETRS89)		GRS80	0	0	0	0	0	0	0
SD42	Pulkowo	Krassovskij 1940	-40.595	-18.550	-69.339	2.5080	-1.8319	-2.6114	-4.3

Table: Current and former national geodetic reference system parameters used in Lithuania

Geographic Coordinate Systems Transformations

Geographic coordinate system transformations involves changing the underlying datum, including the spheroid. Several methods can be used, each having differing levels of accuracy (from centimeters to meters) when transforming between datums. These methods can be categorized into two groups:

- **Grid-based** methods
- **Equation-based** methods

Grid-based methods of a geographic transformation convert geographic coordinates (longitude/latitude) directly. Equation-based methods convert geographic coordinates to geocentric (X,Y,Z) coordinates, transforming the X,Y,Z coordinates, and converting the new values back to geographic coordinates.

Several surveying and GIS software have capabilities to support both types of conversion. For example, ArcGIS supports both types of geographic coordinate system transformations between datums.

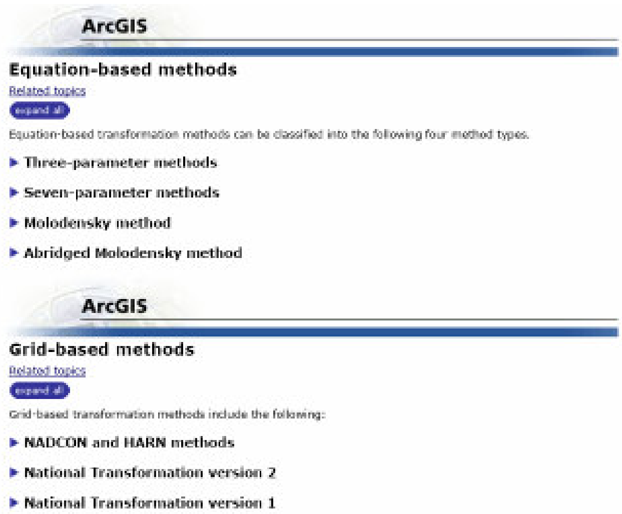


Figure: The ArcGIS options for geographic coordinate system transformations between datums

Equation-based methods convert:

- Original longitude, latitude and elevation into X, Y, Z
- Known X, Y, Z offsets of datums, transforms from old X, Y, Z to new X, Y, Z
- New X, Y, Z to longitude, latitude and elevation of new datum

There are few equation-based methods such as Geocentric translation (three-parameter), Coordinate frame methods (Bursa–Wolf or seven-parameter methods), Molodensky methods, and Abridged Molodensky methods.

Grid-based methods use a **grid** of differences to convert values directly from one datum to another:

- The area of interest is divided into cells
- The different values in decimal seconds are stored in two files: one for longitude and the other for latitude. A bilinear interpolation is used to calculate the exact difference between the two geographic coordinate systems at a point
- These methods are potentially the most accurate (NAD27 to NAD83 accurate to ~0.15m for continental US)

Geodetic Surveying

The main tasks of Geodetic Surveying are to define the surface of a country by:

- Coordinates of a sufficiently large number of control points that
  - can be used for local horizontal datum adjustment
  - can be used for plane surveying
  - include vertical and gravimetric surveys that can be used for local vertical datum adjustment
- Most work in this area is being done by GPS today.

Vertical Datum

Vertical datums are quantities that serve as a reference to accurately define vertical positions. A vertical datum defines the “zero reference” point for elevation (z). For example:

- Baltic Height System Vertical Datum
- North American Vertical Datum of 1988 (NAVD88)

To define a vertical datum, the gravity anomalies between the ellipsoid and the geoid, which are relatively constant, have to be taken into account. These anomalies then need to be mapped.

Earth's gravity field is not the same everywhere due to the fact tat material that makes up the Earth fluctuates from place to place. Measuring gravity field differences make it possible to help figure out the Earth's shape - the **Geoid**.

Mean Sea Level (same as the **Geoid**) is the average level of the ocean surface halfway between the highest and lowest levels recorded. Sea “Level” (geoid) is not a constant level; it is as much as 85 -105 m of relief globally.

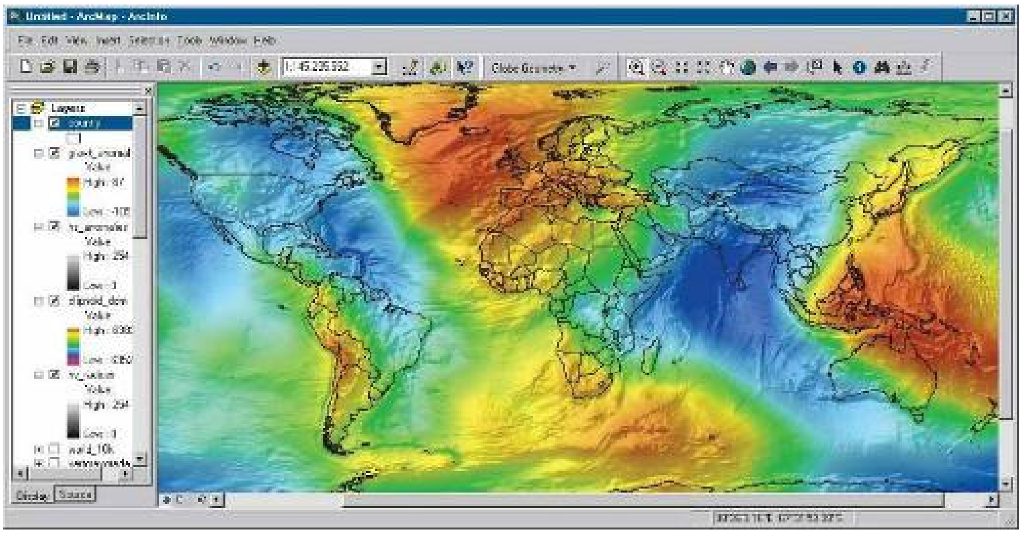


Figure: EGM96 heights (= Geoid – Ellipsoid) range from a low of - 105 m (blue) to a high of 87 m (red). Source: http://www.esri.com/news/arcuser/0703/geoid1of3.html

In traditional vertical surveys, mean sea level is used as a plane upon which surveyors can reference or describe the heights of features on, above, or below the ground. However, today, there are two ways to define height:

- From **Mean Sea Level (MSL)** or from **Geoid**
  - Traditional vertical ground surveying
- From **Ellipsoid** (**HAE -** Height above Ellipsoid)
  - Datum used by most GPS receivers

From the ground surveys, surveyors cannot directly observe the geoid or ellipsoid. So traditionally, the starting point for measuring heights is Mean Sea Level points established at coastal locations.

Starting from these points the heights of points on the Earth's surface can be measured using leveling techniques (e.g. use of a plumb bob to establish a horizontal leveling position).

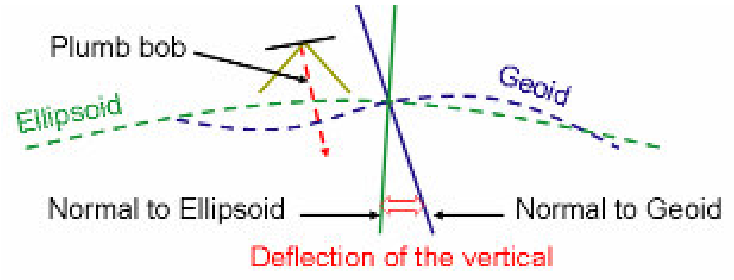


Figure: Use a plumb bob to establish a horizontal leveling position

Today surveyors use GPS receivers to measure height above an ellipsoid. In order to use both types of vertical measurements, we need to connect satellite height measurements with coastal vertical survey measurements. For these, we have to know the geoid surface (geoid height) or measure the height of the sea surface (via satellites) and link that with coastal surveys on land to determine the geoid. The equation that links these height measurements is:

h (GPS) = H (leveling) + N,

or

H (leveling) = h (GPS) – N,

where N = geoid height, h = heights above ellipsoid and H = heights above MSL or geoid (or orthometric heights or elevations)

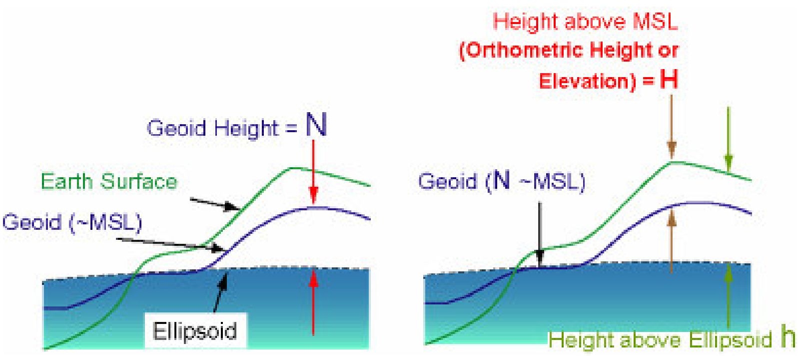


Figure: Geoid (N), Ellipsoid (h) and Elevation (H)

To convert HAE (height above ellipsoid) to Orthometric Height (elevation above MSL), we need an accurate model of the geoid height (e.g. N.G.S. GEOID99 has 1 x 1 minute grid spacing). By using the above equation, we can compute the difference between the HAE and Geoid heights. The current models allow conversions accurate to ~ 5 cm. Conversions that are more precise require a local gravity survey.

The Baltic Height System is used in Lithuania for vertical datum. Elevation is surveyed from Kronstadt zero or surface close to the average sea level (geoid, actually quasigeoid). Present

Lithuanian height reference is a leveling network which, due to many destroyed benchmarks and geodynamic activities, does not correspond to contemporary needs (Source: http://72.14.253.104/search?q=cache:QGMzp8yPl0wJ:www1.apini.lt/includes/getfile.php%3Fid%3D232+Lithuanian+Baltic+height+system&hl=en&ct=clnk&cd=1&gl=ca).

Control Surveys

Control survey refers to the establishment of a geodetic control network. A control network is a series of well-spaced and interconnected markers in the ground that have accurately determined positions, or coordinates, and elevations.

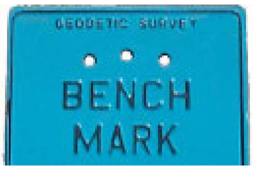


Figure: A control network marker

Control networks provide fixed or anchor points on which to base or "reference" surveys. There are three types or levels of control networks:

- Horizontal Control Network
- Vertical Control Network
- Both Horizontal and Vertical Control Network that accurate define positions and elevations

These control networks can be further categorized by accuracy levels under:

- Primary Control
- Secondary Control
- Local Control

3.4 Principles of Surveying, Measurement Technology

Surveying is the art and science of measuring the surface of the Earth and its features. Survey measurements can be divided into two categories - **horizontal** and **vertical**. These use respective horizontal and vertical datum as points of origin.

In general, there are three types of measurements that can be undertaken in surveying methods:

- Distances
  - Different instruments and technologies can be used for measuring distances: mechanical tools, electronic measuring instruments (EDM), GPS, Remote Sensing (RS) techniques, etc.
- Directions
  - Measured by magnetic compass, theodolite, gyrocompass, radio compass, RS, etc.
- Heights
  - The following techniques can be used: leveling, GPS, inertial system, RS, etc.

Few surveying techniques and methods are used for horizontal and vertical measurement. The most common techniques employed include:

- Theodolite or transit surveys - used for measuring angles and distances and for cacluating positions and heights
- Differential leveling - used for measuring precise elevations and determining relative to the vertical datum by measuring elevational differences between two or more points
- Total station surveys - similar to a theodolite survey; a total station is a theodolite and EDM (electronic distance measurement) wrapped up in one instrument
- Inertial systems - a gyro stabilized platform that starts at a point of known position and elevation
- GPS - a constellation of Global Positioning System (GPS) satellites orbiting the Earth - used to determine the horizontal and vertical position(s) of GPS ground receivers
- LIDAR (LIght Detection and Ranging) are lasers on board an aircraft that 'scan' the ground surface to calculate distances
- Photogrammetry or Remote Sensing - uses stereographic pairs of photographs to indirectly measure objects on the ground and then calculate point coordinates.

3.4.1 Horizontal Survey Methods

Horizontal surveying can be conducted by using the following methods:

- Triangulation
- Trilateration
- Traverse

- Free station
- Resection

Triangulation consists of the measurement of the angles of a series of overlapping triangles, polygons, or quadrilaterals. The lines of this structure are tied together. All triangulation stations are at the vertices of the triangles. The sides of the triangles can vary in length. All triangles' angles are measured precisely. Instrumental errors are either removed or predetermined. More rigorous procedures may be employed to reduce observational errors.

The scale of the network is controlled by measuring the lengths of certain lines called **base lines.** The latitude, longitude and azimuth of one point on the base line are found and from this, the latitude and longitude of all other points in the network can be computed. Except for the base lines, the lengths of all sides are calculated. Thus, triangular figures have to be geometrically strong and permit each side of the triangle to be calculated two ways (stations have to be selected under the same conditions).

The known data for triangulation are:

- Length and azimuth of base line AB
- Latitude and longitude of points A and B

The measured data for triangulation are:

- Angles to new control points

The computed data for triangulation are:

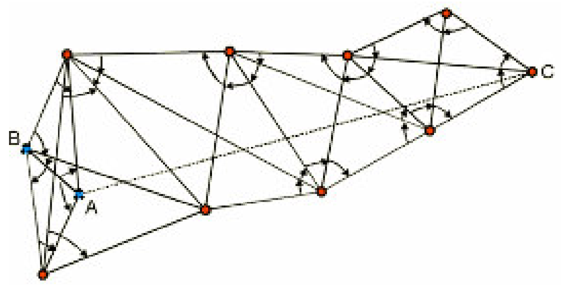
- Latitude and longitude of point C and other new points
- Length and azimuth of line AC
- Length and azimuth between any two points
- 

Figure: Triangulation with AB as a base line

Trilateration determines a position of a point on the Earth's surface with respect to two other points by measuring the distances between all three points. Trilateration uses the known locations of two or more reference points, and the measured distance between the subject and each reference point. Different techniques can be used for trilateration. These include use of the theodolite and EDM, radio, aero-distance (e.g. Shoran) surveys etc.

The known data for trilateration are:

- Length and azimuth of base line AB
- Latitude and longitude of points A and B

The measured data for trilateration are:

- Length of all triangle sides

The computed data for trilateration are:

- Latitude and longitude of point C and other new points
- Length and azimuth of line AC
- Length and azimuth between any two points

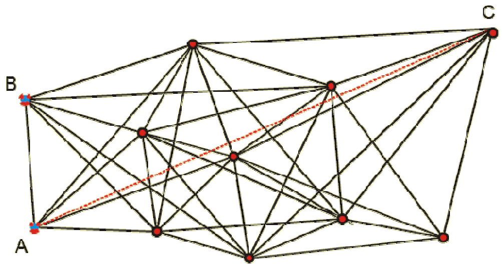


Figure: Trilateration with AB as a base line

Transverse survey is a sequence of instrument setups that start at a known location and end at another known location, with the intermediate setups being at points with unknown coordinates. Distances along each line are measured and an angular measurement is taken at each traverse point. With stand-alone EDM devices or theodolites, the traverse is done in two dimensions. With total station equipment, the traverse can be done in three dimensions,

There are two general classes of traverses:

- *Open* traverse originates at a known point and terminates at an unknown point. There is no way to check the values in this type of traverse and therefore should not be readily used
- *Closed* traverse originates at a known point and closes on another point of known position. Mis-closure (or closure error) is the difference between the computed endpoint coordinate and the known endpoint coordinate.

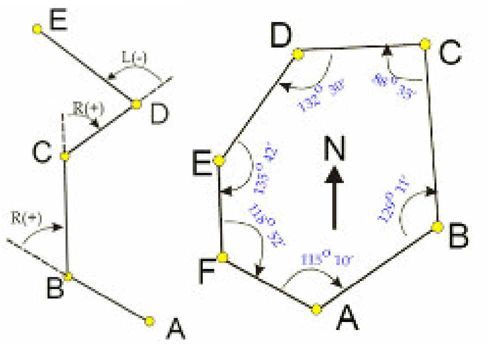


Figure: Open and closed traverses

In the **free station** method of measurement, coordinates are computed for an instrument setup at an unknown location. Distances and observations on the horizontal circle to at least two reference points need to be observed. For **resection,** coordinates are computed for a setup location if known at least three visible reference points and their horizontal circle readings (angles).

For most surveying methods, it is necessary that the number of observed measurements be greater than the number of computed parameters. This condition allows the use of **least squares adjustment** (LSQ) computation techniques to control accuracy of measurements and make adjustments for computation solutions. The LSQ methods define a best-fit solution by finding a minimum for the sum of the squares of the measurement residuals.

There are different techniques for adjusting a traverse:

- *Compass* correction - the corrections are reflected in each distance and direction value
- *Transit* correction - the technique favors the direction measurements over the distance measurements
- *Crandall* correction - this adjustment will preserve all the direction measurements and will alter only the distance measurements

There are surveying computation techniques that can be applied to survey measurements. Thus COGO (COordinate GeOmetry) computers is a set of algorithms for converting survey data (bearings, distances, and angles) into coordinate (GIS) data for mapping and storage within a SDI.

COGO computation can be done with automated mapping software that is used in land surveying. Automated mapping software include different methods to make adjustments and that calculates locations using distances and bearings from known reference points. For example, total surveying station software often includes COGO functions.

GIS software also has a simple tool for the input and processing of COGO and survey measurements (traverse, resection, etc.) that can include, for example, the following simple COGO computations:

- Delta XY - compute coordinates for a survey point based on a known difference in coordinates from a given start point
- Direction distance - the calculation of a coordinate from an existing coordinate using known distance and direction values

- Deflection-angle distance - compute a new coordinate by defining a deflection angle offset, based on a reference direction, and a distance from a known point

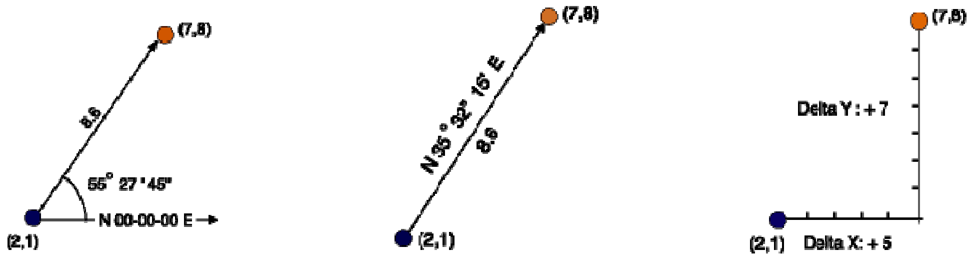


Figure: Simple COGO computations

Figure: ArcGIS Desktop tools for COGO and surveying computations

3.4.2 Vertical Surveys

Vertical control may be established by many different methods (including trigonometric and barometric leveling, and GPS). The primary vertical control network is established using differential leveling.

Figure: Leveling instrument

Leveling permits the precise measurement of elevational differences between known and unknown points on the Earth's surface. The computation of elevation of an unknown point relative to the vertical datum can be used to establish a bench mark.

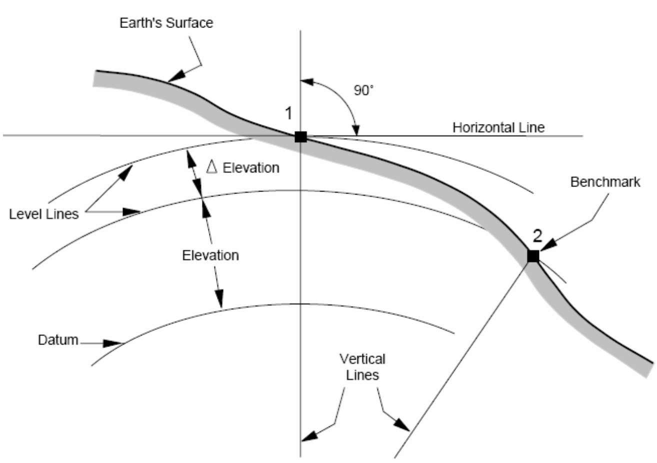


Figure: Leveling concept

In **differential leveling,** precise elevations are determined relative to the datum by a process known as differential leveling. Points established are called bench marks (BM) or temporary bench marks (TBM). Rods are held vertically on an “A” point of known elevation and on a turning “B”point.

The change in elevation is calculated by subtracting the Sum (B.S.) - Sum (F.S.) = D Elevation. Accuracy is related to the difference between forward and backward leveling.

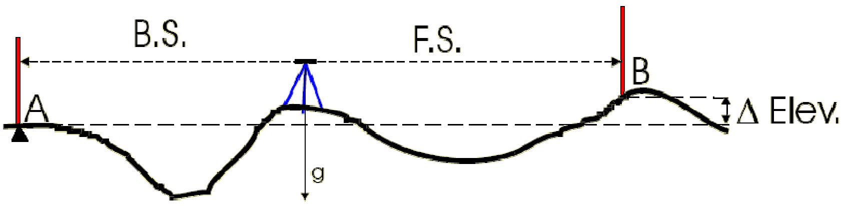


Figure: The differential leveling approach

In **trigonometric leveling,** the vertical angle and either the horizontal or slope distance between two points are known. The difference in elevation between the points can be calculated based on the fundamentals of trigonometry – Elev $_{B\ point}$ = Elev $_{A\ point}$ + H.I. + cos(z) * S – Rod(r)

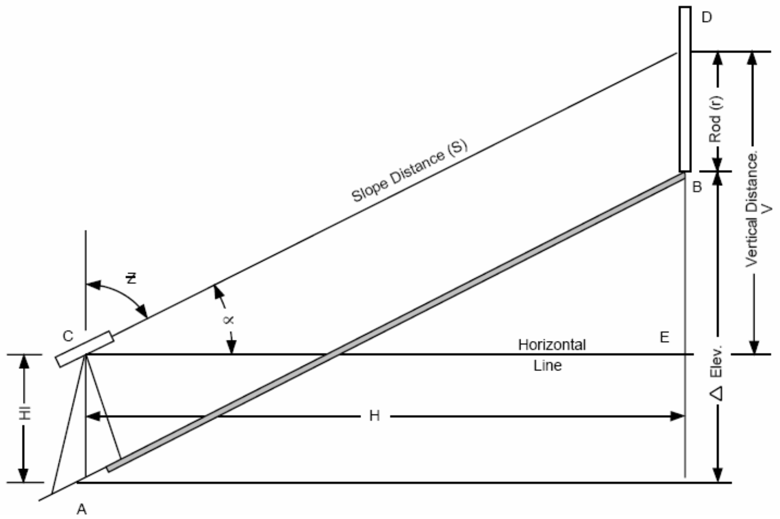


Figure: The trigonometric leveling approach

The different orders of vertical control are defined in terms of the allowable discrepancy between independent forward and backward leveling between benchmarks:

- First Order ± 4 mm /km
- Second Order ± 8 mm /km
- Third Order ± 24 mm /km
- Fourth Order ± 120 mm /km

3.4.3 Total Station Surveys

Total Positioning System (TPS) measurements are the values observed using any analog or electronic device from the theodolite family that measures zenith and horizontal angles and distance, and includes electronics to perform electronic calculations on a site. TPS is a form of electronic theodolite combined with an electronic distance measuring device (EDM).

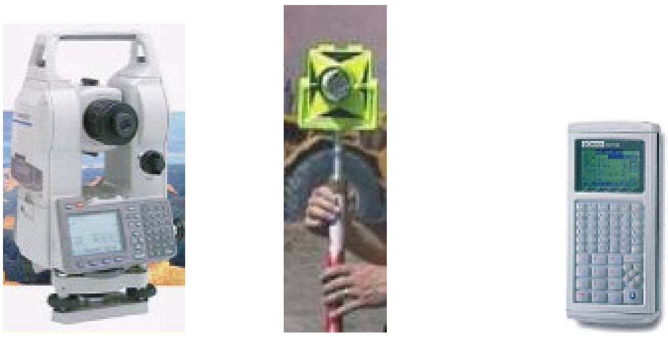


Figure: TPS, prism for EDM, and electronic notebook

The primary function of a TPS is to measure slope distance, vertical angle, and horizontal angle from a setup point to a foresight point. Most total stations use a modulated near-infrared light emitting diode, which sends a beam from the instrument to a prism. The prism reflects this beam back to the instrument. The portion of the wavelength that leaves the instrument and returns is assessed and calculated. Distance measurements can be related to this measurement.

Figure: Electronic display of TPS measurements

Total Station Surveying has some advantages compare to traditional theodolite surveying. These include:

- Relatively quick collection of information
- Multiple surveys can be performed at one set-up location
- Easy to perform distance and horizontal measurements with simultaneous calculation of project coordinates (Northings, Eastings, and Elevations)
- Digital design data from CAD and GIS programs can be uploaded into and from data collector
- Can perform surveying calculations with adjustments: COGO and height calculations

The horizontal accuracy of a total station is dependent on instrument type and in general:

- Angle accuracy (horizontal or vertical) can range from 2" to 5"
- Distance accuracy can range from +/- (0.8 + 1 ppm x D) mm to +/- (3 + 3 ppm x D) mm, where D = distance measured. A ppm (parts per million or mm/km) correction is entered into total station; it takes into consideration changes of the speed of light in air defence of the density of the air.

The vertical accuracy is highly dependent on leveling the instrument. Thus, two leveling bubbles are provided on the instrument and are referred to the circular level and the plate level:

- Sensitivity of Circular level = 10' / 2mm
- Sensitivity of Plate level = 30" / 2mm

Vertical elevation accuracy is not as accurate as using conventional survey levels and rod techniques.

3.4.4 Conclusion

A map is a mathematical transformation of reality that requires two steps of modeling and transformation:

- The Earth has to be model as an ellipsoid (sphere) or/and geoid
- An ellipsoid (sphere) has to be projected into the flat surface

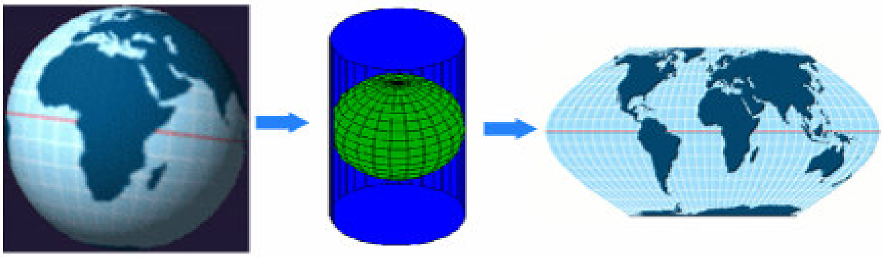


Figure: A map is a mathematical transformation of reality

In this module, we have discussed the first step of such modeling and transformation. In the following module we will discuss the mathematical projection process.

Module Self-study Questions:

- Why is it important to know the datum used for a given map? Which datum(s) are you most likely to encounter for maps used in Lithuania?
- How does a spheroid (ellipsoid) differ from a sphere in approximating the shape and size of the Earth?
- Explain the difference between geocentric and local horizontal datums.
- What is a geodetic datum?
  - The latitude and longitude of FMG origin
  - The data point that defines the location of Greenwich, England
  - A theoretical map project designed to provide accurate scale over the entire surface of an oblate spheroid
  - The set of parameters that define the size and shape of the Earth and the origin of coordinate systems that describe positions on the Earth
- When converted from DMS to DD units, $66^{0}30'00''$ will read:
  - 66.3^{0}
  - 66.5^{0}
  - 66.7^{0}
  - none of the above
- Which of the following spheroids is ground-measured, rather than satellite-determined?
  - Krassovskij 1940
  - WGS84
  - GRS80

Required Readings:

- **Geodesy**, Natural Recourses Canada, http://www.geod.nrcan.gc.ca/geodesy/index_e.php, **correspondent links** What is Geodesy? http://www.geod.nrcan.gc.ca/geodesy/whatis/index_e.php Geoid http://www.geod.nrcan.gc.ca/geodesy/geoid/index_e.php Control Networks http://www.geod.nrcan.gc.ca/geodesy/control/index_e.php Surveying http://www.geod.nrcan.gc.ca/geodesy/survey/index_e.php
- **Chapters 4, 5 and 7**, Information technology - Spatial Reference Model (SRM) http://standards.iso.org/ittf/PubliclyAvailableStandards/C030811e_FILES/MAIN_C030811e/ISOIEC_18026E_TOC.HTM
- The Geographer's Craft notes, Department of Geography, University of Colorado, **Geodetic Datum** http://www.colorado.edu/geography/gcraft/notes/datum/datum.html **Coordinate Systems** http://www.colorado.edu/geography/gcraft/notes/coordsys/coordsys.html
- **Chapters 1, 3 and 7**, Using ArcGIS Survey Analyst, Tim Hodson and Kristin Clark, ESRI Digital Library, 2003

ESRI Virtual Campus Course:

- **Module 1 and 6: Cartography, Sizing up the Earth,** and **Introduction to Datums**, Understanding Map Projections and Coordinate Systems

Assignment:

- **Assignment 3:** Working with surveying measurements in GIS

References

- [1] Elements of Cartography, Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, 6th ed., NY: John Wiley & Sons Inc., 1995.
- [2] Geodesy, Natural Recourses Canada, http://www.geod.nrcan.gc.ca/geodesy/index_e.php
- [3] The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/notes.html

Terms used

- Global geodesy
- Geodetic surveying
- Plane surveying
- Global Cartesian system
- Geographic coordinate system
- Projected coordinate system
- Prime meridian
- Ellipsoid
- Geoid
- Geocentric horizontal datum
- Local horizontal datum
- Coordinate systems transformations
- Vertical datum
- Triangulation
- Trilateration
- Traverse
- COGO
- Leveling
- Total station survey

4 Global Positioning Systems

A Global Positioning System (GPS) is a radio-transmitted and received surveying technique used to acquire spatial data. This module examines the components of a global positioning system system, how radio signals are transmitted, and what a GPS receiver does to translate those signals to calculate a receiver's position. Varying levels of GPS accuracy and possible improvements in accuracy are also examined in this module.

Module Outline

- GPS Basics
- GPS Segments
- How Does GPS Work?
  - Geometry of GPS Systems
  - GPS Frequencies and Codes
  - Types of GPS Positioning
- GPS Accuracy
  - Error sources
  - Differential correction

4.1 GPS Basics

GPS is a space-based radio-positioning system nominally consisting of a 24-satellite constellation. Civilian applications of GPS exist in a number of fields, including surveying, transportation analysis, natural resource management, and agricultural production. GPS technology is also important in spatial data acquisition and surveying controls within a SDI framework.

GPS is basically a type of radio-navigation system. Any device that uses radio waves can be used to determine distance and/or direction. The earliest radio-navigation systems could only determine the bearing (direction or angle) to a beacon (e.g. Radio Direction Finder (RDF), Non-Directional Beacon (NDB), VHR Omnidirectional Range (VOR). Later radio-navigational systems could determine bearing and range (distance) (e.g. Instrument Landing System (ILS), TACAN).

The U.S. Military developed the world's first satellite radio-navigation system, Transit, in the 1960's. The experience of developing and maintaining this system was used to develop NAVSTAR GPS. The 1990 Gulf War proved the effectiveness of NAVSTAR GPS for military operations. Despite its military origins, GPS quickly became popular for a wide variety of civilian mapping and navigational needs. Positions can be reported by a GPS in terms of Latitude (Φ), Longitude (λ) and Elevation (Z).

There are currently 4 GPS systems operational or being built:

- NAVSTAR GPS (U.S., operational)
- GLONASS (Russia, being rebuilt)
- Beidou (China, experimental/under construction)
- Galileo (Europe, experimental)

In this module we will primarily deal with the NAVSTAR global positioning system as it is fully operational and is used around the world. NAVSTAR stands for Navigation Signal Timing and Ranging. NAVSTAR GPS was developed from research conducted at Johns Hopkins University in the 1960's. The system is operated and maintained by the U.S. Department of Defence (DoD) that initiated this GPS program in 1973. The program cost about $12 billion USD to develop. It became operational in December, 1993 and contains a minimum of 24 satellites, plus a few spare satillites. Currently there are 30 operational satellites that operate 24 hours per day in all weather conditions.

4.2 GPS Segments

In general, GPS consists of 3 segments: space, control, and user.

Space segment contains GPS satellites itself. There are currently 24 operational GPS satellites in orbit, with 6 in-orbit spare satellites. The NAVSTAR GPS satellites weigh about 900 kg and are about 5 meters wide with the solar panels fully extended. They are built to last about 10 years, but many have outlasted their original estimated life-span. The sun-seeking solar panels provide primary power and charge nickel-cadmium (NiCad) batteries for secondary power support. On board each NAVSTAR satellite there are four highly accurate atomic clocks - 2 cesium and 2 rubidium. There are numerous antennas on the satellite.

GPS satellites orbit around the Earth. Satellites orbit the Earth at 20,200 km altitude. The high altitude insures that satellite orbits are stable, precise and predictable, and that the satellites' motion through space is not affected by atmospheric drag. Each satellite orbits the Earth every 11 hours 58 minutes, so the GPS satellites cross over any point on the Earth approximately twice per day.

There are four satellites in each of 6 orbital planes, and each plane is inclined 55 degrees relative to the equatorial plane (the satellite path crosses the equator at 317º, 17 º, 77º, 137º, 197º and 257º degree right ascension angle). It also insures satellite coverage over large areas.

Each GPS satellite transmits multiple signals on multiple frequencies. There are two main frequencies used by GPS Satellites (L1 (1575.42 MHz) and L2 (1227.60 MHz). L1 contains C/A code, P code, and Navigation Message. L2 contains P code only. The L3, L4, and L5 frequencies for transmitting navigational signals already exist or are proposed.

The satellites also provide two levels of service: Standard Positioning Service (SPS) and Precise Positioning Service (PPS). These satellite services, signals and codes are discussed in greater detail below.

GPS satellites transmit signals at extremely low power levels. Satellite signals require a direct line to GPS receivers (“line of sight”), thus signals cannot penetrate water, soil, walls or other obstacles. The signals can pass through clouds, glass and plastic but cannot go through most solid objects, such as buildings and mountains.

The NAVSTAR **control segment** is monitored by US DoD (Department of Defence). The control segment consists of five Monitor Stations (Hawaii, Kwajalein [West Pacific], Ascension Island [South Atlantic], Diego Garcia [Indian Ocean], Colorado Springs), three Ground Antennas (Ascension Island, Diego Garcia, Kwajalein) and a Master Control Station (MCS) located at Schriever Air Force Base in Colorado. The main control is at Shriever AFB in Colorado Springs, Colorado.

Figure: The NAVSTAR control segment

Ground stations monitor orbits of satellites in view and accumulating ranging data. Orbits are precisely measured. This information is processed to determine satellite orbits and to update each satellite's navigation message. Updated information, such as an *almanac* (predicted orbits) and orbital corrections (actual orbits), are transmitted to each satellite via the ground antennas. Satellites then transmit the *ephemeris* information (that includes satellite data messages (position and timing), an updated almanac, and orbital corrections to receivers.

The **user segment** covers both military and civilian users. The user segment consists of receivers that provide positioning, velocity and precise timing to users worldwide. The global GPS market is growing at 22% per year, it was estimated a $4 billion industry in 1998, a $6.2 billion industry in 2000, and is expected to exceed $50 billion by 2010.

How Does GPS Work?

How does GPS calculating a position?

GPS satellites are constantly transmitting signals that contain orbital data and timing information. Receivers pick up those signals and use the information to compute positions. GPS receivers calculate its position by measuring the distance to satellites. The distance measurement, calculated by a GPS receiver, is named as a satellite range or "pseudo-range." The satellites are the known points; the GPS receiver on the ground is the unknown point. Receivers don't send signals back to satellites.

4.2.1 Geometry of GPS Systems

As noted above, a GPS receiver calculates its position by measuring the distance to satellites (satellite ranging). The distance can be measured by two techniques: **code phase positioning** or **carrier phase positioning**.

In **code phase positioning**, the range is measured as *elapsed transit time* for signals to travel from satellite to receiver. Since radio waves travel at the speed of light. So, we can multiply the travel time of the GPS signal by 300,000 kilometres per second to get the distance between the GPS satellite and the receiver (this is described in details within "Code phase positioning" section below). It takes about .06 seconds for a GPS radio signal to reach Earth (). GPS receivers work by determining the distance to three satellites and use *trilateration* to compute a 3-D position (latitude, longitude and altitude). Distance measurements to 4 satellites provide improved (adjusted) computations of a 3-D position.

In **carrier phase positioning**, distances can be measured from the receiver to the satellites by counting the number of wave cycles between satellites and receivers. It is the same general idea as for code phase differential positioning except that the carrier wavelengths are shorter (19 and 24cm) and the range estimates are more accurate.

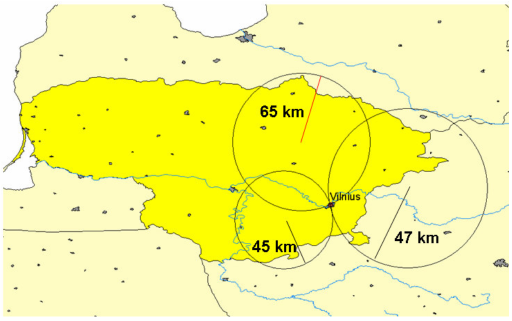


Figure Trilateration principles

Most of GPS measurements have been done by code phase positioning. GPS was not originally designed for carrier phase positioning.

With GPS, trilateration refers to measuring the distance (lengths) from 3 satellites to establish a position on Earth. Receivers compute positions through *trilateration* (measuring *distances* to satellites, not *angles*, as in triangulation). Three ranges will locate a point in two-dimensional space. Considering the three range measurements together, a position must be where the three circles intersect.

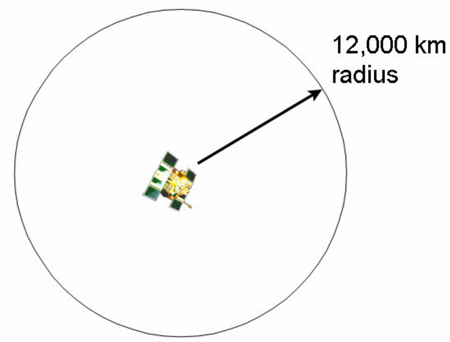


Figure: One measurement narrows down a position to the surface of a sphere. 12,000 km is the radius of a sphere centered on the satellite. A position could be anywhere on the surface of that sphere.

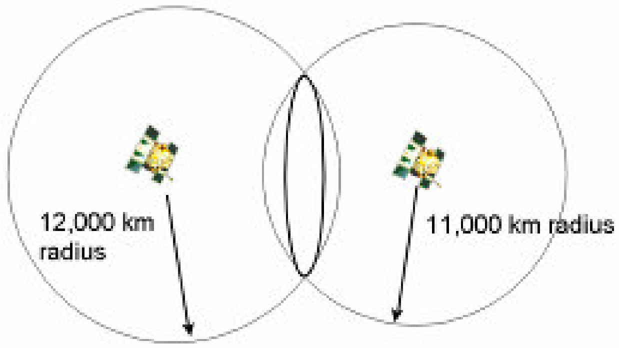


Figure: A second measurement narrows down a position to the intersection of two spheres. The intersection of two spheres is a circle. A position is somewhere on that circle.

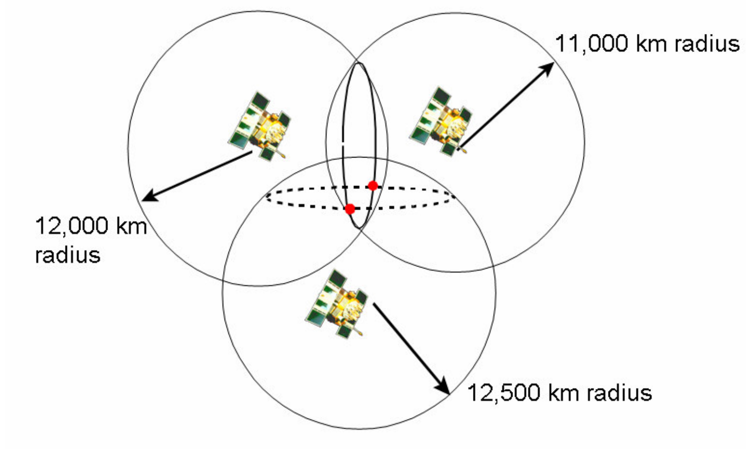


Figure: A third measurement narrows down our position to just two points.

The three spheres intersect at only two points. One of the two points can be discarded because one point might be nowhere near the Earth. The computers in GPS receivers have various techniques for distinguishing the correct point from the incorrect one. But there is a reason we need a fourth measurement.

A fourth measurement is needed to correct for **timing offset** (difference in synchronization between satellite and receiver clocks). Thus satellites use highly accurate atomic clocks, but receivers use only accurate quartz clocks. Timing offset refers to the difference in synchronization between the satellite clock and the receiver clock. Atomic clocks are far too expensive to put in GPS receivers, so a correction must be applied to compensate for the difference between the satellite and receiver clocks.

Thus, establishing the travel time of the GPS signal for code phase positioning is the first step in calculating the distance between the satellite and receiver. The second step is to determine the receiver's position using distance measurements from four satellites. The timing offset correction works by the following principles:

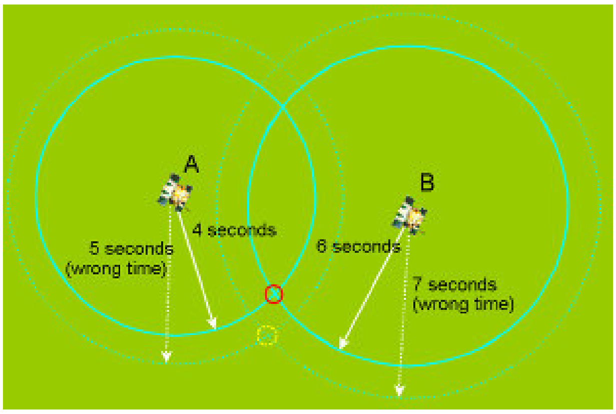


Figure: If there is no timing error, a position is where the 2 circles intersect. In an ideal situation there would be, for example, 1 second from satellite A and 1 second from satellite B timing offsets and a position will be where the two distorted ranging circles intersect.

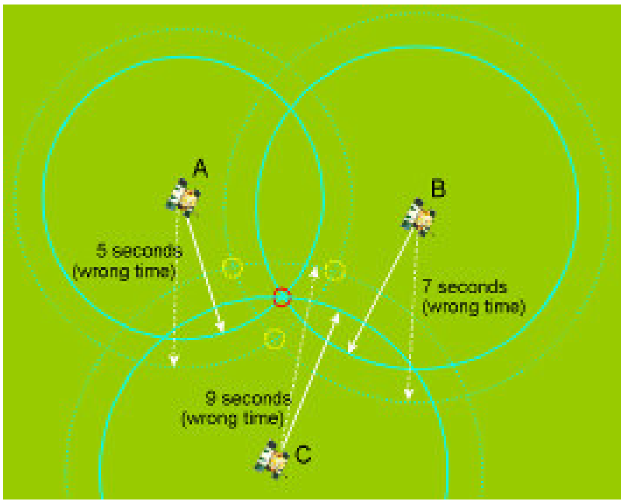


Figure: If we would have ideal clocks, and a third measurement is added, all three ranging circles intersect at the correct point. But with inaccurate clocks, it will be uncertain as to where the zone among the circles intersects to define a specific position.

To find a point position in a timing offset situation, one extra measurement has to be added to correct any consistent clocks' errors. The four measurements used to solve for four variables are:

- Latitude
- Longitude
- Altitude
- Time

4.2.2 Frequencies and Codes

Each NAVSTAR GPS satellite transmits the following multiple signals on multiple frequencies:

- L1 (1575.42 MHz)
- L2 (1227.60 MHz)

- L3 (1381.60 MHz)
    - Used to transmit a signal when a missile is launched, nuclear detonations occur, or when high-energy infrared events are detected
- L4 (1841.40 MHz) (Proposed)
    - Being studied for additional ionospheric correction
- L5 (1176.45 MHz) (Proposed)
    - For Safety-of-life signal
    - Frequency within protected band for aviation, so there is little interference

L1 and L2 are the primary navigational signals. These signals are working as carriers for codes' transmission. The L1's carrier is 1575.42 MHz and carries both the status message and a pseudo-random code for timing. The L1 contains the following 2 positioning codes and *navigation message*:

- C/A (Coarse Acquisition) Code
- P (Precise) Code
- Navigation message

The L2 carrier is a 1227.60 MHz and is used for the more precise military pseudo-random code. The L2 (1227.60 MHz) contains 1 positioning code and respective navigation message:

- P (Precise) Code
- Navigation message used to measure ionospheric delay in military receivers.

The radio signals travel at the speed of light: 300,000 km per second. It takes about 6/100ths of a second for a GPS satellite signal to reach Earth. These signals are transmitted at a very low wattage (about 300-350 watts in the microwave spectrum).

There are two types of pseudo-random codes. The first pseudo-random code is called the C/A (Coarse Acquisition) code. It modulates the L1 carrier. It repeats every 1023 bits and modulates at a 1MHz rate. Each satellite has a unique pseudo-random code. The C/A code is the basis for civilian GPS use. The C/A code (Coarse Acquisition code) is available to civilians as the **Standard Positioning Service** (SPS). Before SA was turned off, SPS provided a predictable positioning accuracy of 100 meters horizontally, 156 meters vertically and time transfer accuracy within 340 nanoseconds (95 percent). SPS now provides average horizontal accuracy of <= 13 meters 95% of the time and average vertical accuracy of <= 22 meters 95% of the time.

Summary of the C/A Code (Coarse Acquisition):

- Modulates the L1 signal
- Is available to civilian users
- Provides Standard Positioning Service (SPS)
- Repeats 1 MHz Pseudo Random Noise (PRN) code
- Generates a "seed" number by algorithm - this allows GPS receiver to generate identical C/A code to satellite

- Repeats every 1023 bits (1 ms)
- Has a different PRN code for each satellite

The second pseudo-random code is called the P (Precise) code. It repeats on a seven-day cycle and modulates both the L1 and L2 carriers at a 10MHz rate. This code is intended for military users (and other authorized users), provides higher accuracy, and can be encrypted. When it's encrypted it's called a "Y" code. Since P code is more complicated than C/A it's more difficult for receivers to acquire. That's why many military receivers start by acquiring the C/A code first and then move on to P code.

Summary of the P (Precise) Code:

- Available only to the military and authorized users
- Provides Precise Positioning Service (PPS)
- Very long (7 day) 10 MHz PRN code
- Anti Spoofing Mode
    - P-Code encoded into Y-Code, which requires a GPS with a cryptographic key to decode
    - Virtually anybody can generate correct C/A code, so it is quite easy to throw off the GPS Position
    - Ability to decode a signal allows user to be sure that it is a "real" GPS signal

Each satellite transmits a *navigation message* as a low frequency signal added to the L1 codes. It contains satellite orbital elements, clock behaviour, and system time and status messages. In addition, an almanac is also provided which gives the approximate data for each active satellite.

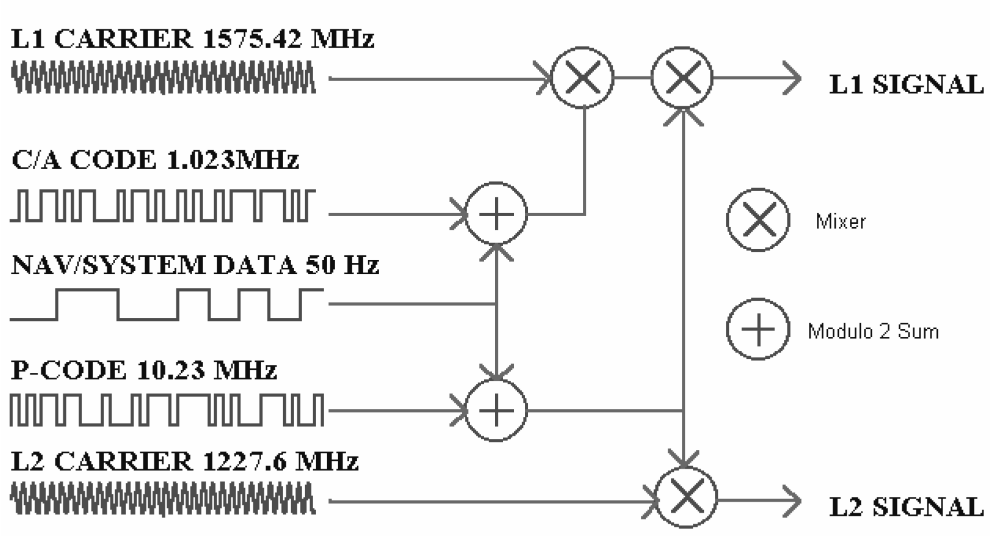


Figure: Mixing of GPS Signals

A navigation message within L1 frequency signal contains three different bits of information: a **pseudorandom code, ephemeris data** and **almanac data.** The **pseudorandom** code is simply an ID code that identifies which satellite is transmitting the information.

Each satellite broadcasts an individual ephemeris that is updated continuously. **Ephemeris data** tells the GPS receiver where each GPS satellite should be at any time throughout the day. The satellite locations have been computed from orbit measurements, along with corrections that are transmitted to the satellites by the DoD. The ephemeris is used by receivers, along with almanac data, to establish precisely the position of each satellite being tracked.

The ephemeris message contains orbit data for an individual satellite, whereas the almanac contains orbit data for all the GPS satellites.

Almanac data, which is constantly transmitted by each satellite, contains important information about the status of the satellite (healthy or unhealthy), a prediction of the orbits of all satellites, and current date and time. Almanac can be obtained from any GPS satellite. A GPS receiver automatically collects an almanac each time it is in operation for about 12.5 minutes. This part of the signal is essential for quick acquisition of satellite positions by the receiver.

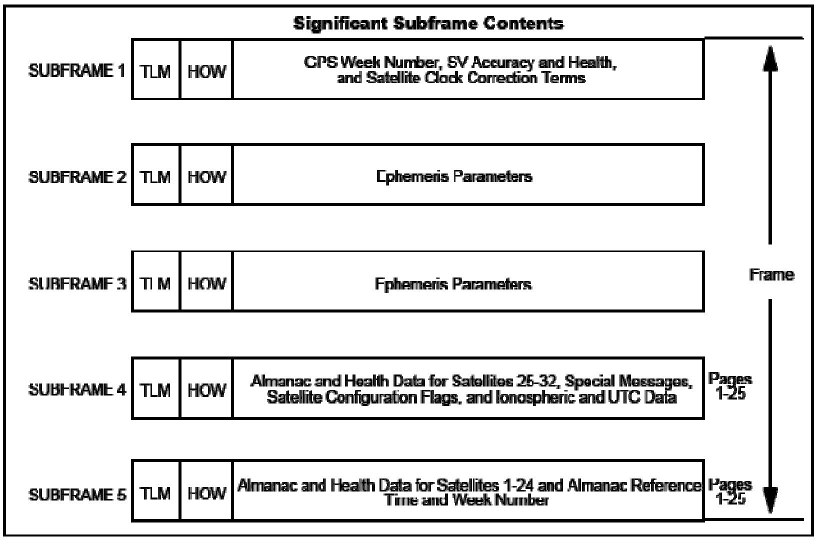


Figure: Content of the GPS navigation message

Summary of navigation messages:

- 50 Hz (i.e. 50 bits per second) signal contains
  - Clock corrections
    - Used to synchronize atomic clocks with GPS time
  - Ephemeris
    - Ephemeris data tells the GPS receiver where each GPS satellite should be at any time throughout the day
    - Each satellite transmits its own ephemeris data, but not that for other satellites
    - Valid for about 30 minutes
    - Updated every 30 seconds
    - Orbital position errors may be present in the ephemeris data, causing errors in the positions a GPS receiver calculates
  - Ionospheric refraction model
    - Characterizes how ionospheric conditions might be expected to refract satellite signals

- The Satellite Almanac is a subset of the clock and ephemeris data, with reduced precision
  - Sent along with position and timing messages
  - Prediction of all satellite orbits
  - Needed to run satellite availability software
  - Valid for about 30 days
  - Constantly transmitted by each satellite
  - Status of the satellite (healthy or unhealthy)
    - GPS receivers ignore signals from unhealthy satellites
    - Commonly set to unhealthy when a satellite needs its orbit to be corrected
  - Current date and time
  - Coarse orbital parameters for each satellite
  - Needed to run GPS Mission Planning software

- Entire message contains 25 frames of 1500 bits, and takes 12.5 minutes to transmit
- Clock correction and ephemeris data repeated in each frame

The navigation message data includes information required to determine the following:

- Satellite time of transmission
- Satellite position
- Satellite health
- Satellite clock correction
- Propagation delay effects
- Time transfer
- Constellation status

Although the satellites are in high, stable orbits, they can still be pushed around by the solar wind. Every so often, the orbit will have to be corrected by firing a rocket thruster.

4.2.3 Types of GPS Positioning

As mentioned above, there are positioning techniques to calculate distances to the satellite from a receiver. They are:

- Code Phase Positioning
  - For military and civilian units
  - Receivers use GPS signals “as designed”
  - Stand-alone or differential
- Carrier Phase Positioning
  - Survey-grade civilian units (very expensive)
  - Takes advantage of variations in wavelength to obtain higher accuracies
  - Inherently differential
  - GPS was not initially designed for this

Code phase positioning techniques use pseudo random number (PRN) code that is generated at exactly the same time by the GPS Satellite and the GPS Receiver, and is broadcast by the satellite. Because the satellite is far away, it takes time for the PRN code to reach the receiver.

So again, to compute a 3D position, *distances* are measured to 4 satellites. To calculate the distances in code phase positioning techniques, *travel time* of satellite signals has to be measured. Question: *How do we find the exact time the signal left the satellite?*

In order to measure the travel time of the satellite signal, it has to be known when the signal left the satellite and when the signal reached the receiver. A receiver can "know" when it receives a signal, but how does it know when the signal left the satellite?

GPS satellites generate a set of digital PRN codes. These codes can be compared easily and unambiguously (see Figure). The codes are "pseudo-random" sequences that actually repeat every millisecond. The GPS satellites and receivers are synchronized so they're generating the same code at exactly the same time.

When a GPS receiver receives codes from a satellite, it looks back to see how long ago the receiver itself generated the same code. The receiver matches the PRN code from the satellite to that from the receiver to determine the time delay. By doing this, a receiver calculates the difference in time the signal took to get from the satellite to the receiver. In other words, the receiver compares how "late" the received satellite code is, compared to the code generated by the receiver itself.

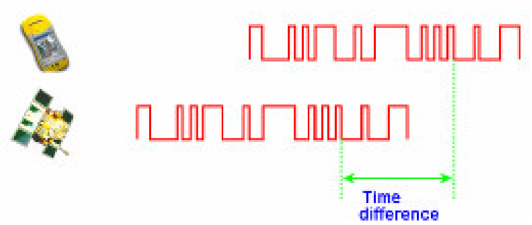


Figure: the PRN codes have been matched by offsetting the PRN code from the satellite.

Then, time delay is converted into distance to satellite with D = c * T, where

- c = speed of light (300,000 km/s)

Thus, for a 0.06 second delay, D = 300000 * .06, or 18,000 km.

Next time offset between the precise satellite atomic clock and the quartz clock of receiver has to be corrected by using extra fourth measurement. This process is also discussed below in the GPS receiver algorithm.

There are 4 atomic clocks in each GPS Satellite 2 cesium and 2 rubidium clocks. Atomic clocks are expensive and large. A cesium clock has 133 vibrates at 9,192,631,770 cycles per second and can maintain time to within 1 billionth of a second. Cesium atomic clocks are more expensive but more accurate than rubidium atomic clocks.

Civilian GPS receivers use an inexpensive quartz clock (probably higher quality than a watch). It can keep time to within 1/1000 of a second per day. Quartz crystal vibrates at a steady rate when electricity is applied, and this is converted into seconds. Quartz satellite clock is constantly adjusted to the atomic clock time, once the navigation signal received from a GPS satellite.

The GPS receiver algorithm for code phase positioning techniques:

- Read internal quartz clock and almanac to determine which satellites should be visible
- Listen for expected signals, according to almanac
  - If expected signals are not found, listen for all signals
- “Lock On” to the four signals that produce the lowest DOP (dilution of precision)
  - Best geometric configuration
  - Define strongest signal
- Download the ephemeris for each “locked on” satellite
- Calculate the pseudo-range to each “locked on” signal by “sliding” the receiver-generated pseudo-random code against the GPS satellite-computed pseudo-random code
  - The amount of “sliding” necessary to make the two pseudo-random codes match indicates the time difference between when the signal was emitted and when it was received
  - Since light travels at 300,000 km/s we can calculate the distance to the satellite using D = c * T
- Calculate the difference between where the spheres are, and where they need to be to meet at a point by using one extra measurement. As it was mentioned above the four measurements are required to solve for four variables: latitude, longitude, altitude and time offset. Convert this to a time, and apply this correction to the GPS receiver clock
- Calculate the exact distance to each satellite
- Using the known position of the satellites from the ephemeris, calculate the position of the GPS receiver by using trilateration principle
- Go to Step 5 to calculate the next position

Carrier phase positioning determines the number of wave cycles between satellites and receivers and is used to calculate distance. The exact wave and fraction of a wave from a satellite allows the exact distance to be calculated from the satellite to a mobile receiver and from the satellite to a GPS base station to within mm.

Usually carrier phase positioning is *differential* if two or more receivers are used to track satellites at the same time (base and rover receivers should be within around 30 km of each other). The corrected difference (e.g. for the ionospheric delay) is derived from the difference in carrier cycle measurements by two or more receivers. Carrier phase positioning techniques are used with the same general idea as for code based differential positioning, but because the carrier wavelengths are short (L1 19 cm and/or L2 24 cm), the range estimates will necessarily be more accurate. The L1 and/or L2 carrier signals are used in carrier phase surveying.

These techniques are primarily for survey use. If tracked and measured these carrier signals can provide ranging measurements with relative accuracy in either centimetres or millimetres, under special circumstances.

Carrier phase differential works on the principle of the Doppler phenomenon to define the number of wave cycles or fractions of waves. Since the satellite positions, velocities, precise carrier frequencies, and time are known from the satellite ephemeris, then the ground position can be calculated. Carrier-phase GPS uses code-phase techniques to get accurate distances. If the code measurement is accurate to a meter, then only a few wavelengths of carrier have to be considered to determine which cycle really marks the edge of the timing pulse. Doppler approach is used to calculate wave cycles and fractions of wave to or from the edge of the timing pulse.

As in standard differential GPS, clock and ephemeris errors can be reduced for the carrier phase method by using two receivers. These techniques are also required a substantially higher occupation time of signal acquisition.

One technique of surveying uses **dual-frequency carrier phase positioning.** Dual-frequency receivers receive signals from satellites on two frequencies simultaneously. Receiving GPS signals on two frequencies simultaneously allows the receiver to determine very precise positions. The measurement number is determined by both signals in L1 and L2 bands. These are tracked by the receiver or two signals from different systems (e.g. NAVSTAR and GLONAS systems). Measurement quality is accounted for by a number of measurement errors, such as noise, multipath, tropospheric and ionospheric errors. An ambiguity resolution is necessary to employ the largest possible number of measurements using network surveying techniques and, thereby, minimizing measurement errors.

There are currently 4 GPS systems that are either operational or being built. Here are the overviews of two other GPSs:

- GLONASS (Global Navigation Satellite System)
  - Russian (formerly Soviet) GPS System
  - Minimum 24 satellites
    - Had full constellation in 1995
    - Unable to build new satellites for a few years after breakup of the Soviet Union
    - Only 8 satellites in operation on April 2002
    - Currently 14 operational satellites
    - To be fully operational by 2008 with 18 satellites
    - Complete constellation to be in operation by 2010
  - Civil Signal L1-C/A (1602 + N*0.5625 MHz)
  - 57-70m horizontal accuracy, 70m vertical accuracy
  - Three orbital planes, 120 degrees apart, 64.8 degree inclination, 19,100 km altitude
- Galileo
  - European GPS under development
  - Total cost is more then €3 billion
  - First test satellite (GIOVE-A) launched December 28, 2005
    - Second test satellite (GIOVE-B) to be launched in 2007
  - Availability timeline: 2010
  - Primarily for civilian use
  - Minimum 27 (+3 spares) satellites
  - Civil signals:
    - L1 (1575.42 MHz)
    - E5 (A/B: 1176.45/1207.14 MHz)

    - E6 (1278.75 MHz)
- Rubidium + hydrogen maser atomic clocks (more accurate than GPS)
    - More accurate positioning than GPS possible
- 23222 km orbit
    - 3 orbital planes, 56 degree inclination
    - 9 operational satellites and 1 spare per plane

4.3 GPS Accuracy

Accuracy is a term that broadly describes the level of uncertainty, or error, associated with experimental measurements. Measurements are more accurate when there are few errors; less accurate when there are more errors.

Many factors affect the accuracy of GPS measurements. Measured positions can be compared with true geographic coordinates to assess their level of error. The accuracy of GPS is determined by the sum of several sources of error.

- The accuracy of GPS is determined by the sum of several sources of error:
  - Atmospheric effects
  - Multipath
  - Satellite geometry
  - Measurement noise
  - Ephemeris data
  - Satellite clock drift
  - Selective availability (SA)

4.3.1 Error sources

The travel time of GPS satellite signals can be altered by the **ionosphere** and the **troposphere**. The ionosphere and troposphere both refract GPS signals, causing the speed of the GPS signal to be different from the speed of a GPS signal in space.

The troposphere is the lower part of the Earth's atmosphere where temperatures decrease with an increase in altitude. It can be < 9 km thick over the poles and >16 km thick over the equator. The presence of neutral atoms and molecules in the troposphere affects electromagnetic signal propagation.

In the ionosphere, which ranges from about 50 km above the Earth's surface to about 1,000 km or more, ionizing radiation (principally from solar ultraviolet and x-ray emissions) causes electrons to exist in sufficient quantities to affect radio-wave propagation.

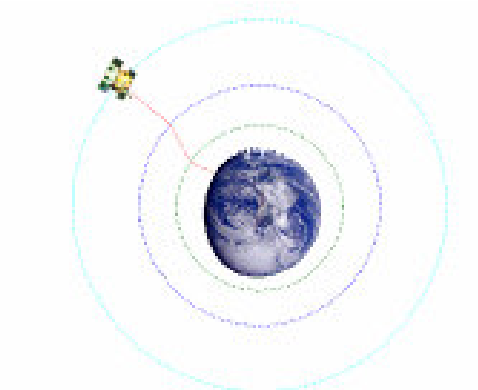


Figure: Ionospheric and tropospheric refraction

Multipath error occurs when GPS satellite signals bounce off other objects before reaching the receiver antenna. The difference in path lengths causes the signals to interfere with each other at the antenna and to contribute an error to the pseudo-range observable. Multipath is usually noted when operating near large reflecting obstacles, such as buildings and fences. Signals can also reflect off the ground and roofs.

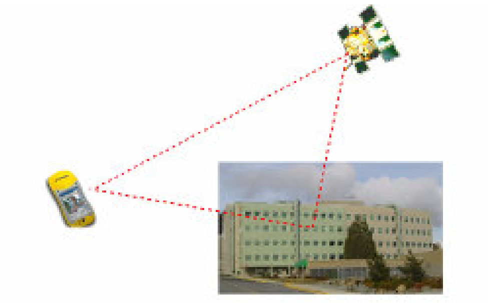


Figure: Multipath phenomenon

Satellite geometry affects the quality of GPS positions computed by the receiver. Satellite geometry can be described by dilution of precision parameters. Thus **geometric dilution of precision** (**GDOP**) is a measure of the quality of the satellite configuration and refers to where the satellites are in relation to one another. Wider angles among satellites are better for measurement purposes. GDOP can magnify or lessen other GPS errors.

GDOP comprises the following DOPs:

- HDOP - Horizontal Dilution of Precision and refers to horizontal measurements (latitude, longitude)
- VDOP - Vertical Dilution of Precision and refers to altitude
- TDOP - Time Dilution of Precision and refers to clock offset

$GDOP^2 = PDOP^2 + TDOP^2$

Where PDOP is **position dilution of precision**. PDOP is a root mean square (RMS) measure of the effects that any given three-dimensional position solution geometry has on position errors.

$PDOP^2 = HDOP^2 + VDOP^2$

PDOP is the most commonly used measure of satellite geometry. It is a unitless measure that refers to the quality of horizontal (HDOP) and vertical (VDOP) measurements (latitude, longitude and altitude). PDOP predicts the accuracy of positions relative to satellite geometry. A low PDOP indicates a higher probability of position accuracy. A high PDOP indicates a lower probability of accuracy.

Standard PDOP values are:

<=4 excellent

5-8 acceptable

>=9 poor

Most of GPS receivers permit to set up a parameter known as the PDOP mask to ignore constellations that have a PDOP higher than the limit of a required specification. Also, most receivers select the satellite constellation that gives the least uncertainty.

Therefore, GDOP refers to three position coordinates plus clock offset in the solution. It is a measure of the quality of a geometric constellation for position and time solutions. A low GDOP value represents a good satellite configuration, whereas a higher value represents a poor satellite configuration. The ideal orientation of four or more satellites is to have one satellite directly overhead and the other three evenly spaced above the receiver.

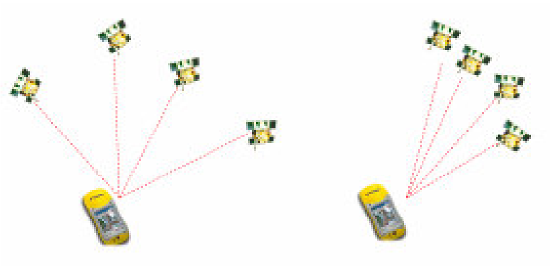


Figure: A good and poor satellite configurations

The DOP changes with time as the satellites move along their orbits.

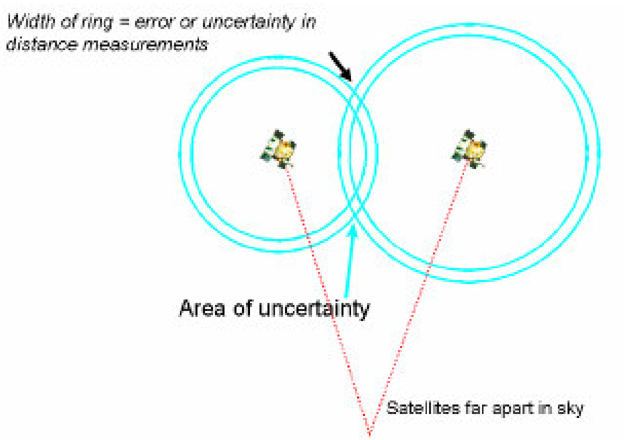


Figure: The width of the ring in this figure represents the error resulting from all possible sources. The diamond-shaped box where the rings intersect represents the area of uncertainty from the two satellite measurements. The diamond-shaped box formed by the two satellites and the receiver on the ground is fairly narrow when the satellites are far apart in the sky

Not incidentally, distance measurements from satellites to receivers are called "pseudo-ranges" because, due to these errors, the measurements are not true ranges.

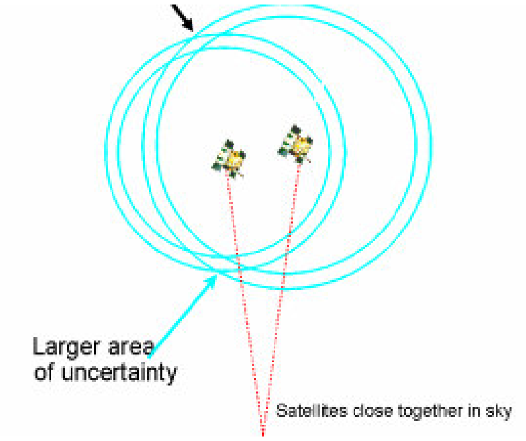


Figure: When the satellites are close together in the sky, the angle is narrow. And, when the rings are close together, the box representing the area of uncertainty becomes long and pointed, and there is greater uncertainty in the position calculation.

The vertical component of a GPS measurement is typically two to five times less accurate than the horizontal component. And it is strictly a matter of geometry. It's a function of where the satellites are with respect to receiver position. The vertical component is difficult to calculate because satellites have a limited perspective when measuring height. If the receiver could use signals from underneath it, the vertical component could be accurately measured, but the Earth blocks these signals.

The arcs representing satellite measurements are more vertical than horizontal, so there is more error in the vertical component. The geometry can be strengthening by putting a satellite directly overhead, but this is usually not the practical case. VDOP becomes larger with fewer satellites overhead.

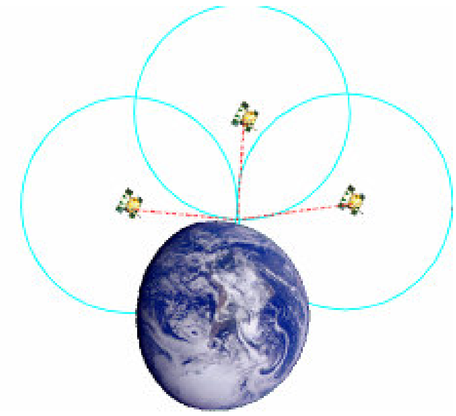


Figure: The arcs representing satellite measurements are more vertical than horizontal

Orbital position errors may be present in the **ephemeris data**, causing errors in the positions a GPS receiver calculates. The ephemeris contains satellite locations that have been computed from orbit measurements, along with corrections that are transmitted to the satellites by the DoD. It is a function of time and tells where the satellite will be and when.

Selective availability (SA) was intentional scrambling of the GPS satellite signals by the U.S. government. SA was turned off at midnight on May 1, 2000. SA was the largest component of GPS error. SA, implemented for national defence reasons, introduced artificial clock and ephemeris errors. This caused position errors up to 70 or even 100 meters. SA was activated on July 4, 1991. On May 1, 2000, President Clinton announced that SA would be discontinued. According to the President, "the decision to discontinue Selective Availability is the latest measure in an ongoing effort to make GPS more responsive to civil and commercial users worldwide. This increase in accuracy will allow new GPS applications to emerge and continue to enhance the lives of people around the world."

New technologies developed by the military enable the U.S. to degrade the GPS signal on a regional basis for national security purposes, making the worldwide degradation unnecessary. According to the White House Office of the Press Secretary, "GPS users worldwide would not be affected by regional, security-motivated GPS degradations, and businesses reliant on GPS could continue to operate at peak efficiency."

Also there are other sources that introduce error in GPS measurements. Thus measurement noise is occurring at the receiver antenna and distorts the signal by electrical interference. Measurement noise is also called receiver error or **receiver noise**.

Small variations in the satellite atomic clocks can translate to large position errors: a **clock error** of 1 nanosecond translates to 0.3 meters user error on the ground.

The final budget of source contributions to possible errors of GPS measurements can be looked like this:

- Ionosphere.................................. 5.0 meters
- Troposphere................................ 0.5 meters
- Ephemeris data........................... 2.5 meters
- Satellite clock drift........................ 1.5 meters
- Multipath.................................... 0.6 meters
- Measurement noise..................... 0.3 meters
- Selective availability..................... 30-100 meters
- Total..~ 10 meters

The satellite geometry can magnify or reduce the effects of other GPS errors.

4.3.2 Differential correction

There are techniques on how to correct some of these errors and achieve better accuracy with GPS receivers. One of such technique is **differential correction**. The main idea of differential correction is to use GPS receivers on the ground in a known location. This receiver is called a **base station** and acts as a static reference point. The coordinate position of based station is precisely measured. Differences between known coordinate values and one acquired by GPS in a particular time frame can be used for error correction in the local area. This error correction can be transmitted.

There are two ways to transmit error corrections.

- The base and roving (user) receivers collect GPS data at the same time from the same satellites. With **real-time** differential correction the base receiver computes timing errors and transmits error correction messages to other GPS receivers in the local area. The transmitted information is instantaneously applied in real-time. Real-time DGPS is necessary when navigation in high accuracy is required.
- With **post-processed** differential correction, the base receiver computes timing errors and stores the error data to a file. The rover file is later processed using differential correction software which uses error data from the base file to correct the rover data. The base and rover receivers have to "see" the same set of satellites at the same time, so the base file has to start before the rover file starts and end after the rover file ends. In general it is assumed that post-processed differential correction is more precise than a real-time one.

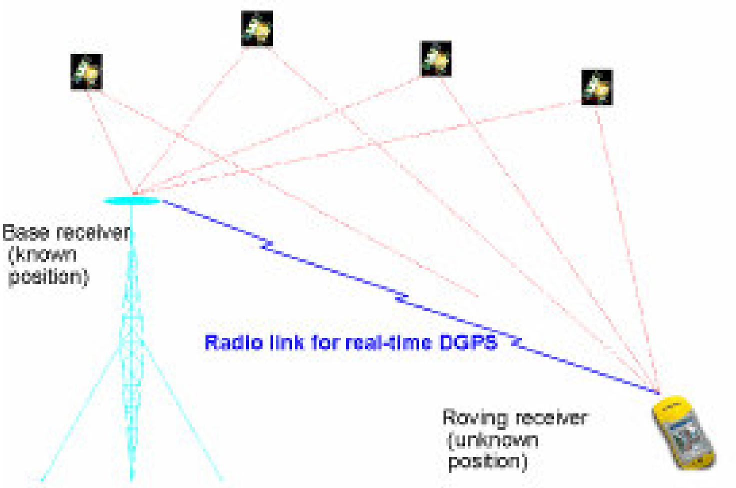


Figure: Real time differential GPS

Differential correction reduces the effects of some GPS errors. But it cannot correct, for example, multipath or receiver error because it counteracts only errors that are common to both reference (user) and roving (base station) receivers.

There are different sources that can be used for real-time and/or post-processed differential corrections. These include:

- Local base station can collect base correction files that can be transmuted for real-time or downloaded post-processed corrections
- Wide Area Augmentation System (WAAS) satellites (this service is specially designed for DGPS) transmitted signal for real-time correction. The WAAS is operated by the Federal Aviation Administration (FAA USA) and is used for aircraft navigation and landing. The WAAS signals are available to other GPS users who have the equipment capable of receiving WAAS corrections. Special receiving device is used to receive the correction signal via a radio link. This device can be part of a GPS unit.

- OmniSTAR communication satellites for DGPS services transmit signals for real-time correction. OmniSTAR provides a highly reliable positioning service with worldwide coverage, 70 reference stations around the globe, and 3 network control centers
- Nationwide Differential GPS (NDGPS) radio-beacons transmitted for real-time or can be used for downloads for post-processed corrections. Beacon on the Belt (BoB) receiver is used for receiving the correction signal via a radio link.

Figure: OmniSTAR base stations used for real-time differential correction (http://www.oxts.com/)

As long as a base station and/or communication satellite is within about 500 kilometres of the roving receiver, its files can be used to correct the roving receiver data. However, the closer you are to the base station, the better the correction.

The following numbers provide some idea of how accurate a GPS is:

- Recreational and mapping grade..................10-15 m
  - C/A code
  - Autonomous. With autonomous data collection, no differential method is applied. The largest source of error with uncorrected positions is atmospheric delay.
- Recreational and mapping grade..................1-5 m
  - C/A code
  - With differential correction. Real-time correction is less accurate than post-processing due to different ephemeris at the base and rover (because of distance between the two), frequency of the output message (latency) – depends on transmission rate of the signal (baud rate) and correction output rate (5 sec to 30 sec), and datum issues – errors occur when reference stations broadcast corrections using a datum other than WGS-84 (datum error is usually small but can be as great as 5 to 10 m in some areas).
- Sub-meter mapping grade.............................10 cm to 1 m
  - C/A code and carrier. Carrier phase receivers measure the distance from the receiver to the satellites by counting the number of waves that carry the C/A code signal; some receivers use what is called "carrier-smoothed code" to increase the accuracy of the C/A code. The carrier wave is a much finer measuring tool than the superimposed code (19 cm vs. 293.1 m), so it yields more accurate satellite ranges
  - With differential correction (e.g. Trimble) GeoXTs are capable of “post-processed carrier phase differential.” In this mode, the receiver measures the code, but also

measures the number of carrier waves between the satellite and the receiver. This method requires more rigorous data collection techniques: an accurate antenna height, a clear view of the sky, 5 satellites in view, rover within 50km of the base station, a base station collecting synchronized measurements over a surveyed-in coordinate, and lock on satellites must be maintained to achieve a minimum “block” of data of at least 10 minutes. With post-processed carrier phase differential, greater accuracies are obtained with longer occupation times. Accuracies are sub-meter in the horizontal after post-processing

- Survey grade...1 cm
  - P code or dual frequency
  - Advanced survey methods

GPS accuracy depends on many factors including type of equipment, time of observation, and position of satellites. It is important to research the type of equipment and receiver settings needed to achieve a specified level of accuracy. Theoretically, a longer occupation time, yielding more positions to average, should result in better accuracy.

4.3.3 Summary

- There are 3 GPS segments
- Formula for satellite ranging is $D = c * T$
- 4 satellites are needed to compute an accurate 3-D position
- The 4th measurement is used to correct the timing offset
- 6 sources of error (and additional factors) affect the accuracy of GPS positions
- Atmospheric error is the largest source of GPS error
- Vertical accuracy is always worse than horizontal accuracy
- Almanac and ephemeris data are different
- Post-processed differential is more accurate than real-time

Module Self-study Questions:

- How does GPS work?
- What s Selective Availability? Why is it used?
- Why does the reported altitude vary more than horizontal positioning?
- What accuracy can one expect from a recreational GPS unit?
- What is DGPS?
- How do some users get centimetre accuracy?

Required Readings:

- **Global Positioning System,** Canadian Spatial Reference System, http://www.geod.nrcan.gc.ca/geodesy/gps/index_e.php
- **GPS Guide,** Canadian Spatial Reference System, http://www.geod.nrcan.gc.ca/software_form_e.php?software=GPS_Guide&filename=GPS_Guide_e.pdf
- **Global Positioning System**, The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/gps/gps.html
- **GPS-System,** all sections, GPS – explained, http://www.kowoma.de/en/gps/
- **All About GPS**, all sections, GPS Tutorial, Trimble, http://www.trimble.com/gps

ESRI Virtual Campus Course:

- **Module 2 and 3: Flattening the Earth,** and **Understanding Aspect and Perspective**, Understanding Map Projections and Coordinate Systems

Assignment:

- **Assignment 3 Cont:** Working with surveying measurements in GIS

References

- [1] Canadian Spatial Reference System, http://www.geod.nrcan.gc.ca/geodesy/gps/index_e.php
- [2] GPS Guide, Canadian Spatial Reference System, http://www.geod.nrcan.gc.ca/software_form_e.php?software=GPS_Guide&filename=GPS_Guide_e.pdf
- [3] GPS – explained, http://www.kowoma.de/en/gps/

Terms used

- Radionavigation systems
- GPS segments
- Standard positioning service
- Precise positioning service
- DoD
- Code phase positioning
- Carrier phase positioning
- Trilateration
- Atomic clocks
- Quartz clocks
- L1 and L2 signals
- C/A Code
- Navigation message
- Ephemeris
- Satellite almanac
- P (Precise) code
- Code phase positioning
- Carrier phase positioning
- Dilution of precision
- Differential correction
- Real-time correction
- Post-processed

5 Map Projections

A map projection is a way of representing the Earth (an ellipsoid or sphere) on a map (flat surface). Map or cartographic projections involve the use of mathematical equations for presenting Earth's features on a flat surface. This module examines the coordinate system used in projecting Earth information on a map; map scales; classification of map projections by examining the shape of normal graticules, distortions, and aspects of developable surfaces. The key points of map distortion theory are also introduced. The most commonly used map projections and standard map coordinate systems are illustrated. Coordinate systems of the Republic of Lithuania are presented.

<u>Module Outline</u>

- Projected Coordinate System
- Map Projections
- Map Scale
- Map Projection Classification
  - By shape of normal graticule
  - By distortions
  - By aspect of developable surface
- Selected Map Projections
- Standard Map Coordinate Systems
- Projections and GIS Software

5.1 Projected Coordinate System

A map is a mathematical transformation of reality that requires two steps of modeling and transformation:

- The Earth has to be modeled as an ellipsoid (sphere) or/and geoid
- An ellipsoid (sphere) has to be projected onto a flat surface

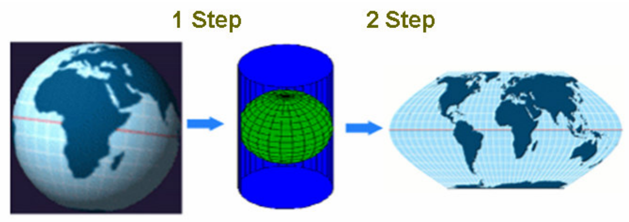


Figure: Two steps of the globe geo-referenced modeling

We have discussed the first step of this modeling and transformation in Module 3. In this module we discuss the mathematical projection process.

Two types of coordinate systems are used for the transformation of Earth to globe, and from globe to map. They are geographic and projected coordinate systems. Thus, a map uses a planar coordinate system to depict entities on a flat surface. A map planar coordinate can be calculated from a geographic coordinate of a globe and backwards.

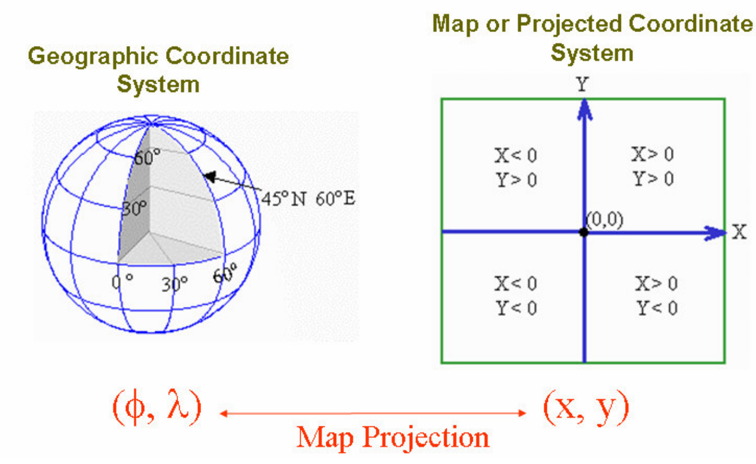


Figure: Geographic and projected coordinate systems

A map planar coordinate system is defined by a pair of orthogonal (x,y) axes drawn through an origin. The origin of a map or projected coordinate system can be defined from any point on the geographic coordinate system. The origin of a map coordinate system does not necessarily have to be in the origin of geographic coordinate system.

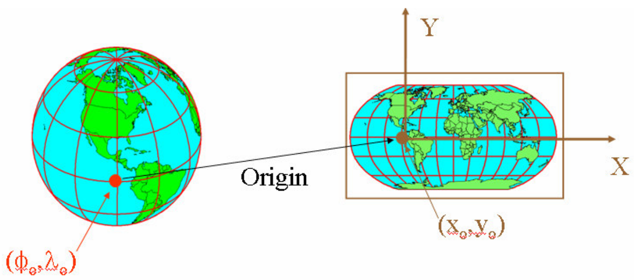


Figure: Origins of geographic and map coordinate systems

5.2 Map Projections

Map or cartographic projections are methods for projecting the spherical coordinates of latitude and longitude to orthogonal coordinates. One may think of this as a metaphorical projection of light through a transparent globe onto a *developable surface* such as a flat piece of paper.

However, strictly or mathematically speaking, a map projection can be represented as a system function or equation:

$$\mathbf{x} = \mathbf{F_1}(\phi,\lambda)$$
$$\mathbf{y} = \mathbf{F_2}(\phi,\lambda)$$

Where **x** and **y** are orthogonal map coordinates, ϕ and λ are angular geographic coordinates of a sphere or ellipsoid, and $\mathbf{F_1}$ and $\mathbf{F_2}$ are projection functions.

All projections have mathematical formulas that define the relationship between the latitude/longitude graticule on the Earth and its representation on the map sheet, or the relationship between the geographic coordinates (latitude and longitude) of points and their projected coordinates (grid or rectangular coordinates) on the projection or map. The mathematics of maps is based on a discipline called *differential geometry*.

Information from a globe can be projected onto a flat surface by using different classes of light sources. One can think of a light shining from inside or outside the Earth that projects an image of the globe onto a flat surface. As noted above, projections are determined mathematically.

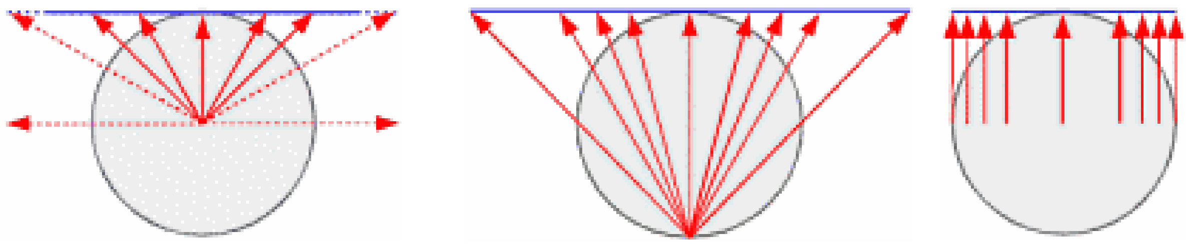


Figure: Three classes of projections by light source (the three possibilities for the projection center when the projection plane is a plane tangent to the sphere at a point)

Cartographic (map) projections try to flatten the curved surface of the globe without stretching or tearing it. However, since all map projections attempt to represent the curved surface of the Earth on a flat sheet of paper *distortions* are inevitable. The degree of distortion differs from point to point on the sphere and from map to map.

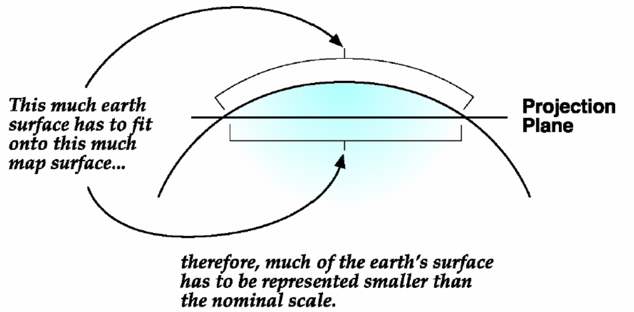


Figure: Map projections attempt to represent the curved surface of a globe onto a flat surface, however, distortions on the resulting map are unavoidable

Any map projection has its area of least distortion. Projections can be shifted around in order to lessen the degree of distortion over an area of user interest. Thus, any projection can have an unlimited number of variations or cases to determine the projection's parameters.

5.3 Map Scale

Map scale is the amount of reduction that takes place in going from real-world dimensions (sphere or ellipsoid) to the same area projected onto a map plane. Due to the projection distortions, map projections **do not preserve** a constant scale everywhere on a map.

There is **principal** map **scale** is defined as the ratio of globe distance to Earth distance, with each distance expressed in the same units of measurement. Principal map scale is defined for map places (points or lines) where there are no the projection distortions. For all other map places with projection distortions, **local scale** can be used to describe distortion in a point with consideration of the ratio of map distance and correspondent globe distance. We will discuss local scale in a general framework in following sections.

The inevitable distortion on a map differs from point to point on the sphere, and from map to map. These concepts will discussed within a general framework below.

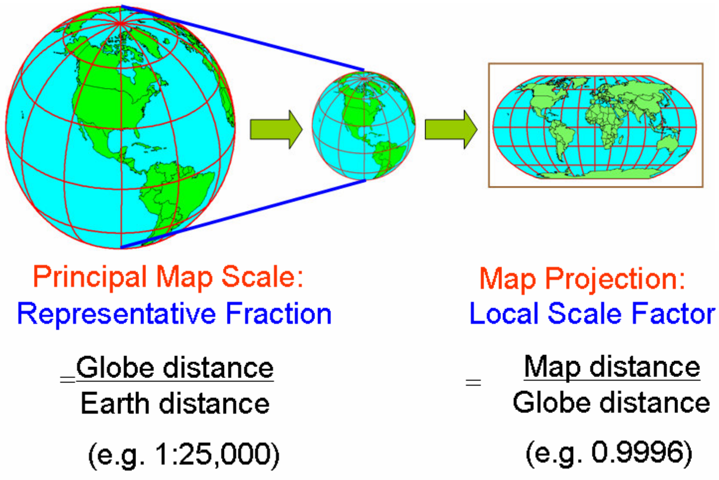


Figure: Principal and local map scales

Principal Scale is the ratio of map distance to ground distance.

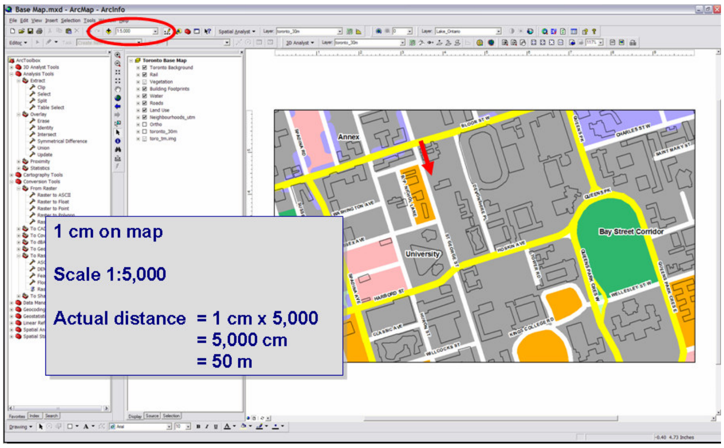


Figure: Principal map scale

Principal scale can be expressed in different forms on a map such as:

- Representative fraction (RF)
- Statement scale
- Bar scale

Representative fraction (RF) is shown as ratio. For example,

Always 1

- 1:250,000 or $\frac{1}{250{,}000}$ $\frac{\text{Map distance}}{\text{Ground distance}}$

RF always uses same linear units $\frac{1\ \text{mm}}{250{,}000\ \text{mm}}$.

Statement scale uses words to state ratio: “1 mm to 1 km” (1:1,000,000) or “one inch to one mile” (1:63,360). In addition, it incorporates unit conversion.

Bar scale is used to estimate scale on a map.

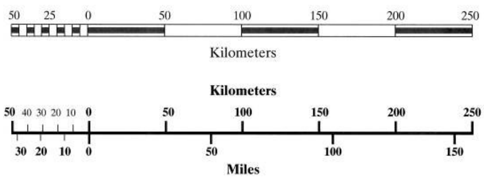


Figure: Bar scales

People often confuse expressions of “large” and “small” scales. The larger the area shown on a map, the smaller the scale. For example, a scale of 1: 10,000 is considered a large scale map and a scale of 1: 1,000,000 is considered a small scale map.

$\frac{1}{10}$ Large ratio

$\frac{1}{10{,}000}$ Smaller ratio

$\frac{1}{10{,}000{,}000}$ Really small ratio

Figure: Large Scale vs. Small Scale Maps

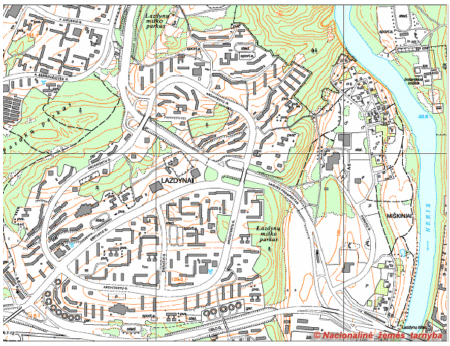


Figure: Large Scale Map 1 : 10,000

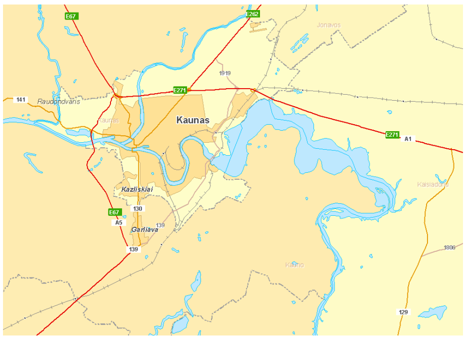


Figure: Smaller Scale Map 1 : 200,000

Figure: Small Scale Map 1 : 10,000,000

One of the useful functions of GIS software is that it can display a map layer within a predefined scale range. For example, an urban area that extends over 5 km can only be displayed on maps with scales between 1:250,000 and 1:3,000,000, using specific map symbology.

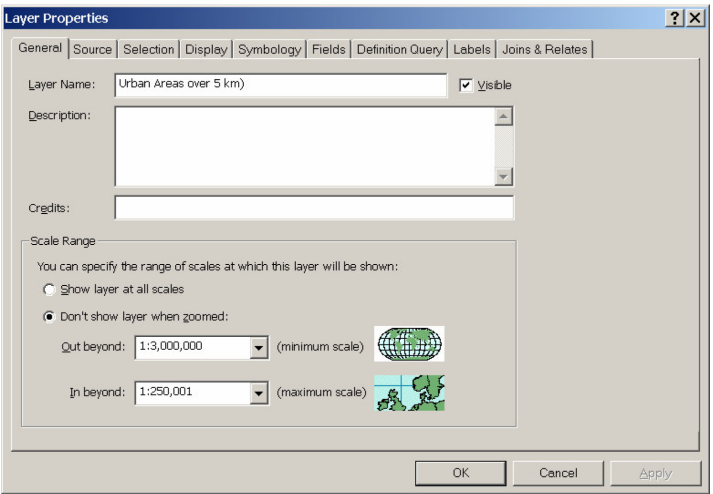


Figure: Scale range for display within ArcMap layer control dialog

A scale affects how entities are represented and scales define level of represented details on a map or level of generalization. Some maps (e.g. topographic) are created in a series of map scales

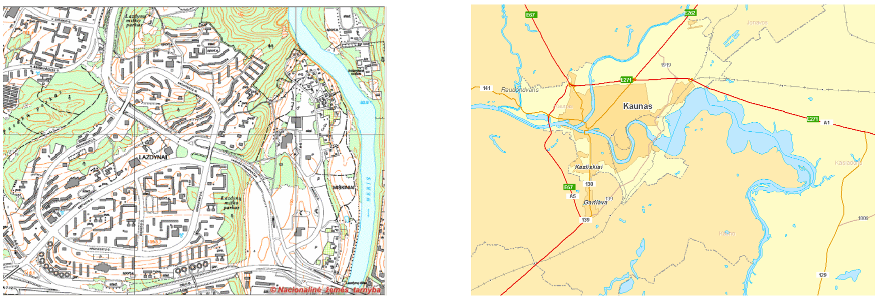


Figure: Two levels of representation of Kaunas

5.4 Map Projection Classification

Map projections are classified by:

- Shape of normal graticule (meridians and parallels or developable surfaces)
- Distortions (map properties)
- Aspect of developable surface

5.4.1 Classification by Shape of Normal Graticule

Projections by shape of normal graticule is based on meridians and parallels when the central axis of the developable surface is oriented north-south.

The most common projections based on shape are:

- Cylindrical projections
- Conical projections
- Azimuthal or planar projections (gnomic, stereographic, orthographic)
- Polyconic projections
- Pseudo-cylindrical projections
- Pseudo-conical projections
- Pseudo-azimuthal projections
- Arbitrary projections

Three types of map projections are used to project different *developable surfaces. These* include (1) cylinders, (2) cones or (3) planer surfaces.

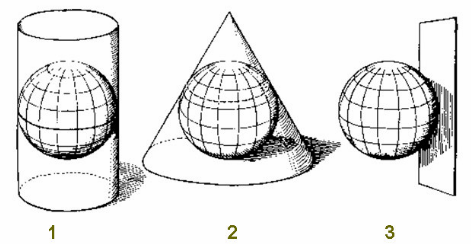


Figure: Three types of projected *developable surfaces*

Cylindrical projections are based upon the projection of a globe onto a cylinder. A classic example of such projection is the Mercator projection.

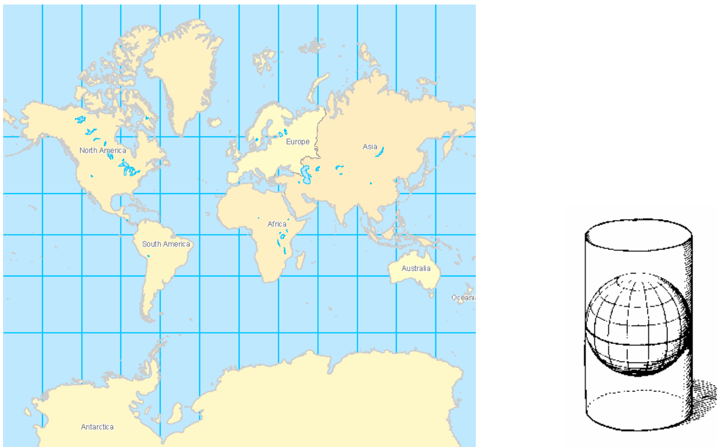


Figure: The Mercator projection and its developable surface

There are two cylindrical projection variations: 1) variation of projections where one line of the globe touches the cylinder and 2) where two lines of the cylinder intersect the globe. These lines are called standard parallels. A standard parallel is the circle of latitude where a cylinder touches or intersects the globe. This parallel(s) will be at true scale with increasing distortions to the north and south (in case of one standard parallel) and distortion variation between two standard parallels (in case of two standard parallels). If one standard parallel is used, it is normally the equator.

Meridians and parallels of normal cylindrical projections are respectively parallel lines that are orthogonal to each other.

A **conical** projection projects information from the globe onto a conic surface. Meridians of normal conical projections are straight radial lines. The parallels are sectors of circles that are centered in the point of meridian's intersection.

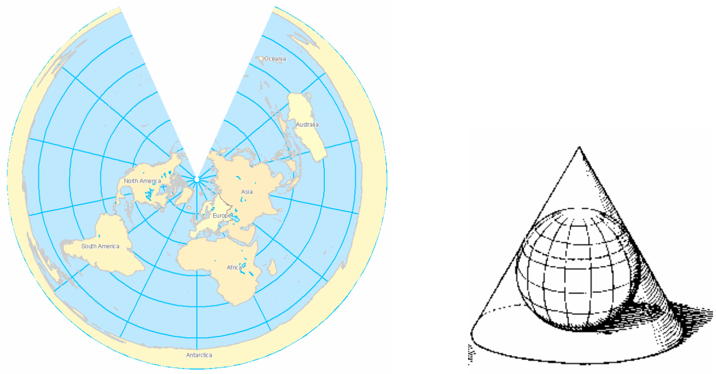


Figure: A conical projection and its developable surface

There are also two variations of conical projections when one and two lines (standard parallels) touch or intersect the globe on the conic surface. A standard parallel is usually the circle of latitude when a cone touches or intersects a sphere that the cone is put over. This parallel will be at true scale, with increasing distortions to the north and south.

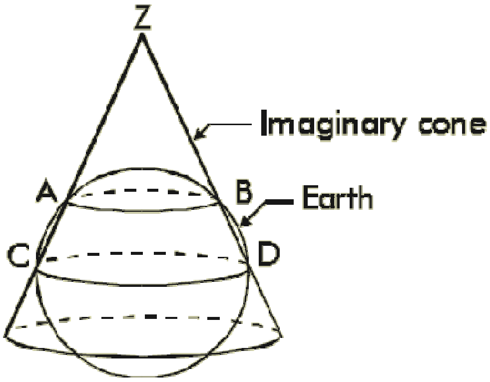


Figure: A conical projection with two standard parallels where two lines (AB and CD) both intersect the globe

Azimuthal or **Planar** projections project information from the globe onto a flat surface when only a point of the planar surface touches the globe. The projection plane is a plane tangent to the sphere at a single point. Meridians of normal azimuthal projections are straight radial lines, and parallels are circles that are centered in the point of the meridian's intersection.

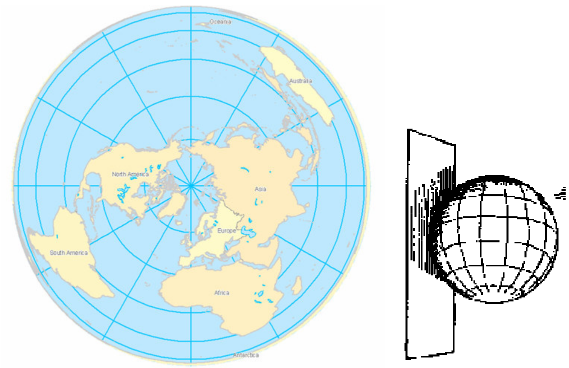


Figure: A Lambert equal-area azimuthal projection and its developable surface

Three possibilities for this type of projection, using differing central points on the globe, are: gnomic, stereographic, and orthographic.

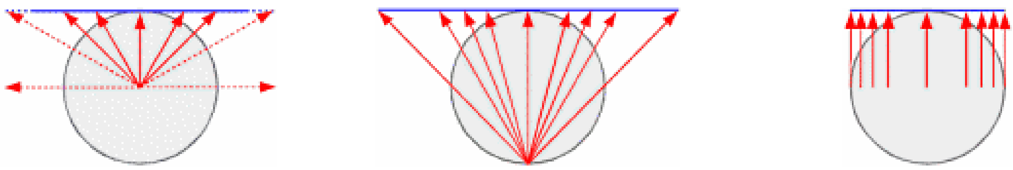


Figure: Three possibilities for the projection center: gnomic, stereographic, and orthographic projections

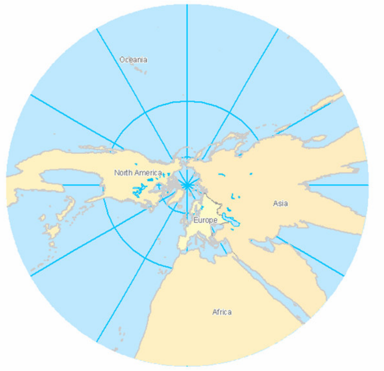


Figure: Gnomic planar - the projection center is located in the center of globe

Figure: Stereographic planar- the projection center is located in the surface of globe and opposite to a plane surface

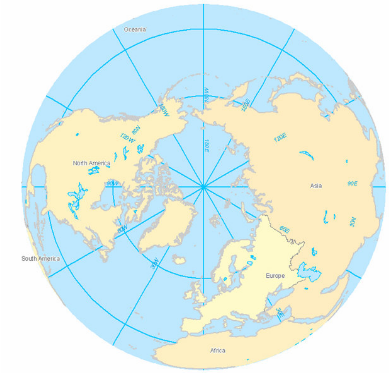


Figure: Orthographic planar - - the projection center at infinity and opposite to a plane surface

Meridians and parallels in normal **polyconic** projections produce some curved meridians and parallels, with the exception of the prime meridian and equator, which remain straight lines.

Figure: Polyconic projection

Figure: Aitoff polyconic projection

Meridians of **pseudo-cylindrical** projections produce some curved meridians with straight lined parallels, with the exception of the prime meridian (also a straight line).

Figure: Pseudo Cylindrical Robinson projection provides solution for the "Greenland via Australia Problem" – relative sizes of the landmasses are preserved

Arbitrary map projections do not fall into standard categories. Some examples are shown below.

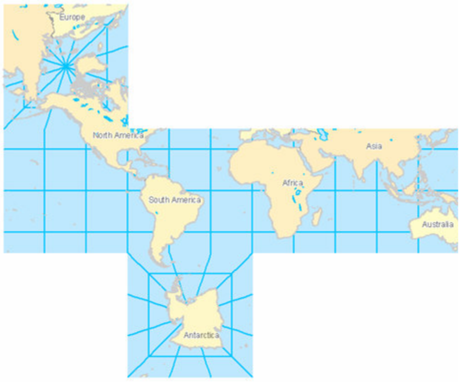


Figure: Cube projection

Figure: Bonne projection

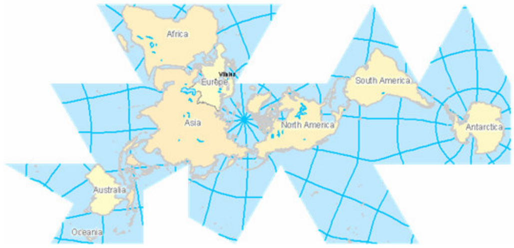


Figure: Fuller projection

5.4.2 Classification by Distortions or Projection Properties

It is not possible to completely avoid the distortion of length, in all directions, in any one projection. However, projections can preserve the following Earth properties on a map:

- Local scale or distances along particular lines - the scale distance cannot be kept a constant all over the entire map, however, one can maintain:
  - Constant scale along meridians
  - Constant scale along parallels
- Conformity of shape - a map preserves shape over a small area if it is conformal or orthomorphic ("right" in Greek**)**
- Local area distortions (an equivalent or equal-area map) - the area represented is in proportion to the area it represents
- Angle distortions - compass directions are "right" or "correct" everywhere on a map

In order to define projections by distortion types (conformal, equal-area, or equidistance), the degree of distortion has to be quantified. The local scale of lengths along parallels and meridians can be defined as μ **= ds/dS** , where **ds** is local distance (infinitely a small liner segment) in the map projection, **dS** is correspondent local distance (infinitely a small liner segment) on the ellipsoid or sphere. In such cases, distortions of lengths in percentage (%) can be defined as $\upsilon = (\mu - 1)*100\%$**.** Distortion of lengths is a function of geographic coordinates **(**ϕ, λ**).** Thus if $\upsilon = 1$ then there is no distortion in length, if $\upsilon = 0$ then the projection representation has disappeared, but $\upsilon \neq 1$ is somewhere on a map**.**

When the local scale of an area is defined as **p = df/dF,** where **df** is the local area in a map projection and **dF** is the correspondent local area on the ellipsoid or sphere, then the distortion of the areas in % can be defined as $\upsilon_p = (p - 1)*100\%$**.** Distortion of areas is also a function of the geographic coordinate **(**ϕ, λ**).**

An angle's distortion is defined as Δ**u = u - U,** where **u** is the value of an angle in a map projection and **U** is value of a correspondent angle on the ellipsoid or sphere. Maximum distortion of angles ω in a point is used to define the angle distortion in a point.

The values of distortions in lengths, areas and angles are used as a main criterion for classifying and choosing map projections for specific regions.

An easy way of visualizing the distortion of a map at a given point (*P*) is Tissot's indicatrix, which is an ellipse centered at a point. If the indicatrix is a circle, then the map is conformal (angle preserving) at a point, otherwise the main axes of the ellipse give the directions of maximal **a** and minimal **b** length distortions. All indicatrices, though they are of a different shape, have the same area.

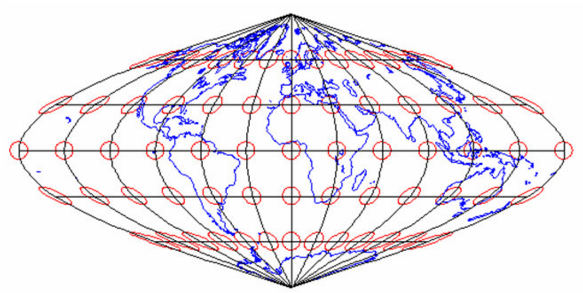


Figure: Tissot's indicatrix for Sanson's projection. The image is undistorted along the equator and the Greenwich meridian. Sanson's projection is area preserving.

In order to preserve shape, area and/or distance (length), key **map properties**, projections can be classified as:

- Equal-area or equivalent projections – relative areas are preserved
- Conformal projections – directions and relative shapes are preserved
- Equal-distance projections – relative distances (lengths) are preserved along particular lines
- Compromise projections - attempts to minimize distortions in areas and directions

Conformal projections preserve <u>directions</u> and relative <u>shapes:</u>

- Maximum angle of distortion ω **= 0** and **p** $\neq$ **1**
- **m = n = a = b =** μ**,** where **m** - local scale of lengths along meridians, **n** - local scale of lengths along parallels, **a** and **b** are respectively maximum and minimum local scales of distances in a map point
- Local scale of area is $\mathbf{p = a^2}$

In conformal projections, a circle on a globe will remain a circle on the map, but the relative scale or size may change. For example, if projections preserve a continents' shapes then the size of the continent is distorted. Conformal projections are used to create maps for different types of navigation.

Figure: Conformal Mercator (cylindrical by shape of graticule) projection. Greenland looks larger than Australia

Figure: Conformal Transverse Mercator (cylindrical by shape of graticule and transverse by aspect of developable surface)

Equal-area or equivalence projections preserve areas and depict correct relative size

- Local scale of area is **p = constant** and often **p = 1**
- Maximum angle of distortion ω **> 0**
- **a = 1/b**

In these projections, if South America is eight times larger than Greenland on the globe, it will also be eight times larger in the projection. Australia will be larger than Greenland. However, if sizes of continents are correct, shapes and directions will be greatly distorted.. Equivalence projections are used for cadastral purposes, area-measurements and visual comparisons.

Figure: Cylindrical equal area projection

Figure: Azimuthal equal area projection

One cannot make a single projection over which all distances are maintained. However, **equidistant** projections maintain relative distances from one or two points only. The length of a straight line between two points represents the great circle distance between them and along this line μ **=1.** In this type of projection **a = p** or **b = p**. This means that the scale of lengths is a constant along one of the major distortion directions **a = 1** or **b = 1.** Shapes and sizes may be distorted in equidistant projections.

If a graticule of a projection is orthogonal, where **a** or **b** coincide with **m** or **n,** such projections are called equidistant projections along meridians or parallels respectively.

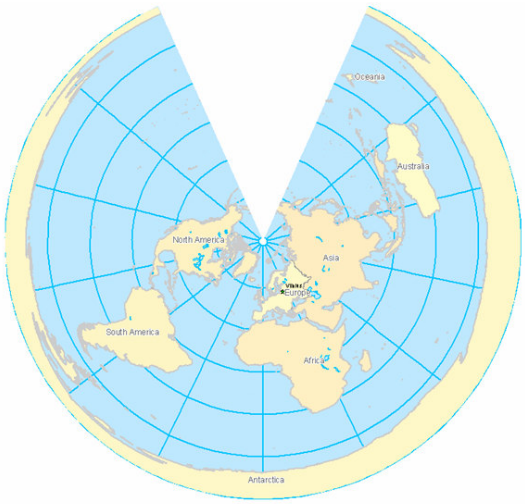


Figure: Equidistant Conic projection

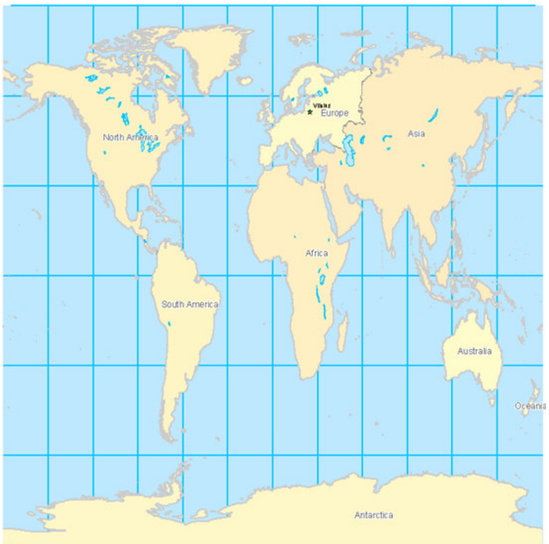


Figure: Equidistant Cylindrical projection

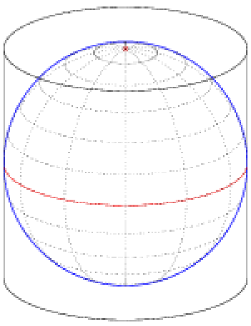
Compromise projections fail to preserve map properties - they are not conformal, equal area, equidistant, or true-direction. There are all kinds of distortions: **p** $\neq$ **1,** ω **> 0, and** μ $\neq$ **1.**

Figure: Robinson Projection is a compromise projection. Such a projection provides a solution to the "Greenland via Australia Problem"

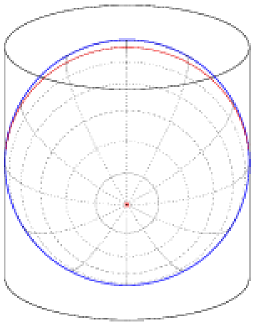
5.4.3 Classification by Aspects of Developable Surface

Map projection's classified on the basis of *aspect,* assume distinct orientations of projection developable surfaces. Map projections can be oriented to the following aspects:

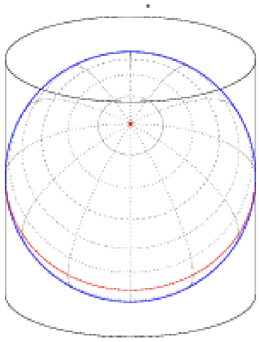
- Normal - central axis of the developable surface is oriented north-south
- Transverse - central axis of the developable surface is oriented east-west (perpendicular to the earth's axis)
- Oblique - assumes a variety of *oblique* aspects



Normal

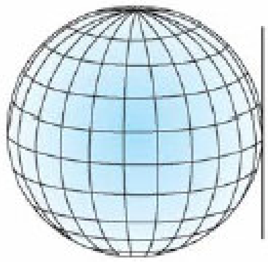
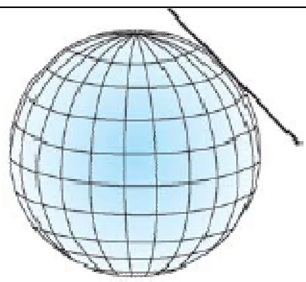


Transverse



Oblique

Figure: Cylindrical projections by orientation

		
Polar	**Equatorial**	**Oblique**

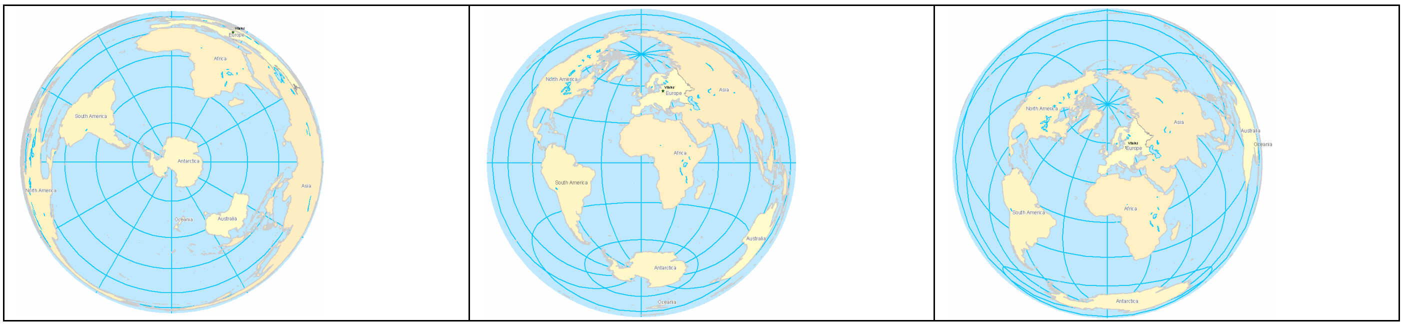


Figure: Azimuthal Projections by Orientation

Figure: World from space in Oblique Orthographic projection

5.5 Selected Map Projections

In this section, the following commonly used projections are discussed:

- Mercator projection
- Transfer Mercator projection
- Gauss-Kruger projection
- European projections

The Mercator projection is a conformal projection where the equator is the single standard parallel. The meridians are equally spaced vertical lines, and the parallels are horizontal lines whose spacing increases towards the poles. This projection is rarely used for land mapping purposes but is universally used in the production of navigation charts. It provides the basis for the transverse and oblique forms of the Mercator projection.

As well as being conformal, it has the particular property that straight lines drawn on it are lines of constant bearing. This is called a *rhumb* line or *loxodrome*. Thus, navigators may derive their course from the angle the straight course line makes with the meridians.

There are two classes of special curves on the sphere (or ellipsoid) which are *loxodrome* and *orthodrome*.

An orthodrome shows the shortest distance between two points on the globe. An orthodrome on the sphere is a segment of a *great circle* (or *great ellipse*), that is., the intersection of a plane through the center of the sphere with the sphere itself.

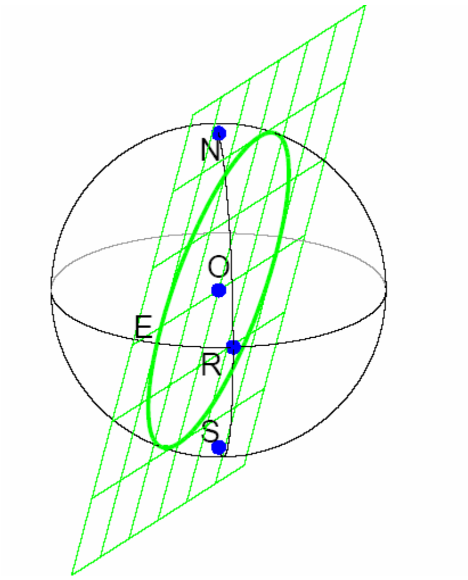


Figure: A great circle

The second important type of curve is the rhumb line or loxodrome. A loxodrome is a curve which intersects all of the meridians of longitude at the same angle.

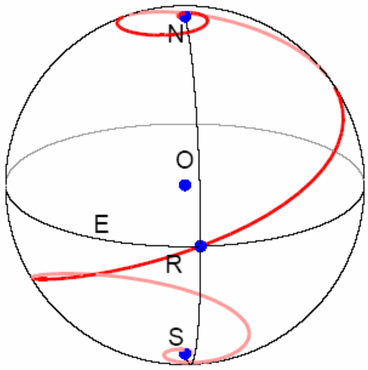


Figure: A loxodrome

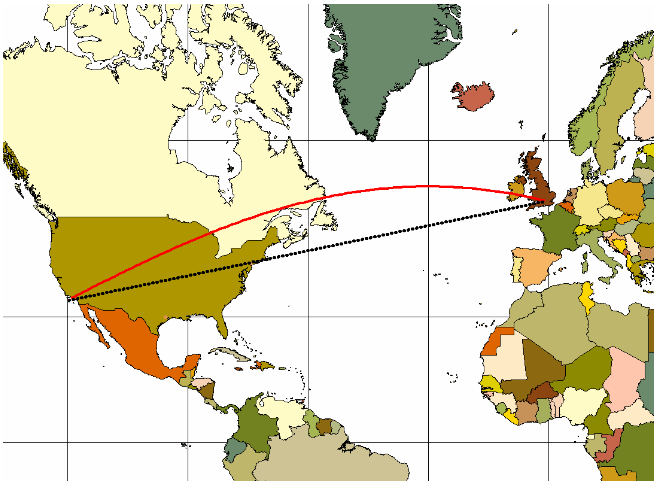


Figure: An orthodrome and loxodrome

The formula for producing a Mercator projection for a sphere is:

x = R*Ln(Tan(450 + ϕ/2)
y = R* λ
ω = 0

Where, **R** is radius of the Earth sphere.

Figure: The Mercator projection

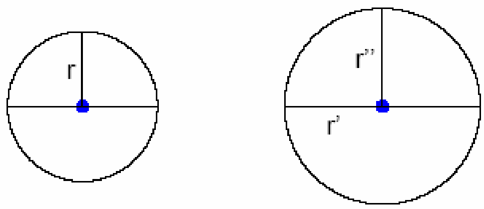


Figure: Distortions in Mercator projection: r = R sec ϕ

The **Transverse Mercator (TM)** projection, in its various forms, is the most widely used projected coordinate system for world topographic and navigational chart mapping. All versions of TM have the same basic characteristics and formulae. However, different forms of the TM projection are applied in different countries. Variations in the parameters used include:

- The latitude of the origin
- The longitude of the origin (central meridian)
- The scale factor at the origin (on the central meridian)
- The values of False Easting and False Northing, which change the coordinate origin relative to the projection origin – done in order to avoid negative coordinate values
- Possible variations in the width of the longitudinal zones for the projections used in different territories. Usually it is 6° zones.

The Transverse Mercator is similar to a Mercator projection except it is turned 90° so that it is related to a central meridian and it does not retain the straight rhumb line property of the Mercator projection.

Figure: The Transverse Mercator projection

There are two variations of a Transverse Mercator projection: with one and two lines (standard meridians) touching or intersecting the cylinder and globe.

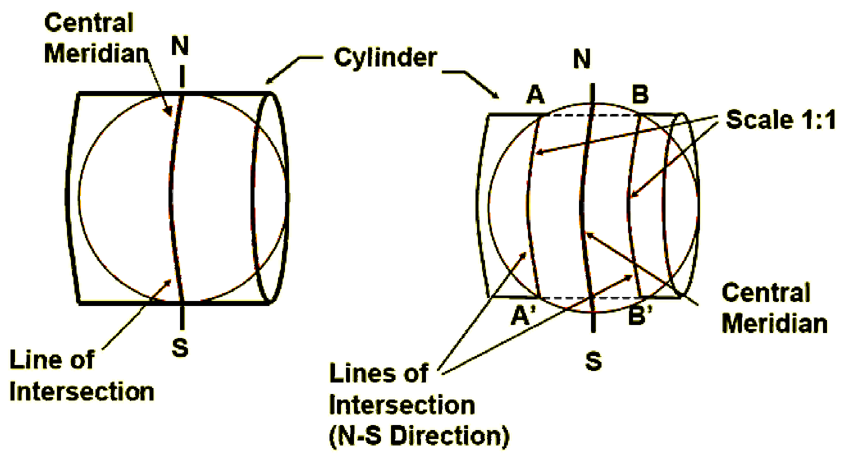


Figure: TM projections with single standard meridian and double standard meridian

In many countries, TM projection employs two standard meridians. In this case, the scale factor (μ) at the origin (on the central meridian) equals to 0.9996.

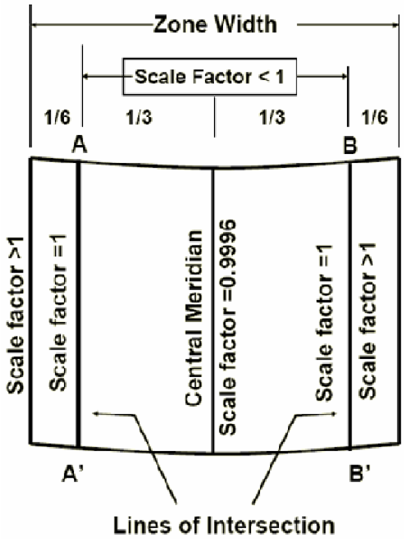


Figure: Scale length factor distribution in TM projection within 6° zone.

TM coordinates are measured in meters. By convention, the central meridian is assigned a value (*false easting*) of 500,000 m. The 500,000 m is added to all **x**-coordinate values. In the northern hemisphere, we measure values (northings) from the equator. In the southern hemisphere, coordinates are measured north to the equator, which is arbitrarily given a *false northing* of 10,000,000 m. The 10,000,000 m is added to all **y**-coordinate values.

The Lithuanian Coordinate System LKS-1994 uses the Transverse Mercator projection with the following parameters:

- False Easting: 500000,000000
- False Northing: 0,000000
- Central meridian: 24,000000
- Scale factor: μ **=** 0,999800
- Latitude of origin: 0,000000 (equator)
- Linear unit: Meter

The Gauss-Kruger projection is similar to the Transverse Mercator projection except that there is only one meridian that touches the cylinder from the globe. In this case, the scale factor $\mu = 1$ on the touching meridian - the central meridian.

Figure: Gauss-Kruger projection

The following projections are commonly used to show Europe on a map:

- Europe Lambert Conformal Conical projection
- Europe Albers Equal Area Conical projection
- Europe Equidistant Conical projection

European **conical** projections use two standard parallels. The choice of the two standard parallels will usually be made according to the latitudinal extent of the area that one wishes to map. The parallels usually chosen this way lie proportionally inboard of the north and south margins of the mapped area. The projections with two standard parallels normally maintain the nominal map scale along the two parallels of latitudes that are the lines of intersection between the imagined cone and the ellipsoid or sphere.

The **Lambert conformal conical** projection maintains key conformal properties. The spacing of the parallels is variable and increases with increasing distance from the standard parallel, while the meridians are all straight lines radiating from a point on the prolongation of the spheroid axis or ellipsoid's minor axis.

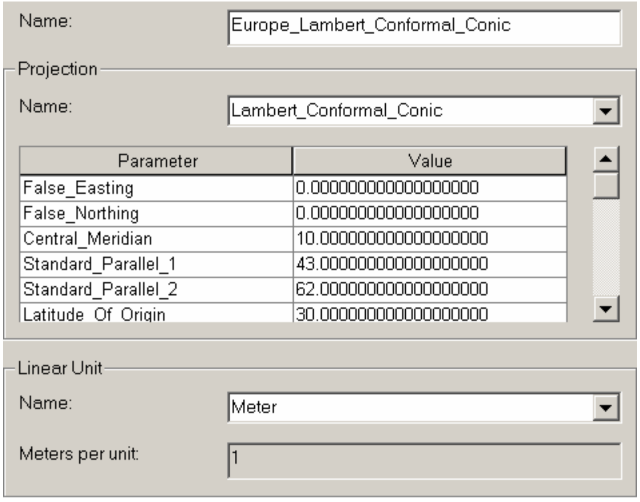


Figure: Parameters of Europe Lambert Conformal Conical projection

Figure: Europe Lambert Conformal Conical projection

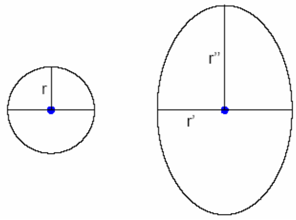


Figure: Distortion of conformal conic projections

The **Albert equal area conical** projection maintains the equivalent properties – the spacing of the parallels is variable and decreases with increasing distance from the standard parallel, while the meridians are all straight lines that radiate from a point on the prolongation of the spheroid axis or ellipsoid's minor axis.

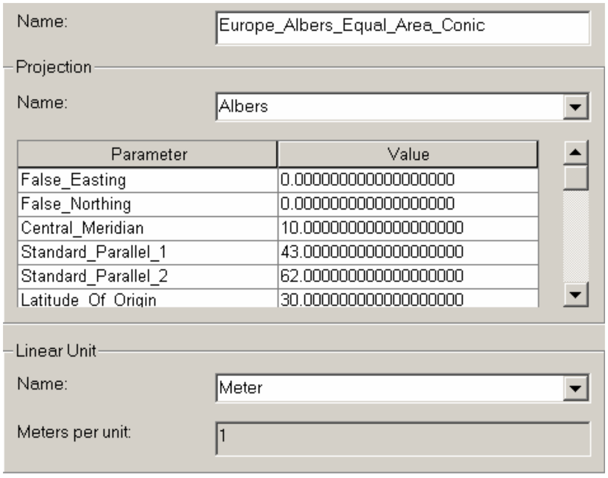


Figure: Parameters of Europe Albert equal area

conical projection

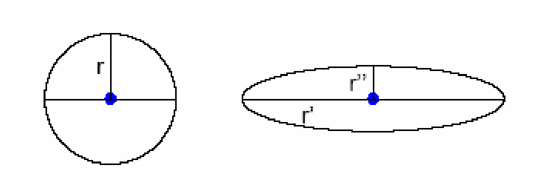


Figure: Distortion of equal area conical projections

Figure: Europe Albert equal area conical projection

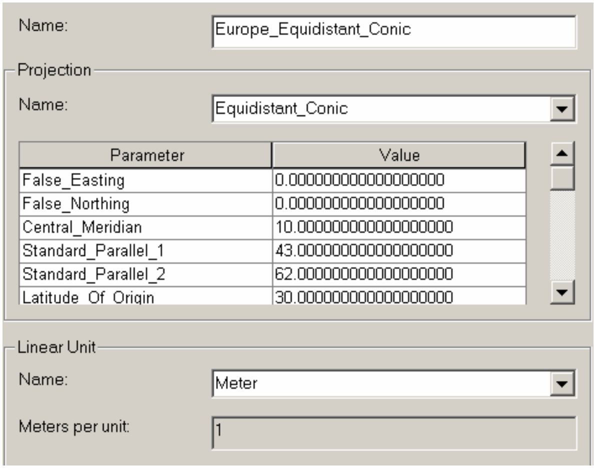


Figure: Parameters of Europe equidistant conical projection

Figure: Europe equidistant conical projection

The Equidistant conical projection maintains the equidistant property *only* along all meridians and along one or two standard parallels. Directions, shapes and areas are reasonably accurate, but distortion increases away from the standard parallels. Maps that use this projection are not conformal or equal area, but a compromise between the Lambert Conformal Conic and Albers Equal Area Conic projections.

Conical and transverse cylindrical projections are also can be used for mapping the Baltic Region. Some samples of these projections are shown below:

Figure: Albers equal area conical

Central meridian: 20.000000

Standard parallel 1: 56.000000

Standard parallel 2: 62.000000

Latitude of origin: 40.000000

Origin scale: 1 6 000 000

Figure: Lambert conformal conical

Central meridian: 20.000000

Standard parallel 1: 56.000000

Standard parallel 2: 62.000000

Latitude of origin: 40.000000

Origin scale: 1 6 000 000

Figure: Cylindrical equal area

Central meridian: 0.000000

Standard parallel: 60.000000

Origin scale: 1 6 000 000

5.6 Standard Map Coordinate Systems

Standard map coordinate systems use particular projections with particular parameters over **zones** of the Earth's surface. There are different types of standard coordinate systems that are used by different countries.

The most common standard state map coordinate system is the Universal Transverse Mercator (UTM) that is set of 6° wide zones that cover the entire globe or a particular state or region. UTM is piecewise (consist of pieces) coordinate system that comprises 60 zones for global coverage using the Transverse Mercator projection. UTM uses the Transverse Mercator projection with particular parameters for each zone.

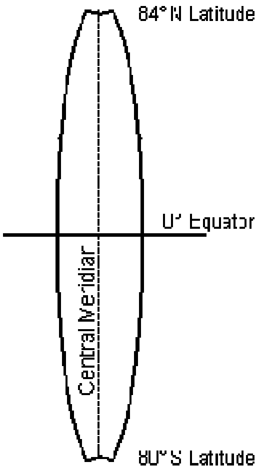


Figure: UTM zone

The UTM is the base map coordinate system adopted by many countries when producing national topographic maps. Each 6° wide zone has a central meridian (λ_o), and goes from pole to pole. The zone's origin is the International Date Line (Standard Meridian) and proceeds eastward, so that Zone 1 is 180º-174ºW, Zone 30 is 6º-0ºW, Zone 31 is 0º-6ºE and Zone 60 is 174º-180ºE. A total of 60 zones cover the Earth from east to west.

In this system, the Reference Latitude (ϕ_o) is the equator. The UTM is again using false easting (Xshift) and northing (Yshift) so it never has a negative coordinate. False easting and northing is expressed in meters. A false easting of 500,000 meters is assigned to the central meridian for each zone. For the southern hemisphere, a false northing of 10,000,000 meters is given to the equator. A **30'** overlap is provided between zones.

Commonly, a UTM uses the Transverse Mercator secant case projection (two standard meridians), with the scale factor of 0,9996 at the zone central meridian. However, there are systems that use different scale factors and/or one standard touching meridian.

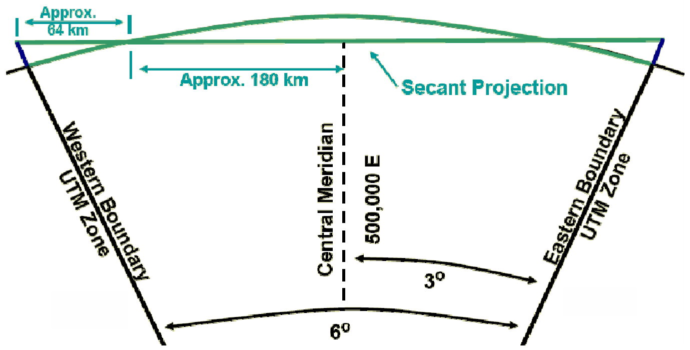


Figure: Structure of a UTM zone (profile)

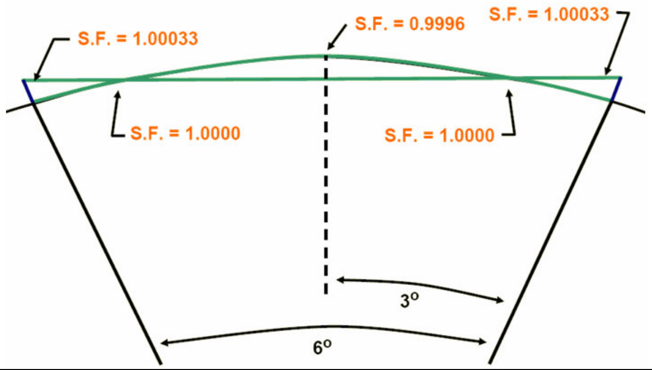


Figure: Distribution of scale length factors within UTM zone. Scale factor for the central meridian is 0.9996 (but it is not an absolute rule)

UTM zones are formed from a rotated cylinder. The relationship between a zone central meridian and zone number is $\boldsymbol{\lambda_o = (N - 30)*6^0 - 3^0}$, where **N** is a zone number.

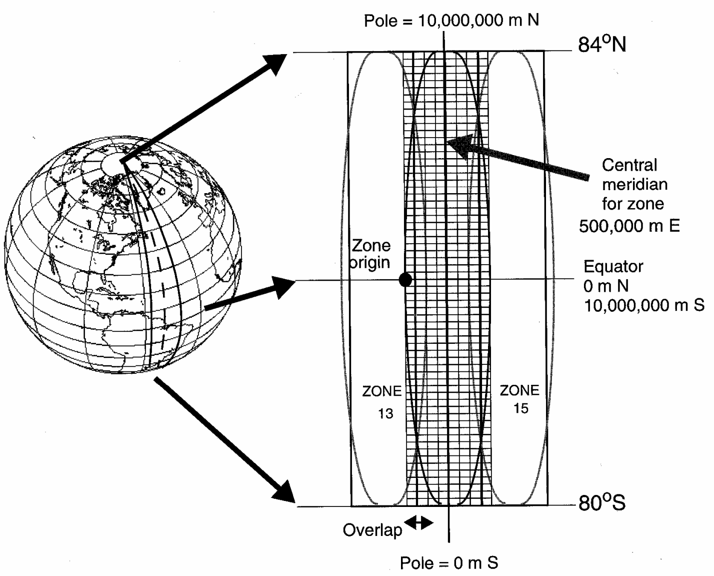


Figure: A UTM zone's formation

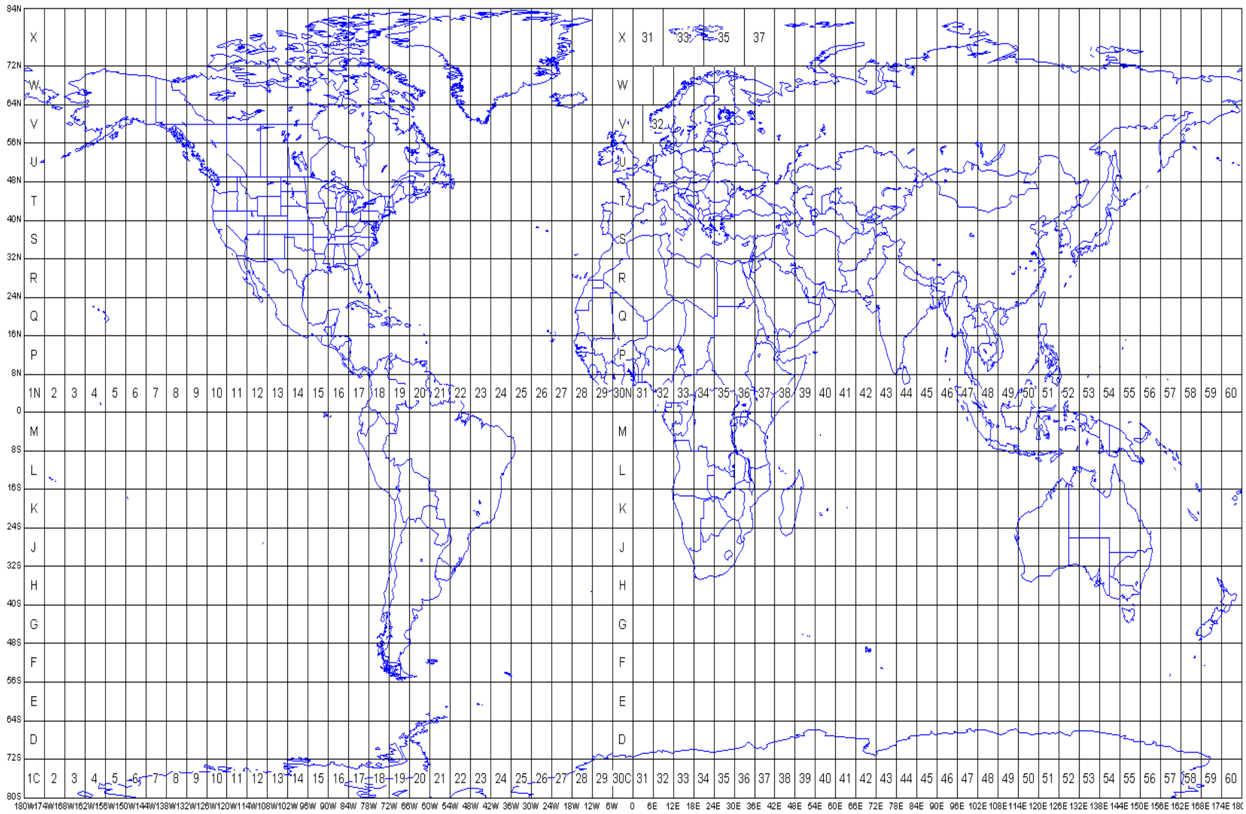


Figure: UTM zone numbers of the World

The Lithuanian LKS-1994 system is not coinciding with standard UTM zones. The complete description of the Lithuanian Coordinate System 1994 is:

- Projection: Transverse Mercator
  - False easting: 500000,00
  - False northing: 0,0
  - Central meridian: 24,00
  - Scale factor in the central meridian: μ = 0,9998
  - Latitude of origin: 0,0 (equator)
  - Linear unit: meter
- Geographic coordinate system: GCS LKS 1994
  - Angular unit: degree (0,0174532925199433299)
  - Prime meridian: Greenwich (0,0)
  - Datum: d_lithuania_1994
  - Spheroid: grs_1980
  - Semimajor axis: 6378137,00
  - Semiminor axis: 6356752,314140356100
  - Inverse flattening: 298,2572221010000200
  - Vertical datum:
    - Baltic height system

UTM coordinate systems of UTM34N/UTM35N zones are used in Lithuania for nautical charts and military maps.

For territories with limited latitudinal extent, but wide longitudinal width, it may sometimes be preferred to use a single projection rather than several bands or zones of the Universal Transverse Mercator. If the latitudinal extent is also large, there may still be a need to use two or more zones if the scale distortion at the extremities of the one zone becomes too large to be tolerable.

5.7 Projections and GIS Software

In GIS, projections are often grouped into two categories: "Geographic" (or Lat/Long) coordinate system and Projected coordinate system. Thus, a spatial data can be displayed in GIS without a projection. GIS can only use geographic coordinate system parameters (ellipsoid and datum) and plots longitude and latitude as a rectangular coordinate system. So called "Lat/Long projection" has the following system of equitation: $\mathbf{x} = \lambda$ and $\mathbf{y} = \phi$. In this "projection", distance, area, and shape are all distorted.

A projected coordinate system uses geographic coordinate system parameters and projections with parameters (e.g. origin, central meridian etc.).

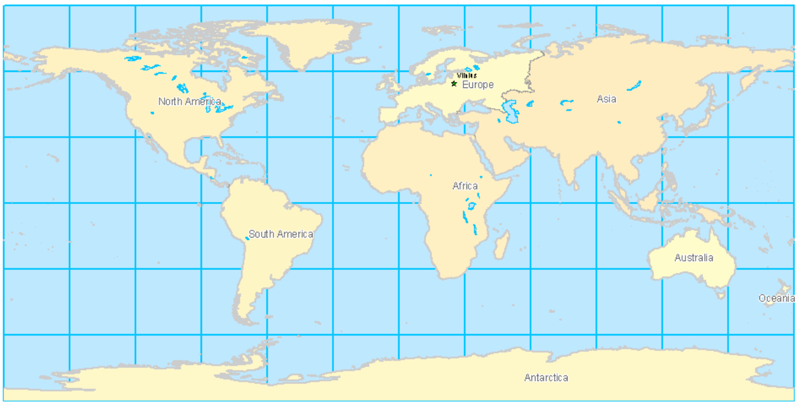


Figure: "Geographic" (or Lat/Long) coordinate system with parameters of WGS 1984

Some GIS software is able to change or display projections interactively by re-projecting one projection to another "on the fly". Coordinates within the actual data file are not changed during such "on the fly" re-projection.

Almost all GIS software has the ability to re-project actual data files from one projection to another projection permanently. Coordinates in the actual data file are recalculated and adjusted to match the projection desired.

Careful attention must be paid to the projection, datum and coordinate system for every piece of GIS data used. Failure to use data from the same system or change the data (re-project) it to the desired system will result in overlay errors. These errors can range some small to significant. The real danger is when the errors are small (possibly unnoticed).

In GIS files, images, grids all have projected data inherent in their very creation. Usually projection data are included in a system of files known as "metadata" (e.g. xxxxxx.prj files) or recorded within coverage or geo-database files.

5.7.1 Summary Concepts

To prepare a map, the Earth is first reduced to a globe and then projected onto a flat surface. A datum, a projection type and a set of parameters for projection define a particular map coordinate system. Standard map coordinate systems use particular projections over zones of the Earth's surface.

Types of standard coordinate systems include:

- UTM
- State Plane
- Others (too numerous to consider)

Do not confuse the standard map coordinate system with its projection.

The main (first) task of Mathematical Cartography is to select on appropriate projection. Consideration of a few factors can assist in selecting the appropriate map projection:

- Position, shape configuration and size of mapping regions:
  - For whole world, pseudo-conical, pseudo-cylindrical and other similar projections can be used
  - Equatorial and extended along the parallels regions are mapped in cylindrical projections
  - For polar areas, azimuthal projections are used
  - Mid-latitude compact regions are mapped with conical projections
- Purpose and applications of mapping:
  - Conformal projections are used for angular measurements applications such navigations
  - Equal-area projections are appropriate for area measurement applications such as cadastral, geographical comparisons. etc.
- Audience:
  - For example, for primary education for world map production, Robinson's World compromise projection shows right proportions of land masses

Various schemes and formulae have been developed to make this selection such that the maximum scale distortion within the mapped area is minimized.

A "good" map is one that is being successfully used for its intended purpose and was created in a precise and accurate manner. There is always a trade-off in the following projection errors:

- Shape and direction (conformal)
- Distance (equidistance)
- Area (equivalence)

It can only keep one or two of these properties accurate at one time. There is always one or more of these properties compromised. Errors may not be significant for small study areas but they do exist.

Module Self-study Questions:

- Briefly explain how a UTM zone is defined in terms of its central meridian, standard meridian, and scale factor.
- List the four types of map projections by the preserved property.
- Map projections involve a process of converting from:
  - a three-dimensional surface to a two-dimensional surface.
  - a two-dimensional surface to a three-dimensional surface.
  - a two-dimensional surface to a two-dimensional surface.
  - none of the above.
- The secant case means that a cylindrical projection has _____ line(s) of tangency:
  - 1
  - 2
  - 3
- Which of the following statements is **not** true about a meridian with a scale factor of 1?
  - The meridian must be a standard meridian.
  - There is no projection distortion along the meridian.
  - The meridian must be the line of 0^{o} longitude.
  - None of the above.

Required Readings:

- **Map Projection**, The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/mapproj/mapproj_f.html
- **Chapter 1-3**, Understanding Map Projections, ESRI digital books, 2004
- **Map Projections**, USGS, http://erg.usgs.gov/isb/pubs/MapProjections/projections.html

ESRI Virtual Campus Course:

- **Module 3 and 4: Understanding and Controlling Distortion Geographic** and **Planar Coordinate Systems,** Understanding Map Projections and Coordinate Systems

Assignment:

- **Assignment 5:** Working with map projections in GIS

References

- [1] *Elements of Cartography*, Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, 6th ed., NY: John Wiley & Sons Inc., 1995Canadian Spatial
- [2] *Thematic Cartography and Geographic Visualization*, Terry A. Slocum, Robert B McMaster, Fritz C. Kessler, Hugh H. Howard, 2nd ed., 2004
- [3] *Map Projections: A Reference Manual*, L.M. Bugayevskiy and J. P. Snyder, 2002

Terms used

- Geographic and projected coordinates
- Map coordinate system
- Map projection
- Principal map scale
- Local scale
- Shape of normal graticule
- Developable surface
- Map properties
- Projection distortions
- Aspect of developable surface
- Cylindrical projections
- Conical projections
- Azimuthal or planar projections
- Equal-area or equivalent projection
- Conformal projection
- Equal-distance projection
- Compromise projection
- Orthodrome
- Loxodrome
- TM
- Origin
- Central meridian
- False easting and false northing
- Scale factor
- UTM
- UTM Zone

6 Cartographic Principles and Topographic Cartography

This module is an introduction to basic cartographic principles. Concepts such as map symbolization, dimension, levels of attribute measurement for symbolization, and visual language use in cartography are discussed. Map generalizations or abstractions of reality, including differing generalization technique and implementation procedures are also discussed.

This module also examines topographic map production. The mathematical basis of topographic mapping, including concepts of scale and coordinates are outlined. National topographic grids, elements presented on a topographic map, map symbology, and map specifications are discussed. The key steps in the production of topographic maps are examined. Elements of topographic maps of the Republic of Lithuania are reviewed.

<u>Module Outline</u>

- BASIC CARTOGRAPHIC PRINCIPLES
  - Entities and Symbology Dimensions
  - Levels of Attribute Measurements
  - Visual Variables
- GENERALIZATION
  - Driving Factors
  - Techniques
  - Algorithms
  - Implementations
- TOPOGRAPHIC MAPPING
  - Scale Series and Map Projections
  - Elements
  - Symbology
  - Generalization
  - Technological Processes

6.1 Some Cartographic Principles

Cartography is “The art, science, and technology of making maps, together with their study as scientific documents and works of art …” (The International Cartographic Association). As you may recall from a previous module, there are three main properties of maps:

- All maps are mathematical transformations of reality
- All maps use symbolization to represent reality
- All maps are abstractions of reality

We have discussed the mathematical base of maps in Module #5. In this module, we begin a discussion about map symbology and we will continue this discussion in subsequent modules. This module also examines map generalization processes – the third property of any map.

All maps use symbols and/or text to represent elements present in the real world. Symbology of this kind distinguishes maps from aerial and satellite imagery, which also have a mathematical base (scale and coordinate system) and are also abstractions of reality (natural abstraction), however, they do not use symbology.

We begin our discussion of mapping by examining formal descriptions of some properties of real world entities or phenomenon and how these are represented on a topographic map. Real-world features exist in two forms:

- Entity or object
    - Discrete and definite (e.g. a building, a road)
- Phenomena
    - Continuous distribution over an area (e.g., terrain, temperature)

Space is populated by discrete entities. To create a map, cartographers need to recognize and define a specific entity (e.g. “Is it a building or a road?”). In addition, one has to define the entity’s boundaries and location, and list the entity’s attributes. An example of discrete entitles are land parcels, wells, streets, oil spills, populated places, etc.

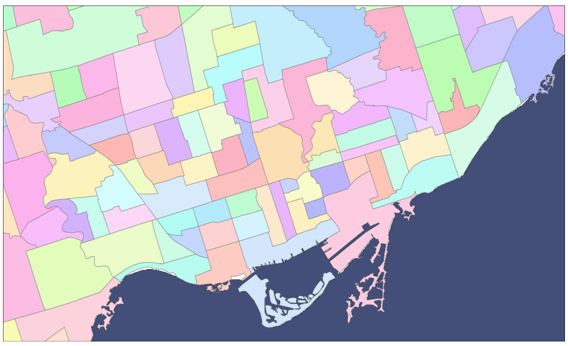


Figure: Land parcels as discrete entities

For a continuous field of any phenomenon, attribute values vary continuously over space. There is no single object. The examples geographic continuous phenomena would apply to the mapping of air pressure, air temperature, elevation, population densities, etc.

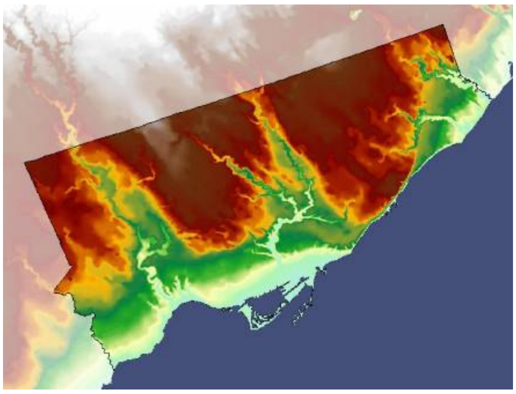


Figure: Elevation as continuous field of phenomena

6.1.1 Geographic and Cartographic Dimensions

The first step in mapmaking is to determine what types of geographic features one wishes to represent on a map. Real-world features can occur at points, along lines, or over areas. Real-world features can be described as having **geographic dimensions**. Thus, discrete real world entities have different spatial dimensions or extent:

- 0- or point-dimension, such a discrete entity is dimensionless and can be defined by geographic or/and map coordinates in 2-, 3- or 4- dimensional space
  - e.g. oil well
- 1- or linear-dimension, such a discrete entity has a length and can be defined by geographic or/and map coordinates in 2-, 3- or 4- dimensional space
  - e.g. boundary
- 2- or areal-dimension, such a discrete entity has a length and width and can be defined by geographic or/and map coordinates in 2-, 3- or 4- dimensional space
  - e.g. parcel
- 3- dimension, such a discrete entity has a length, width and height (or z- value) and can be defined by geographic or/and map coordinates in 2-, 3- or 4- dimensional space
  - e.g. cartogram map of incomes (z- values) per country
- 4- or time-dimension, such discrete entity has a length, width, height (or z- value) and time stamp and can be defined by geographic or/and map coordinates in 2-, 3- or 4- dimensional space
  - e.g. map of population density per district for 1992, 2002 and 2006

Continuous phenomena can be 3- or 4-dimensional:

- 3- or volumetric-dimension, such a field phenomenon has a length, width and height (or z- value)

  - e.g. elevation
- 4- or time-dimension, such a field phenomenon has a length, width, height (or z- value) and time stamp
  - e.g. population census density for 2002

In a spatial database, features are represented as objects that are a model of the real-world feature. Objects in the real world are represented as symbols on a map. Features have a shape and location; symbols may only represent some of those features characteristics, e.g. point symbol for building only represents building location and do not show a real building shape.

In GIS, three kinds of objects are represented:

- Points: An object that is too small to be shown as an area.
  - e.g. stop signs, coffee shops, crime locations
- Lines: An object that has length but is too narrow to depict as an area.
  - e.g. roads, pipelines, and rivers.
- Polygons: An object that has a visible extent in both length and width.
  - e.g. provinces, census tracts, parks.

Maps represent discrete entities and continuous phenomenon by symbols of different dimensions. **Cartographic** or conventional sign **dimensions** may not correspond to **geographic dimensions of respective** real world entities. A powerful tool of cartographic abstraction is to map with fewer geometric dimensions than the data exhibit. A dimension can be reduced during the cartographic modeling process. For example, a building can be mapped as a 0-demensional cartographic symbol that defines only the position of a building, but dimensionless in terms of building size. Sizes of the symbol are defined, but these sizes only depict cartographic representation of the symbols and do not correspond to real building dimensions. A contour line represents volumetric data, not linear data. Dimensions of symbols may not correspond to what they spatially express.

The following cartographic symbols' dimensions can exist:

- 0- or point-symbols - defined by geographic or map coordinates in 2-, 3- or 4- dimensional space. It represents an object that is too small to be shown as an area
  - e.g. a shaded gray circle represents an oil well, stop sign, coffee shop, or crime location
- 1- or linear-symbol - has length and defined by geographic or map coordinates in 2-, 3- or 4- dimensional space. An object that has length but is too narrow to depict as an area will be symbolized by a linear symbol
  - e.g. dashed black line represents a field road or trail
- 2- or areal-symbol - has a length and width, and defined by geographic or map coordinates in 2-, 3- or 4- dimensional space. An object that has visible extent in both length and width will be symbolized by an areal symbol
  - e.g. black rectangle represents a building

- 2.5- or pseudo-volumetric-symbol - has a length and width, and defined by geographic or map coordinates in 2-, 3- or 4- dimensional space. Height or time-changes are simulated by graphical means
    - e.g. brown contours lines are shaded by toned areas between contours, thereby, representing changes in elevation
- 3- or computer graphic volumetric-symbol - has a length, width and height, and controls changes in perspective, and defined by geographic or map coordinates in 2-, 3- or 4- dimensional space
    - e.g. flight simulation over a river valley
- 4- or animation-volumetric-symbols with time-series slices - has a length, width, height (or z- value) and time stamps, and defined by geographic or map coordinates in 2-, 3- or 4- dimensional space
    - e.g. computer animation of population census over a time period

In computer cartography and GIS, representations, using map symbols, are based on vector or/and raster graphic objects and their cartographic visualization (graphic variables). Thus, discrete entities are represented by:

- Vector objects such as points, lines or polygons or any combination of these three
- Raster grid objects such as a set of pixels that store integer values for discrete objects

Continuous phenomenon can be represented by:

- Raster grid objects such as a set of pixels that store floating values
- Triangulated irregular network objects such as a set of interconnected points, edges and faces

Table: Vector or/and raster graphic objects represent different dimension types of any feature or phenomenon in a spatial database

Feature or phenomenon types	Point Objects	Line Objects	Polygon Objects	Surface
Vector	No dimension	Length	Area	TIN, point DEM, set contours or hypsometric polygons
Raster	Single pixel	Cluster of integer pixels	Cluster of integer pixels	Floating grid

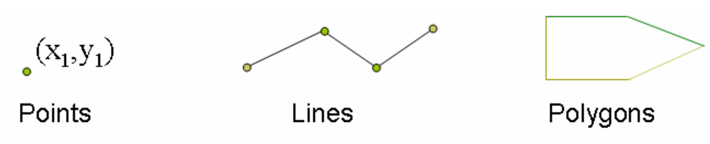


Figure: Discrete vector objects

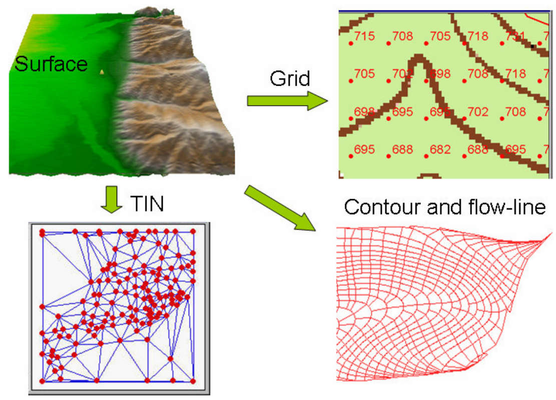


Figure: Continues representations by vector or raster objects

Vector or raster objects can be symbolized on the basis of their attributes by using different visual or graphic variables such as size, shape, orientation, texture or pattern, orientation pattern, hue of color, value or color, etc.

6.1.2 Levels of Attribute Measurements

Another challenge for the cartographer is to determine how the data are measured. Data can be described by a measurement level.

In digital cartography, each graphical object that corresponds to real-world entity can be linked to its attributes within a database. This attribute data can be either qualitative or quantitative. Mapmakers are generally concerned with four levels: nominal, ordinal, interval, and ratio data.

- **Nominal** (categorical) scale - assignment of objects into categories based on data type or data qualitative characteristics (e.g., male or female, conifers or deciduous forest)
- **Ordinal** (ordered or ranked on a continuum) scale - assignment of objects into categories based on ranking where data only have relative values (e.g. small, medium, large; safe, iffy, dangerous)
- **Interval** scale - assignment of objects into categories based on data magnitude or numerical values in relative scale (e.g., temperature 32°C but also equals 0°F, or calendar dates: 1st, 2nd, 3rd. Thus, it is possible to measure the difference between things as well as rank them
- **Ratio** scale - assignment of objects into categories based on data magnitude or numerical values in absolute scale (with a true zero) (e.g., volume, population density, income, distance etc). This scale has best precision of all four.

Qualitative measurements are nominal, quantitative measurements are ordinal, interval or ratio.

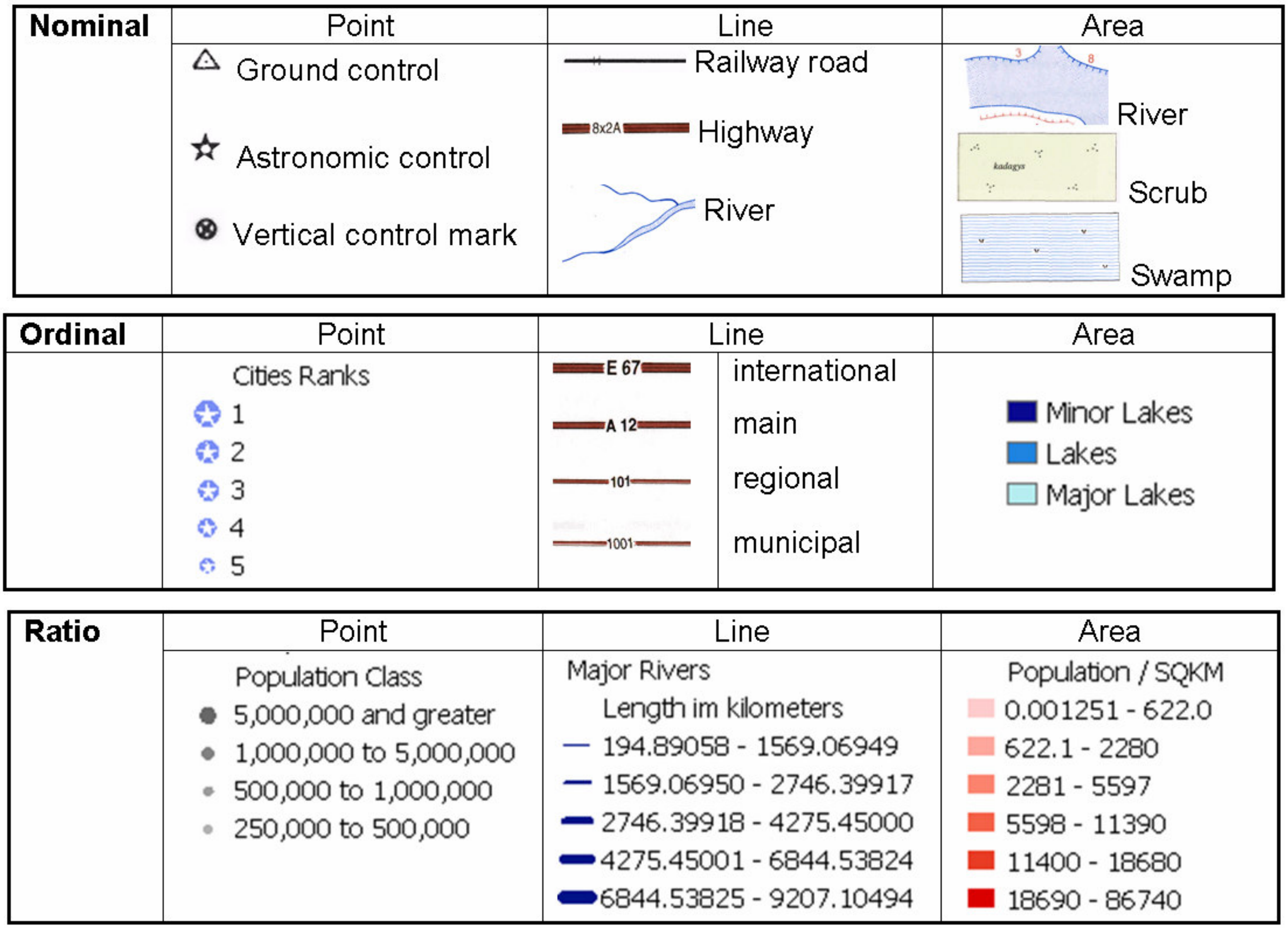


Figure: Differentiation of point, line and area features using symbols on nominal, ordinal and ratio scales of measurement

6.1.3 Visual Variables (Map Symbol Language)

Several characteristics of a symbol can be manipulated. These are often referred to as the visual variables. Visual characteristics of a symbol can be manipulated depending on the respective object's attributes. There are several graphic variables (Bertin's) that can be considered as map language elements or fundamental units when building a map image.

Conventional static visual variables include size, shape, orientation, pattern (arrangement, texture, and orientation), hue, value, and saturation. Visual variables introduced with GIS and digital cartography include transparency along with dynamic visual variables, such as moment, duration, frequency, order, rate of change, and synchronization.

The form of a graphic mark provides shape. Size describes apparent geometric dimensions - length, height, area, volume. A direction of lines and elongated shapes define a symbol orientation. Color has three dimensions: hue, tone (or value) and saturation (or chroma).Hue of symbol color is defined by length of reflected light. Value or tone of color refers to the relative lightness or darkness of a mark in any hue. Chroma or saturation refers to the degree to which a hue departs in "colorfulness" from a gray tone of the same value or purity of hue.

A pattern is repetition of basic graphic elements (marks) representing various combinations of above primary visual variables. A pattern exhibits the characteristics of arrangement, texture (spacing), and orientation Arrangement refers to the shape and configuration of component marks that make up a pattern. Texture refers to the size and spacing of component marks that make up a pattern. Pattern orientation refers to the directional arrangement of parallel rows of marks as they are positioned with respect to some frame of reference.

In GIS, visual variables can be stored as that object's symbolization attribute in a database. The following matrix illustrates which main visual variables are useful to represent different types of geographic data.

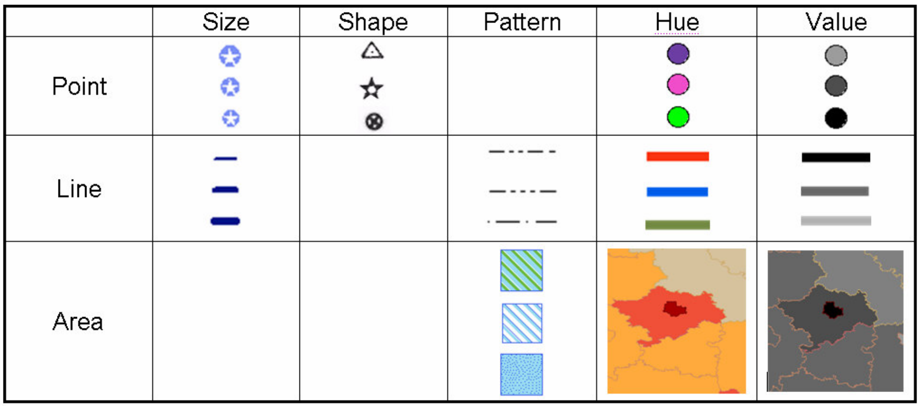


Figure: Samples of static visual or graphic variables

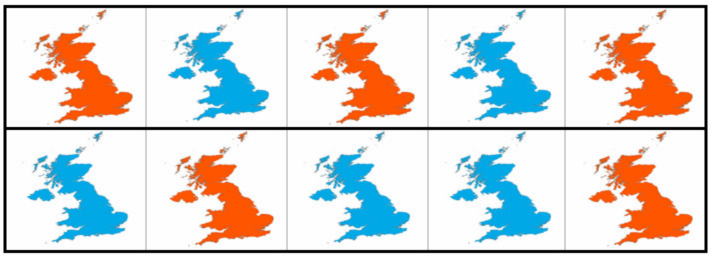


Figure: Samples of dynamic visual or graphic variable - frequency

In addition, not all variables apply equally well to different types of data. The following table summarizes which kinds of visual variable symbols should be used with each data type (measurement scale). When making thematic maps, a primary goal is to choose symbols that are intuitive to the map-reader and follow fundamental guidelines of map design. This theme will be addressed in subsequent modules.

Table: Kinds of visual variables that should be used with each data type

Visual Variable	Level of Measurement		
	Qualitative	Quantitative	
	Nominal	Ordinal	Interval/Ration
Size		x	x
Shape	x		

Orientation	x		
Texture (pattern)	x	x	x
Orientation (pattern)	x		
Hue (color)	x	x	
Value (color)		x	x

The following table summarizes which geographic and cartographic dimensions and scales of measurements are used in order to represent real-world features or phenomena on maps.

Table: Summary of Elements of Cartographic Representation

Geographic Entities and Phenomena		GIS Representation	Measurement Scales of Attribute Data	Map Symbols
Discrete Continues	point line area surfaces time	Vector point line area TIN	nominal ordinal interval ratio	point line area volume animation
		Raster integer floating		cell

6.2 Generalization

In cartography and geodesy, three processes of the transformation of the Earth's surface are recognized:

- Primary is semi-geometric – transformation of geoid to ellipsoid
- Secondary is geometric (map projections) – transformation of an ellipsoid (sphere) to a flat surface
- Tertiary is generalization

The first two processes have already been discussed. Here we will discuss the generalization process.

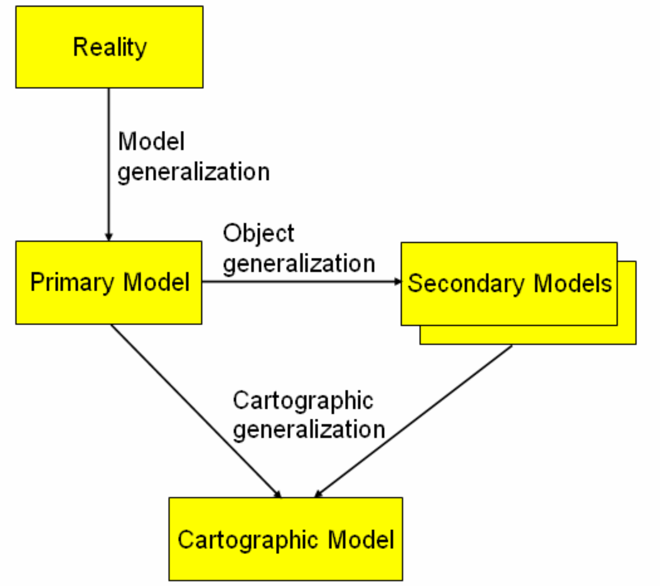
Cartographers have always had the task of attempting to visualize spatially-related phenomena. The challenge of handling scale appropriately is a problem that is aggravated by the fact that visualization necessarily takes place on a drawing medium of restricted size.

In general, the process of reducing the amount of spatial information and adjusting the data to a given purpose, theme and map scale is called generalization. This abstract transformation reduces the complexity of information from reality while maintaining spatial and attribute accuracy, along with significant characteristics of an area's topology and meeting user purposes and requirements. It encompasses a reduction of the complexity in a map (or database), emphasizing the essential while suppressing the unimportant, maintaining logical and unambiguous relations between map objects, and preserving aesthetic quality. Generalization is a non-reversible process and, therefore, must be carefully managed.

In geographical information systems and digital cartography, generalization functions are needed for a variety of purposes, including the creation and maintenance of spatial databases at multiple scales, cartographic visualization at variable scales, and data reduction.

Generalization occurs in three different forms along the path from the real-world phenomenon to its representation in a digital model or on a display medium, respectively. Three **steps** of the generalization process in digital environment are:

- **Model-based** generalization – geometrical transformations of real entities for purposes of their representations in digital databases
- **Object-based** generalization - geometric transformations of digital databases objects for the purpose of their representation in different scales
- **Cartographic** generalization – graphical transformations for visualization representation purposes and map production



•

Figure: Three steps of the generalization process

6.2.1 Driving Factors of Generalization

The main groups of factors or controls that drive the generalization process are:

- Phenomenon-based factors that take into consideration the nature of the modeled entity (e.g. the peculiarities of the area, relations among entities, etc.)
- Purpose-oriented factors (e.g. user needs and purposes, map contents, scale, etc.)
- Graphic media and format factors (e.g. visualization purpose, technology of map compilation, types of cartographic objects, methods of cartographic representation, rules for map design, etc.)
- Computational factors (e.g. efficiency of used information system)

It is necessary that the first factor be taken into consideration as the first model-based step in the generalization process. Factors 2 and 4 are accounted for during the second object-based stage of the generalization process. Factors 2 and 3 drive cartographic generalization. However, factor 1 has the highest priority when resolving cartographic generalization conflicts.

The following knowledge is required to accomplish the generalization process:

- Structural knowledge - spatial and semantic structure and topology of objects and phenomena
- Geometrical knowledge - graphical limits of representations
- Procedural knowledge - how generalization operations are selected and its running sequence
- Cartographic knowledge – relation to map design (e.g. aesthetics and visual balance)
- Application requirements - map theme, purpose, and intended audience

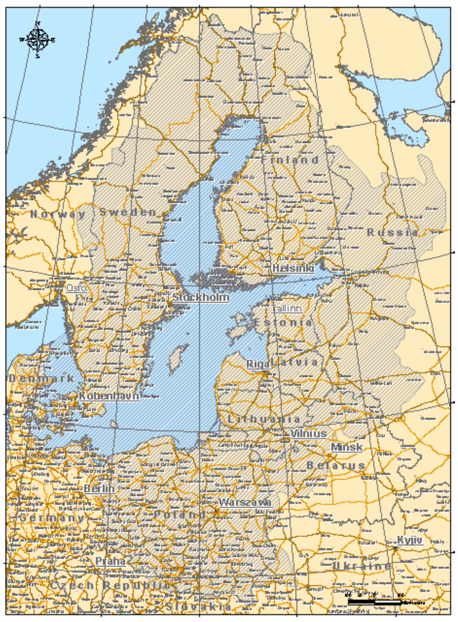
6.2.2 Generalization Techniques

The following operations, methods or techniques can be used in the generalization process:

- Elimination of an entire entity or part of its geometry:
    - **Selection**
    - **Simplification**
    - **Collapse**
- Merging of entities or objects:
    - **Aggregation,**
    - **Amalgamation**
- **Typification** of a feature or object's pattern
- Display operations:
    - **Exaggeration**
    - **Displacement**
    - **Smoothing**
    - **Classification**

These generalization techniques can be implemented by vector or/and raster algorithms or a combination of these algorithms within an automated environment, such as GIS or computer-assisted mapping software. There is software designed specifically for spatial and cartographic map generalization (e.g. Radius Clarity of Inc., formerly Laser-Scan, http://www.1spatial.com/products/radius_clarity/).

Selection operators select the most important (due to some criteria) features or/and objects and eliminate the not-so-important whole features or objects. These operations could be implemented by spatial SQL queries and more complex algorithms.



Capitals, large, medium and small cities, European Highway system and major country roads

Capitals, large, medium and small cities, European Highway system

Capitals and large cities only, no roads.

Figure: Examples of selection and elimination processes

Most often used criteria or measures for selection usually based on:

- Qualitative, quantitative and spatial characteristics of objects or features such as area, length, width, height, etc.:
  - For example, lakes will be shown on a map for an area more than 2 sq. mm at map scale
  - Population places will be shown on a map with a place's population more than 5,000 people
  - Only paved roads will be shown on a map
- Object's importance or rank:
  - For example, only administrative centers will be shown on a map, and other populated places will be omitted from a map image
  - Only water wells in desert areas will be shown
  - Cost functions based on a few characteristics of importance
- Object's density can be used to control how many objects can be shown per sq. unit of map area:
  - For example, only less than 10 geodetic control points have to be shown per 10 sq. cm at a map scale

The Topfer's radical law $\mathbf{n_c = n_s^* \, sqrt(S_c/S_s)}$ can be used to quantify level of generalization between maps of two scales, where **n** is number of objects on **c** and **s** maps with respective scale fractions of $\mathbf{S_c}$ and $\mathbf{S_s}$**.** This law only measures level of objects or an object's elements (e.g. zigzags of a river) reductions. It does not guide what objects or elements need to be omitted during the selection or simplification processes.

Simplification operators selectively reduce the number of lines or outline vertices required to represent a feature or object. Lines or outlines of objects are simplified. There are many algorithms for simplification and they usually take into consideration the specifics of distinct feature classes, (e.g. specific algorithms can be applied to simplify coastal lines, roads etc.). Usually vector-based analysis algorithms are used for simplification.

Simplification algorithms can work as:

- Independent algorithms (e.g. every *n*th vertex will be eliminated)
- Local algorithms that take into consideration positions of neighboring vertices (e.g. distance or angular measurements to immediate neighbors, such as the Jenks algorithm)
- Global algorithms that process entire outlines (e.g. Douglas-Peucker)

Thus, different constraints and measurements can be used by employing different algorithms.

For example, Jenks algorithm works in the following way: (1) If the distance from *point 1* to *point 2* is less than *min 1*, or (2) the distance from *point 1* to *point 3* is less than *min 2*, *point 2* is omitted. If both are larger, the angular check is calculated using *ang*. If an angle is smaller than *ang* this will result in the removal of *point 2*.

The mostly widely known global algorithm is Douglas-Peucker. This algorithm starts by connecting the two end-points of the original line with a straight line (termed the base line or anchor line). If the perpendicular distances of all intermediate vertices are within the threshold ε from the baseline, these vertices may be eliminated and the original line can be represented by the base line. If any of the intermediate vertices falls outside ε, however, the line is split into two parts at the furthest vertex and the process is repeated recursively on the two parts.

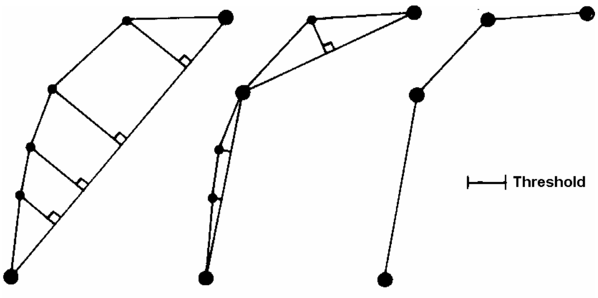


Figure: Douglas-Peucker global algorithm

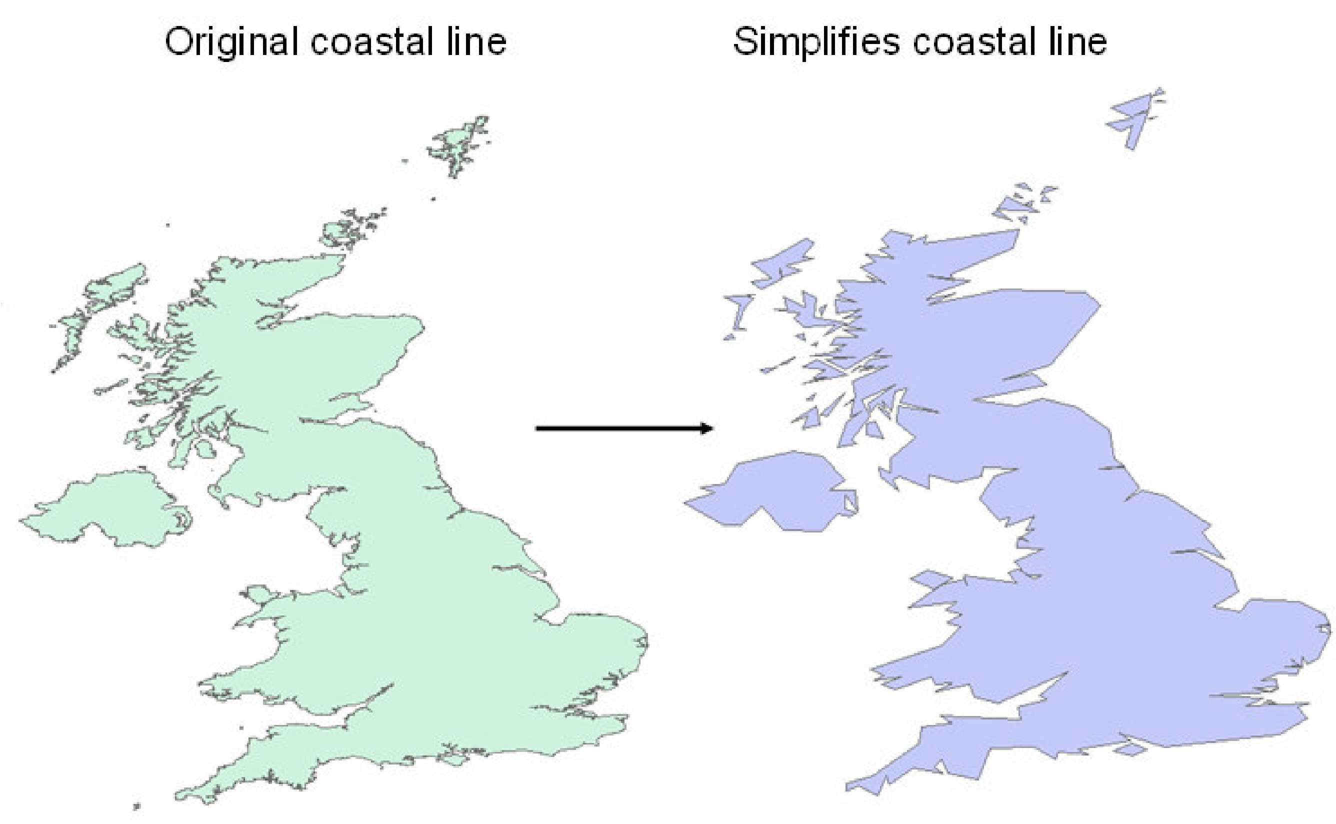


Figure: Simplification operation

Collapse operators replace feature or object physical details with abstract geometrical objects or details of low dimension. An object's geometry is completely changed. For example,, collapse reduces line or area features to point objects, or area features to line objects. Collapse operations are used when some objects or their details have to be preserved for some reason, such as importance or preservation of a typical feature's structure.

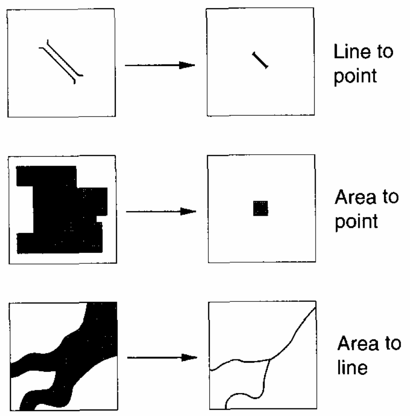


Figure: Collapse operations

Aggregation operators group several homogeneous objects or features into one object on the same level of the object's hierarchy. For example, aggregation operators can group areal objects that represent individual gardens into one large orchard or group several point objects that represent trees into one linear object that represents a hedgerow.

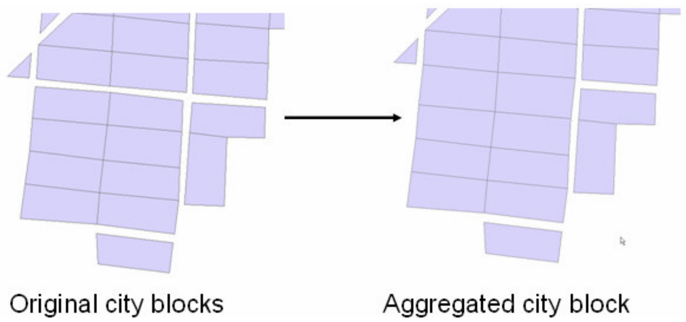


Figure: Aggregation operation

Amalgamation operators group several homogeneous objects into one object on the higher level of the object's hierarchy. For example, an amalgamation operation groups a few point objects that represent buildings into one large city block or groups two lanes of road into one highway linear object. Amalgamation and aggregation operations are used when distance between adjacent objects is too narrow to represent graphically on a map and no map space is available to displace these objects.

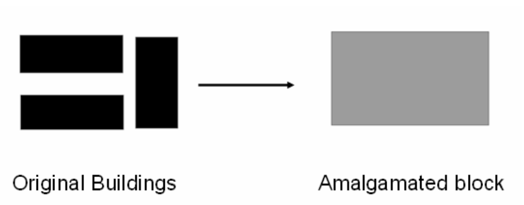


Figure: Amalgamation operation

Exaggeration operators amplify the whole object or a specific portion by not preserving its size. This is done to highlight distinguishable shapes or details of an object. This operation is used to enlarge objects equally in all directions. The resultant shape is preserved but it will be scaled by some magnitude. Enlargement is mainly applied to objects that reach a minimum size, but still have to be shown on a map due to it importance. Exaggeration is used to enlarge parts of objects, either because they do not satisfy the geometric constraints, or because such parts are of special interest.

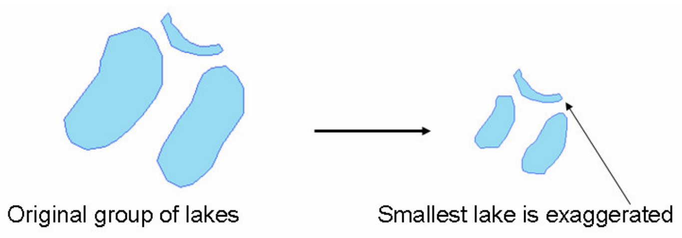


Figure: Exaggeration operation

Displacement operators employ a shifting of object(s) from original position(s) in order to resolve topological conflicts of more than one object overlapping. This operation might be important to solve conflicts between objects that are too close or to preserve important neighborhood relations.

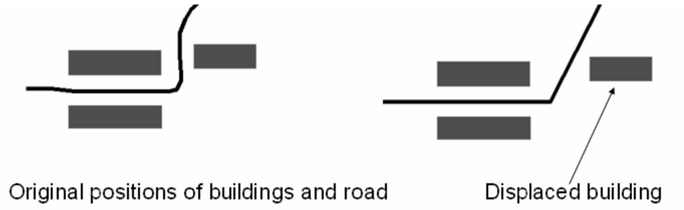


Figure: Displacement operation

Smoothing allows an aesthetic refinement by reducing the angularity of angles between segments of linear objects. These operations reduce sharp angularity from objects having smooth shapes. This is a completely aesthetic process of cartographic generalization.

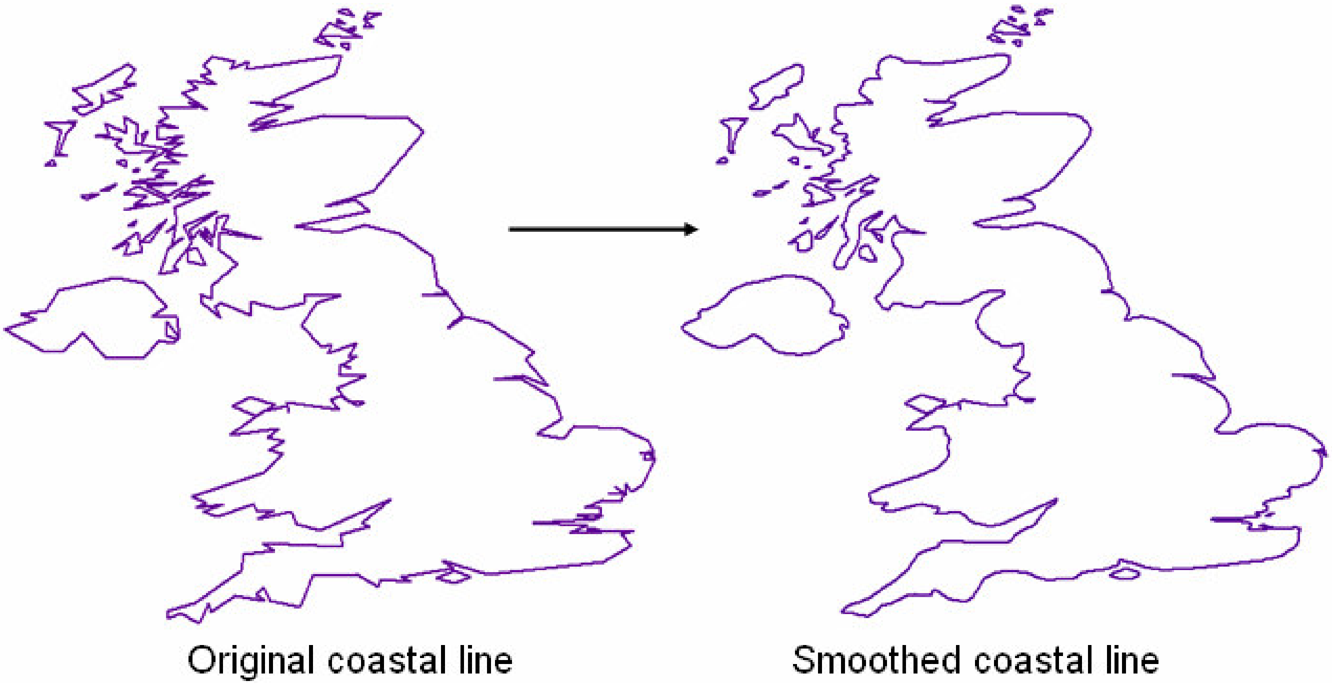


Figure: Smoothing operation

Typification operators often generalize a set of objects to a new reduced group of objects that has to show similar characteristics of pattern and arrangement (e.g., density and orientation). These techniques can be implemented by combining several basic algorithms, such as selection, aggregation and simplification into one algorithm.

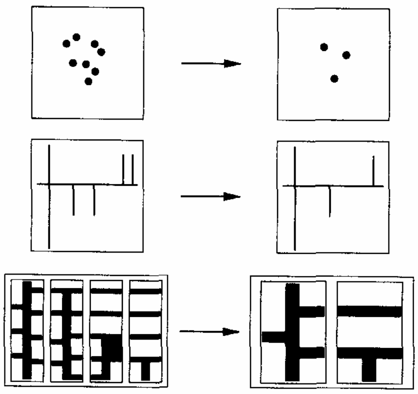


Figure: Typification operations

Classification operators group features, objects or symbols into new groups based on their quantitative or qualitative attributes. These generalization techniques are based mainly on attribute values, not geometry or position. Attribute transformations, including reclassification and thematic simplification, can be performed. For example, conifers or deciduous forest types are simplified into one forest land cover type. Alternatively, populated places can be reclassified based on population size or administrative functions. Objects that have adjacent boundaries can be merged into one 'polygon' after reclassification.

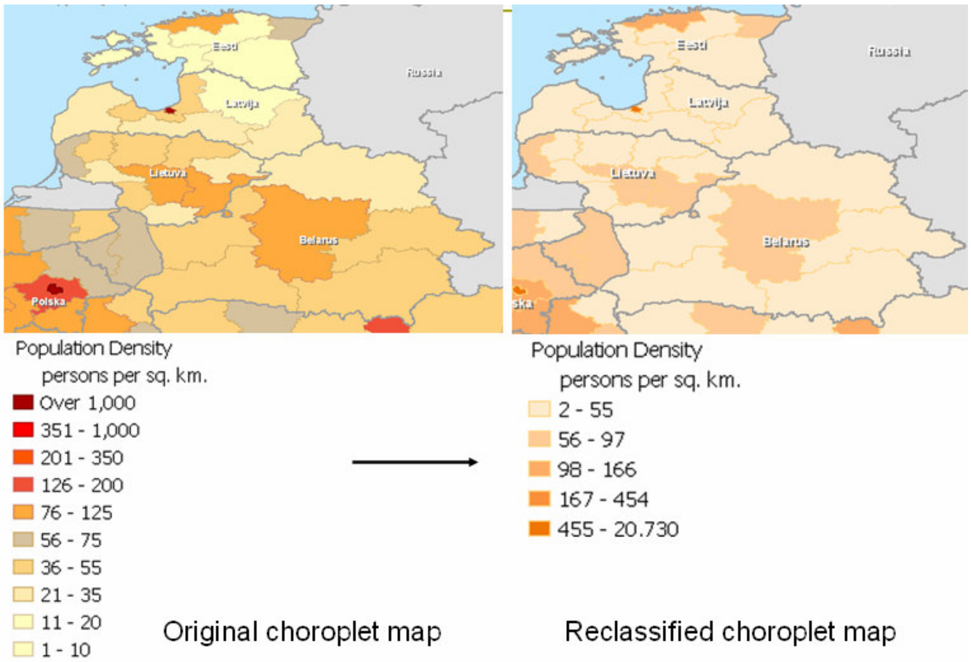


Figure: Classification operations

Symbolization is a process of applying symbology to vector or raster graphic objects. It may also require generalization. For example, the size of symbols must be chosen to satisfy map design principles or scaled properly for graduated symbols to be chosen appropriately to convey map clarity and readability.

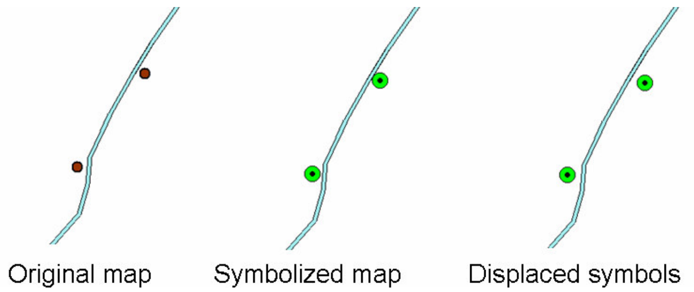


Figure: Symbolization generalization

6.2.3 Generalization Algorithms

Generalization algorithms are formal mathematical constructions that solve a generalization problem by changing an object's geometry or attribute. The role of algorithms is to provide procedures that allow for the database or visual re-organization of information in such a way as to communicate the aspects of space that are salient to a map's scale and specific theme.

Algorithms of generalization techniques can work locally, within a neighborhood and/or globally. Local algorithms process objects one by one on an individual basis, as they are loaded at the beginning of the process.

Neighborhood operations work with groups of objects. These groups are created during the generalization process to allow contextual operations and to maintain some group properties.

Neighborhood operations comprise a group of local objects or a group of neighborhood objects. Local objects generalize themselves when they are autonomous or accomplish orders given by neighborhood operations for contextual operations (e.g. road and building conflicts are when several generalization operators can be applied but topological relations between objects has to be preserved).

For example, a smoothing operator transforms only individual objects. However, during simplification operations, one has to take into consideration the locations of neighbors to avoid topological conflicts (e.g. overlapping). Displacement operators have to be able to work within a neighborhood in order to solve topological conflicts and preserve local patterns. Selection/elimination is a global technique that can be based on global object patterns (e.g. Topfer's radical law).

Generalization algorithms use **measures** that are procedures for computing measurements, which are used as a basis for evaluating the characteristics of a geographical entity and assessing the need for, and the success of, generalization. A measure might consist of a simple formula or it may involve a complex algorithm, including computation of auxiliary data structures or representations. For example, the position, orientation, shape, distances, alignments, and characteristics of topology can assist in choosing generalization algorithms and their parameters and/or they can be used as an algorithm's parameters.

Another component of generalization algorithms is **constrains**. They reduce the number of possible results of a generalization process, while at the same time increasing the proportion of acceptable ones. The two basic constraints are:

- Geographic and cartographic - arise from characteristics of data and map specifications
- Process – arise from resource limitations and workflows.

Some examples of constraints include minimum length or width of objects, separability between objects and rules such as avoid self-intersection.

6.2.4 Implementations

Spatial and cartographic map generalizations can be performed:

- Manually by using computer-assisted on-screen drawing tools
- With the help of computer-assisted generalization tools in two modes:
  - Batch mode - when one algorithm can be run to process a set of similar objects (e.g. to simplify river after river). A cartographer or GIS technician can choose a particular algorithm and its parameters, but can see and estimate results only after the algorithm has been completed
  - Interactive generalization - when the computer assists an operator in real time. An operator can see a result of individual operations and have some control on decision-making. This process is still time consuming
- Automated generalization - performed in any of the following ways or in combination:
  - On-the-fly generalization - when the user requests a map in an appropriate scale and for a particular purpose, and the system returns a result after real-time processing of data from a single scale database

- Multi-scale or multi-detail database – map may be pre-built at a requested scale and then can be retrieved in real-time without heavy processing
- Combination of multi-scale and on-the-fly generalization approaches can be used
- Use of object-oriented and agent generalization systems - design principles to implement fully automated generalization

6.3 Topographic Mapping

The topographic map is one of the most widely used maps. Topographic maps usually portray both natural and cultural features. The wide range of information provided by topographic maps make them extremely useful to professional and laypersons alike. Topographic maps are used for engineering, energy exploration, natural resource conservation, environmental management, public works design, commercial and residential planning, and by outdoor recreationalists when hiking, camping, and fishing.

Topographic cartographers produce maps, charts, and plans at differing scales and they are classified into map series. Topographic maps generally show real-world physical and cultural objects whose locations have been accurately determined by surveyors. These objects are generally visible in the real world, except for such things as government administrative boundaries,. Elevation, buildings, roads, settlements, hydrography (creeks, rivers, and lakes), and general land uses are represented on topographic map sheets.

Topographic maps are one of the most important tools when studying landforms and the processes that shape them. These maps are two-dimensional, scaled representations of the Earth's surface. These maps show the size, configuration, and spatial relationship of landforms and cultural features in some detail. They allow very accurate measurement of elevations and horizontal distances.

The important characteristic of topographic maps is the emphasis placed on positional accuracy of real-world features that is repeatable in a quantitative way, through coordinates, grids, and markers. Topographic maps belong to the class of general reference maps.

Topographic maps are classified on the basis of scale and are identified as belonging to distinct map series. Each country has its national system of topographic map series. The professional cartographer usually identifies three general scales for maps:

- **Large-scale topographic plans** (scale <= 1:2000); projections not applied; found in neighborhood plans and cadastres
- **Standard base** or **topographic maps (scale** 1:2,000 to 1:100,000)
- **General topographic maps** (scale 1:200,000 to 1:1,000,000); considered by some traditional cartographers to be outside of topographic representation because too much geographic information has to be generalized to make "topographic" representation almost meaningless.

6.3.1 Scale Series and Coordinate System

Topographic maps are created as a series – at differing scales designed to satisfy the needs and requirements of different users. The maps must be produced in such as way that the scales of one series will mesh with that of the smaller and larger scale maps.

Although slightly different map series are used in different countries, the following scales often comprise discrete series:

- 1 : 500 – Lithuanian municipality level topo-plane
- 1 : 1000 – Lithuanian municipality level topo-plane
- 1 : 2,000
- 1 : 5,000
- 1 : 10,000
- 1 : 25,000
- 1 : 50,000
- 1 : 100,000
- 1 : 200,000
- 1 : 250,000
- 1 : 500,000
- 1 : 1,000,000

Topographic map sheets are classified and indexed in a predefined hierarchical system starting from primary 1:1,000,000 map sheets quadrangles. Such **national topographic** divisions and indexing **systems** (or grids) are very convenient when searching and/or storing cartographic information.

Primary quadrangles of 1:1,000,000 series can be created by the following rules:

- A map of entire Earth can be divided on the basis of quadrangles 4° wide along the meridians and 6° wide along the parallels
- Division along the parallels can coincide with 6° division of UTM zones
- When above φ=80°N and below φ=80°S each sheet can cover 4° of φ and 12° of λ

A **National Topographic System (NTS)** provides general-purpose topographic map coverage for a nation. Primary quadrangles of 1:1,000,000 series can be indexed created by the following rules:

- Each map of 1:1,000,000 series is indexed with two indexes – meridian zone number (can be the same as UTM zone number) and English character (in upper case) for parallel zone number
- Numeration of parallel zone can start from the equator in alphabetical order and may not coincide with UTM numeration
- For example, in Lithuania, a topographic map with a scale of 1:1,000,000 can be **N-34**, where **N** is 4° zone number of φ, and **34** is 6° zone number of λ in accordance to UTM numeration

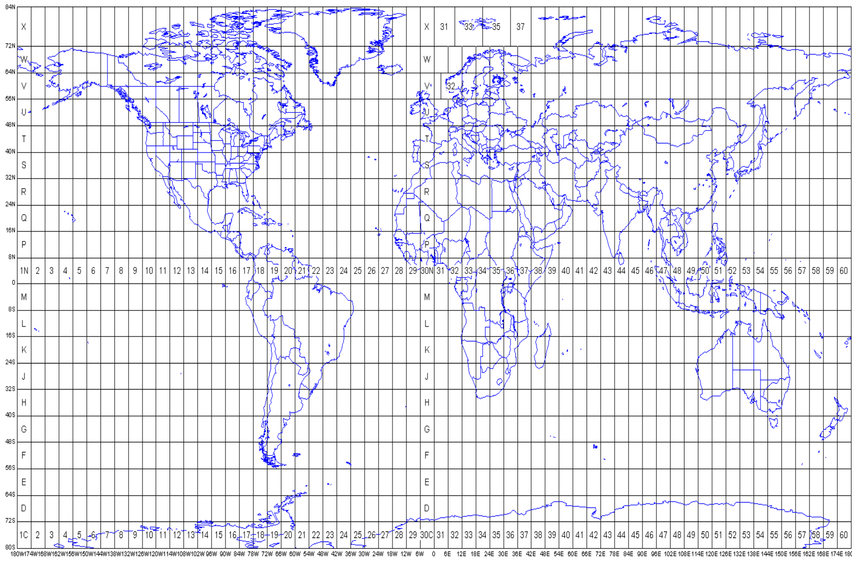


Figure: Global indexing system of UTM zones

The first subdivision of the primary quadrangles is to quarter them to produce maps at a scale of 1:500,000. Each 1:500,000 map sheet can cover 3° of λ and 2° of φ. Each 1:500,000 map sheet within a 1:1,000,000 quadrangle can be indexed as A, B, C or D (e.g. **N-34-B).**

Each 1:1,000,000 quadrangle can be subdivide into 36 quadrangles with a scale of 1:200,000. Each 1:200,000 map sheet can cover 1° of λ and 40' of φ. Each 1:200,000 map sheet within a 1:1,000,000 quadrangle can be indexed hierarchically by Roman numbers sequentially from I to XXXVI (e.g. **N-34-VIII).**

The next subdivision is the 100,000 map scale series. This series is produced from the 1:1,000,000 quadrangles by subdividing it into 144 quadrangles. Each map sheet at 1:100,000 scale can cover 30' of λ and 20' of φ. Each 1:100,000 map sheet within 1:1,000,000 quadrangle can be indexed hierarchically by numbers sequentially from 1 to 144 (e.g. **N-34-48).**

The next subdivision is to quarter the 1:100,000 quadrangles to produce the 1:50,000 series. Each sheet with a 1:50,000 map scale can cover 15' of λ and 10' of φ. Each 1:50,000 map sheet within 1:100,000 quadrangle can be indexed hierarchically by characters (in upper case) sequentially from A to C. (e.g. **N-34-48-C),** one 1:50,000 map sheet can contain joint quadrangles (e.g. **N-34-48-CD).**

To produce the 1:25,000 series each of the 1:50,000 subdivisions of the lettered quadrangles are divided into 4 parts. Each sheet with a 1:25,000 map scale can cover 7'30'' of λ and 5' of φ. Each 1:250,000 map sheet within a 1:50,000 quadrangle can be indexed hierarchically by characters (in lower case) sequentially from *a* to *c* (e.g. **N-34-48-C-a).**

The next subdivision is to quarter the 1:25,000 quadrangles to produce the 1:10,000 series. Each map sheet with a scale of 1:10,000 can cover 3'45'' of λ and 2'30'' of φ. Each 1:10,000 map sheet within the 1:250,000 quadrangle can be indexed hierarchically by numbers sequentially from 1 to 4, (e.g. **N-34-48-C-a-2).**

Map sheets in the 1:10,000 series can be subdivided into 4 quadrangles to produce 1:5,000 map sheet scales. Each sheet of a 1:5,000 map can cover 1'52.5'' of λ and 1'15'' of φ. Each 1:5,000 map sheet within a 1:10,000 quadrangle can be indexed hierarchically by numbers sequentially from 1 to 4 (e.g. e.g. 42/56-4**).**

Map sheets in the 1:10,000 series contain 25 map sheets at a scale of 1:2,000. Each sheet of 1:2,000 scale map covers 37.5'' of λ and 25'' of φ. Each 1:2,000 map sheet within 1:10,000 quadrangle can be indexed hierarchically by numbers sequentially from 1 to 25 (e.g. 42/56-21**).** See "Geodetic engineering investigations for constructions" - GKTR 2.08.01, 2000.

Typical national grids, at several scale series, are produced below.

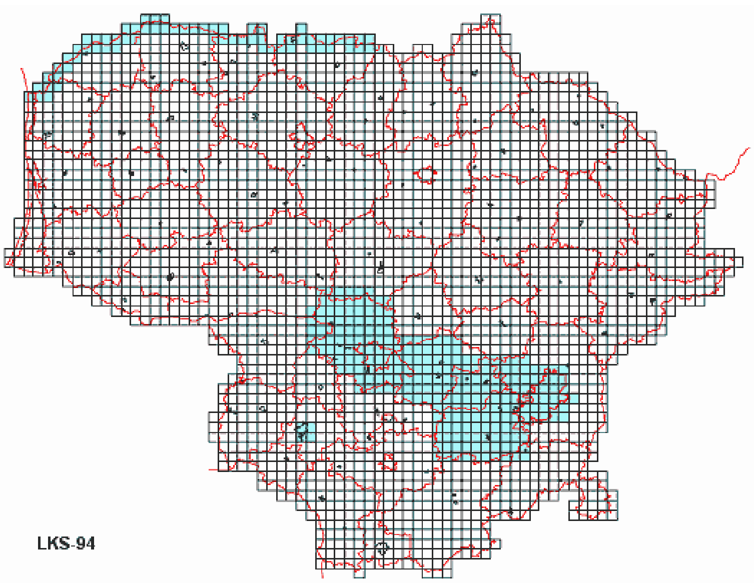


Figure: National Lithuanian topographic map grid for 1:10,000 map scale series in LKS-94 coordinate system

Figure: National Lithuanian TOP50LT topographic map grid for 1:50,000 map scale series in LKS-94 coordinate system

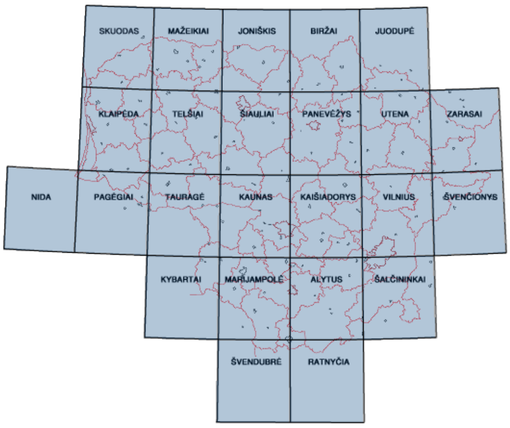


Figure: National Lithuanian topographic map grid for 1:200,000 map scale series (in KS-1942 coordinate system)

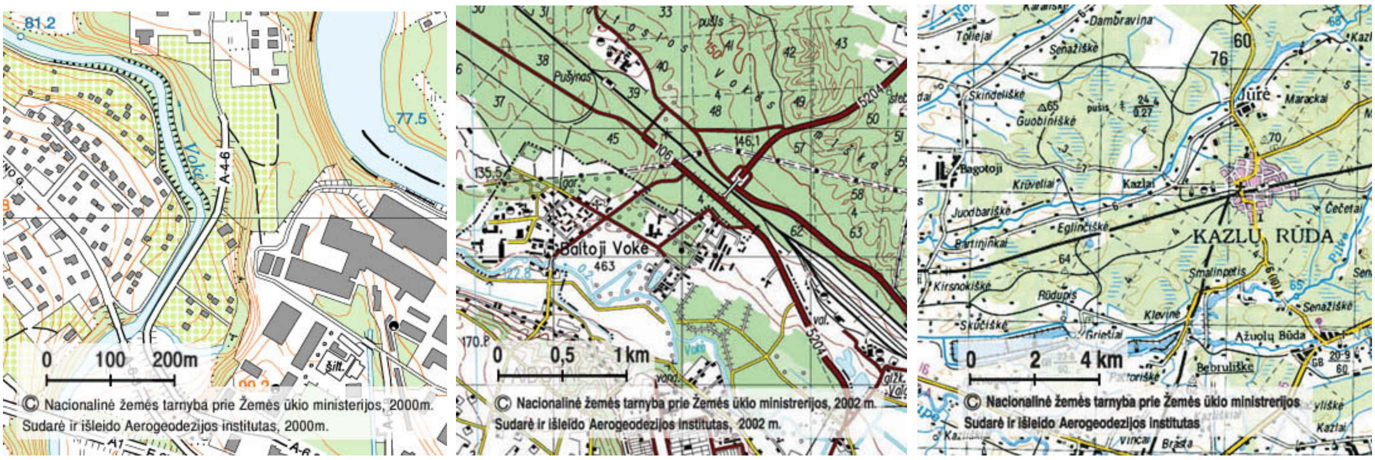


Figure: Samples of Lithuanian topographic map sheet of 1:10,000, 1:50,000 and 1:200,000 scale series

Topographic maps are produced using specific geographic and projected coordinate systems. Usually, the UTM coordinate systems are used. The following parameters can vary for national UTM grids of different countries: scales of central meridians of zones, systems of zone delimitations and widths of zones, False Northings and False Easting.

Topographic maps can show control geodetic points, ticks and labels of geographic graticule, measured grid lines and labels (e.g. kilometrical UTM grid), neatlines, national topographic system index and subdivisions, all these elements help to work with the mathematical base of a map and assist in making measurements.

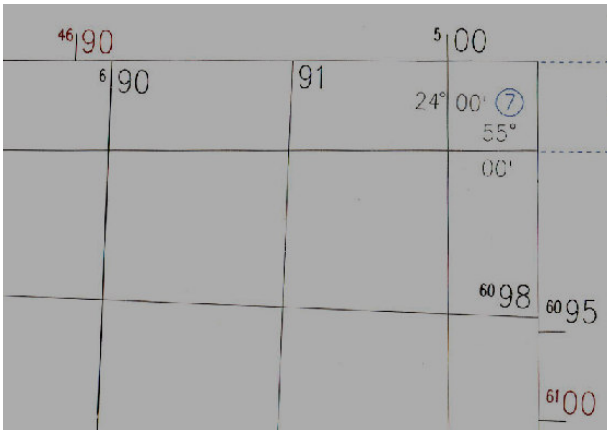


Figure: Ticks and labels of geographic graticule, measured grid lines and labels (kilometrical UTM grid)

The parameters of the Lithuanian Topographic Coordinate System used for Lithuania's new topographic maps:

- LKS-1994 Lithuanian TM:
  - Projection: Transverse Mercator
  - False Easting: 500000,000000
  - False Northing: 0,000000
  - Central Meridian: 24,000000
  - Scale Factor in the Central Meridian: μ = 0,9998
  - Latitude of Origin: 0,0000 (equator)
  - Linear Unit: Meter
- Geographic Coordinate System: GCS LKS-1994
  - Angular Unit: Degree (0,017453292519943299)
  - Prime Meridian: Greenwich (0,0)
  - Datum: D_Lithuania_1994
  - Spheroid: GRS_1980
  - Semimajor Axis: 6378137,00
  - Semiminor Axis: 6356752,31414035610
  - Inverse Flattening: 298,2572221010000200
  - Vertical datum:
    - Baltic height system

This system does not coincide with standard UTM zones.

6.3.2 Topographic Map Specifications and Elements

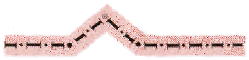
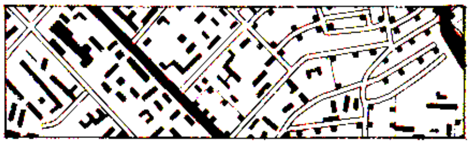
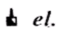
Topographic cartography is clearly defined by the following specifications:

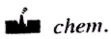
- A list of the elements presented on a topographic map
- Generalization rules for those elements
- Symbology, text labels and annotation design rules (e.g. color schemas, font styles and sizes, etc.)
- Layout elements (e.g., size, style, position, margins, etc)

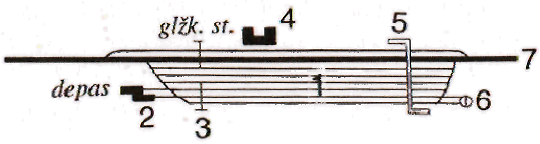
For example, the “Specifications for Lithuanian 1:50 000 scale topographic maps” has to be used for compilation, design and production of topographic maps of that series.

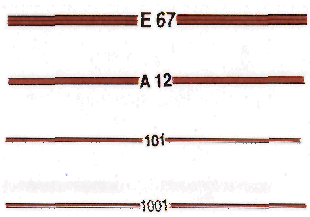
A detailed list of the elements presented on a topographic map depends on map scale and specifications. For example, Lithuanian 1:50,000 scale topographic maps show the following groups of map elements:

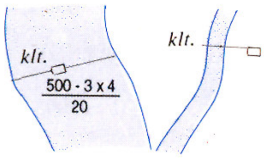
- Ground Control Points 

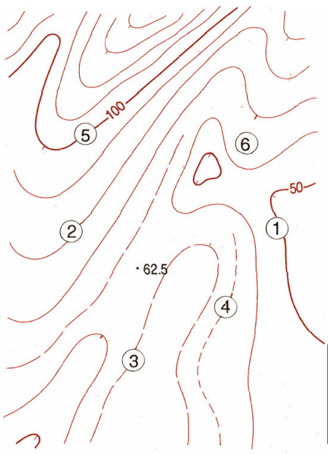
- Boundaries 
- Populated Places, Buildings and Construction  

- Cultural Objects 

- Railroads and Related Data 

- Motor Roads, Gravel Roads, Paths 

- Drainage and Water Conduits 

- Relief 

Metric Contour Interval Guidelines

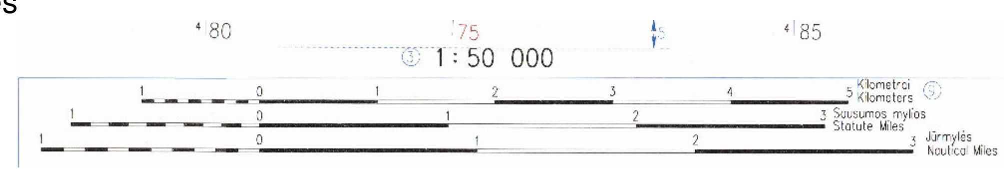
MAP SCALE	TERRAIN TYPES		
1/50 000	Flat or normal	Hilly	Steep
	10 metres	20 metres	40 metres

The most common way of indicating relief on a topographic map is by using a *contour line*. A contour line is a line on a topographic map that connects points of equal elevation. *Contours* are drawn by interpolating between points whose positions and elevations have been measured and plotted. The vertical distance between level surfaces forming the contours is called the *contour interval*. The contour interval selection depends on the diversity of relief in the area being mapped, as well as the purpose and scale of the map. Every fifth or tenth contour may be drawn a little darker for easier reading. In addition, intermediate or half- or quarter-interval contours can be drawn to depict variation for a highly planar surface.

The main characteristics of contours:

- Lines must close on themselves, either on or off a map. They cannot dead-end
- Lie perpendicular to the direction of maximum slope
- Slopes between adjacent contour lines are assumed to be uniform - therefore it is necessary that breaks in grade be located in topographic surveys
- Distance between contours indicates the steepness of a slope. Wide separation denotes gentle slopes; close separation, steep slopes; even and parallel spacing, uniform slope
- Rough and rugged terrain represented by irregular contour lines. Smooth lines imply gradual slopes and changes
- Hills represented by concentrically closed contours as elevation increases. A contour forming a closed loop around lower ground is called a *depression contour*
- Steep cliffs, walls and natural bridges may have touching contour lines; theoretically, contour lines are not allowed to touch one another.
- A contour must be a single continuous line and should not branch into two contours of the same elevation.
- Contour lines point upstream and form V's when crossing a stream; they point down the ridge and form U's when crossing a ridge crest

- Vegetation and Soil
- Map layout's elements:
  - Scales



CONTOUR INTERVAL 10m

  - Contour interval

- Coordinate system

  Ellipsoid: WGS 84
  Projection: Universal Transverse Mercator
  Horizontal Datum: World Geodetic System 1984
  Vertical Datum: Baltic

- Map index for National grid

  Serija N 7510 (3) N-34-48-AB
  Leidimas 1-VŽGT

- Name of a map sheet

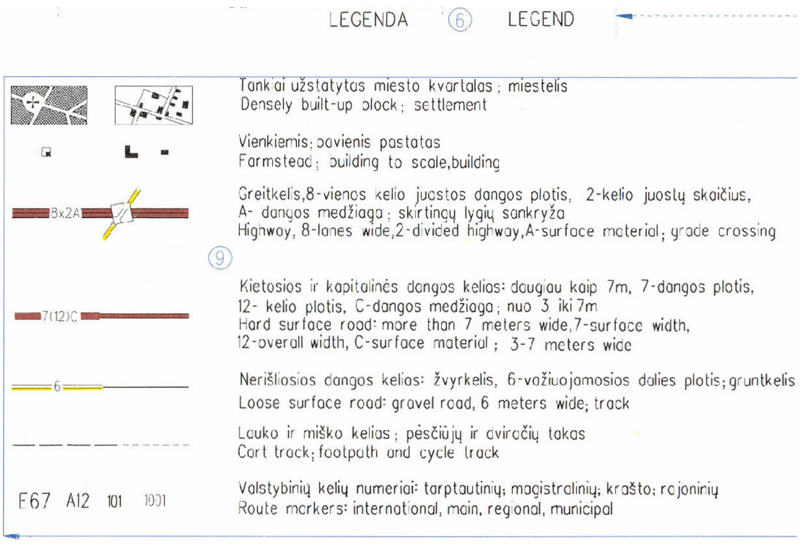
  KAUNAS

- Adjoining sheet positions

  (5) GRETIMI LAPAI
  ADJOINING SHEETS

N-34-35-CD	N-34-36-CD	N-35-25-CD
N-37-47-AB	N-34-48-AB	N-35-37-AB
N-34-47-CD	N-34-48-CD	N-35-37-CD

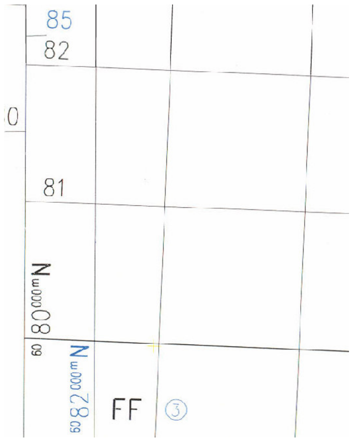
- Legend



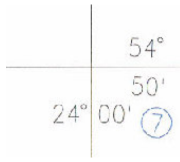
- Measured coordinate grids for different systems

  Grid: 1000 Meter UTM Zone 34 (inside sheet)
  5000 Meter UTM Zone 35 (outside frame)
  5000 Meter LCS 94 (outside frame)
  5000 Meter System 1942 (outside frame)

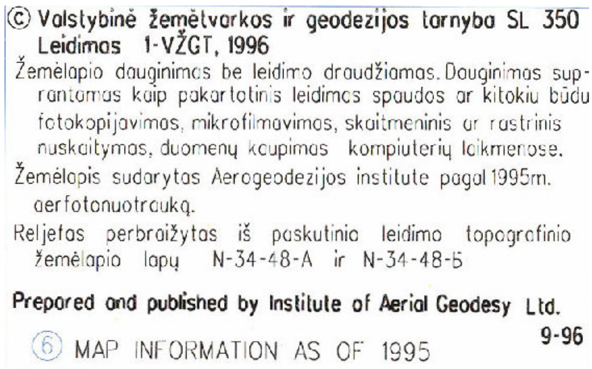
▪ Metric UTM grid



▪ Geographic coordinates for quadrant corners and tics



▪ Ancillary data and information


© Valstybinė žemėtvarkos ir geodezijos tarnyba SL 350
Leidimas 1-VŽGT, 1996
Žemėlapio dauginimas be leidimo draudžiamas. Dauginimas suprantamas kaip pakartotinis leidimas spaudos ar kitokiu būdu fotokopijavimas, mikrofilmavimas, skaitmeninis ar rastrinis nuskaitymas, duomenų kaupimas kompiuterių laikmenose.
Žemėlapis sudarytas Aerogeodezijos institute pagal 1995m. aerfotonuotrauką.
Reljefas perbraižytas iš paskutinio leidimo topografinio žemėlapio lapų N-34-48-A ir N-34-48-B
Prepared and published by Institute of Aerial Geodesy Ltd.
9-96
6 MAP INFORMATION AS OF 1995

6.3.3 Symbology

Standard symbols are used to represent special topographic features, thereby making it possible to show many details on a single sheet. A few of the many symbols employed in topographic mapping are shown below.

Topographic map specification clearly defines symbology including the:

- Symbol's size, shape, color, orientation, pattern
- Classification of features

AUTOMOBILIŲ KELIAI, GRUNTKELIAI, TAKAI **Motor Roads, Gravel Roads, Paths** Greitkelis (8-vienos važiuojamosios dalies plotis metrais, 2-juostų skaičius, A-dangos medžiaga) [1] Highway (8-lanes wide, 2-divided highway, A-surface material)	8x2A	Išorinė linija 0.2 Vidinė linija 0.1 Bendras plotis 1.4	Helvetica Narrow Normal 6 p.
Kietosios ir kapitalinės dangos keliai: [1] Hard surface roads: 1) dangos plotis didesnis kaip 7m (8-dangos plotis, 12-kelio plotis metrais, C-dangos medžiaga) surface width more than 7m (8-surface width, 12-owerall width in meters, C-surface material) 2) dangos plotis nuo 3 iki 7m (5-dangos plotis, 8-kelio plotis, G-dangos medžiaga) surface width 3-7m (5-surface width, 8-owerall width in meters, G-surface material)	8(12)C 5(8)G	Linija 0.15 Bendras plotis 1.0 Tarpas 0.7 Linija 0.1 Bendras plotis 0.6 Tarpas 0.4	Helvetca Narrow Normal 5 p. Helvetica Narrow Normal 5 p.
Nerišliosios dangos kelias: Loose surface road: žvyrkelis (7-važiuojamosios dalies plotis metrais) [2] gravel road (7-surface width in meters)	7	Linija 0.1 Bendras plotis 0.5 Tarpas 0.3	Helvetica Narrow Normal 5 p.
Gruntkelis [3] Track		Linija 0.2	
Lauko ir miško kelias [4] Cart track		Linija 0.15, atkarpų ilgis 3.0 Tarpas 0.6	

Figure: The road classification and symbology from "Regulation of Specifications for Lithuanian 1:50,000 scale topographic maps. First edition"

- Labeling and annotation style, size and font type

UŽRAŠŲ PAVYZDŽIAI **EXAMPLES**			Balt Helvetica
Miestų pavadinimai: [1] Names of towns:			
1) kai gyventojų daugiau kaip 500 000 first class	**VILNIUS** VILNIUS	19 p. 12 p.	Bold Normal
2) nuo 100 000 iki 500 000 second class	**ŠIAULIAI** ŠIAULIAI	16 p. 11 p.	Bold Normal
3) nuo 50 000 iki 100 000 third class	**MARIJAMPOLĖ** MARIJAMPOLĖ	14 p. 10 p.	Bold Normal
4) nuo 10 000 iki 50 000 fourth class	**UKMERGĖ** UKMERGĖ	12 p. 9 p.	Bold Normal
5) iki 10 000 fifth class	**MOLĖTAI** MOLĖTAI	10 p. 8 p.	Bold Normal

Figure: The label styles for population places from "Regulation of Specifications for Lithuanian 1:50,000 scale topographic maps. First edition"

- Colors and screens (patterns) for printouts

Juoda Black SPC 58600 *		tonas tone	Sutartinių ženklų kontūrai, charakteristikos, pavadinimai Symbols of features, dimensions, names
		20% 541 in./cm 21% 120 D*	Tankiai užstatytų miestų kvartalų tonas Tone of densely built-up blocks
Raudonai ruda Red-Brown SPC 61121*		tonas tone	Greitkelių ir kietosios dangos kelių juosta ir reljefas Band of highways and hard surface roads, relief
		20% 481 in./cm 21% 120 D *	Valstybės sienos juosta Band of international boundary
Geltona Yellow SPC 57377 *		tonas tone	Žvyrkelių juosta Band of loose surface roads
Mėlyna Blue SPC 47651 *		tonas tone	Hidrografija Drainage
		20% 541 in./cm 21% 120 D*	Vandens plotai Water area
		1:2 241in./cm LP-13 *	Valstybės saugomos teritorijos ribos juosta Band of protected area boundary
Žalia Green SPC 52813 *		40% 481 in./cm 42% 120 D *	Miškai ir sodų tonas Forests and tone of orchards
		20% 481 in./cm 21% 120 D*	Jaunuolynų ir krūmynų tonas Tone of nurseries and scurbs

Figure: Colors and screens for representation features on the topographic map from "Regulation of Specifications for Lithuanian 1:50,000 scale topographic maps. First edition"

External layout's elements such as scales, legend, neatline, text, border etc.

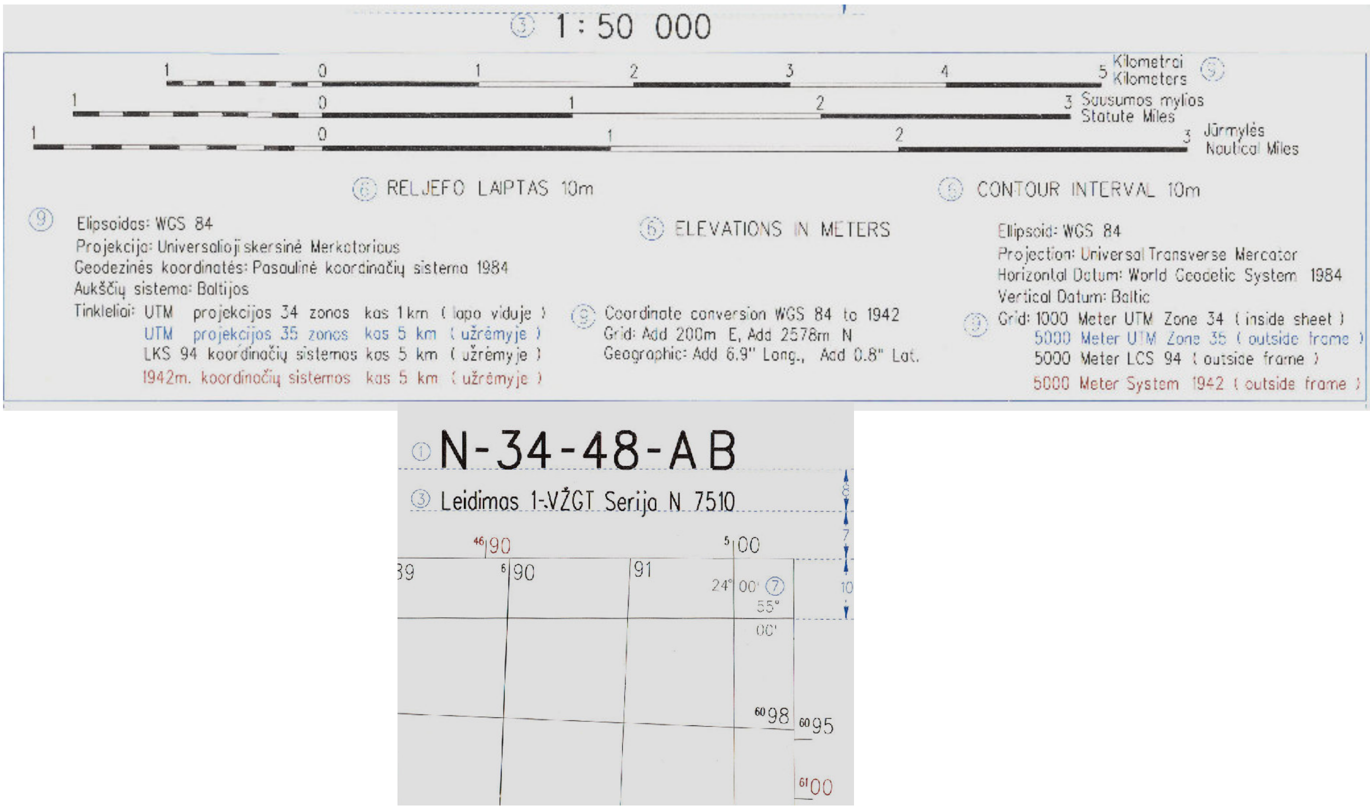


Figure: External map layout's elements

6.3.4 Generalization

Topographic cartography specifications or regulations also have defined generalization rules, constraints and measures:

- For example, lakes have to be shown from 2 sq. mm area in a map scale, rivers are shown from 1 cm length and longer
- Group buildings are joined into blocks if the distance between them is less then 0.2 mm
- Streets are preserved within populated places if the street are wider than 0.2 mm
- Contour lines show at 10 meter intervals
- Contour lines are drawn with a minimum line separation of 0.2 mm between contours
- Shows 10 points of geodetic control per 10 sq. cm on the map
- Shows only main zigzags of the costal line

The sequence in the application of generalization techniques needs to be defined. For example, the first has to be generalized elements of hydrography, followed by populated places, then roads, etc.

6.3.5 Technological Processes

Topographic map production involves a sequence of technological operations that are also defined by cartographic regulations. These regulations cover all aspects of topographic map creation.

Several procedures are involved in the creation of maps. These fall into three broad categories: planning, execution, and reproduction. Within these categories, the processes involve research, compilation, generalization, design and layout, symbolization, construction, editing, and reproduction. In the actual creation of a map, all these considerations are closely interrelated.

In general, technological processes involved in map production include the following groups of operations:

- Editorial-preparation work - includes the gathering of data (in the field, from statistical sources, from other maps, from imagery, and so on), materials and specifications, studying the mapped area, and drafting a plan that includes compilation steps and instructions for digitization, coding, generalization, editing, data cleaning, symbolization, etc.
- In the compilation and design stages, the cartographer or GIS technician "drafts" and provides symbolized sheets for the topographic map (that also can include generalization) in accordance with conventional sign specifications, mapping regulations and the editor instructions
- The next stage involves map editing and revisions by an editor
- The last stage involves reproduction. This stage includes a sub-stage of preparation for map printing and actual polygraphic map reproduction. Cartographers may not be involved in the actual processes of making printed maps, but because reproduction has a

bearing on all other stages in map making, they must be familiar with reproduction methods. With increased use of computers to produce maps, cartographers are more likely to be involved in all of these phases.

Similar stages can be used for updating existing topographic maps.

New technologies are altering the production and use of traditional maps. A few technological methods for topographical map creation today include:

- Large scale topographic maps can be created using ground or field topographic survey measurements
  - Several ground methods are used for topographic mapping (e.g. plane-table surveying and total stations etc)
  - The data obtained from these surveying techniques is used in the creation of a suitably scaled topographic map
  - Survey data can be plotted as points, lines and polygons and stored in a topographic database and later used for topographic mapping

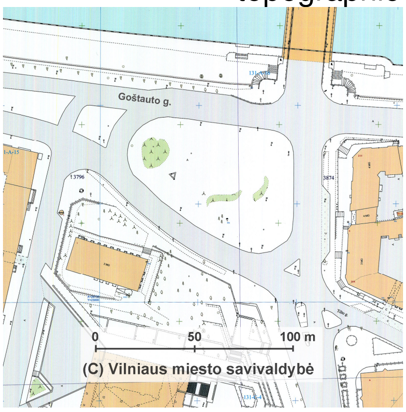


Figure: Municipal topographic map at 1:500 scale

- Aerial photographs and/or satellite imagery can be used for large- and medium-scale maps of varying scales
  - Photogrammetric and remote sensing techniques can be used for image corrections and height extraction (e.g. automatic stereo-plotting); digitizing techniques area used to extract vector planar topographic data. All data can be stored in a topographic database and later used for topographic mapping
  - This method is also used for topographic map updating
  - Thus, satellite imagery was used for compilation of Lithuania’s 1:50,000 topographic base maps; 133 maps give complete coverage of Lithuania

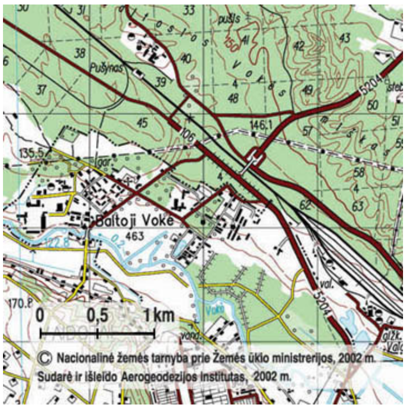


Figure: Lithuanian topographic map at 1:50,000 scale

- Existing topographic digital spatial data sets or databases can be used as a source of spatial and attributive data for medium- and large-scale topographic maps
    - Most digital spatial data are collected from existing topographic maps through digitizing
    - Data from digital mapping can be used:
        - If spatial resolution (level of details) of digital dataset are close to mapping the scale then not much cartographic generalization will be needed
        - If the level of source details does not corresponded to the scale of the compiled topographic map, a lot of generalization has to be applied and it can be time consuming. This mapping process is used for data at large scales to create small scale maps
    - Spatial data also can come from digital databases that store data obtained from ground survey and/or photogrammetric methods
        - These data also can be used for topographical mapping of smaller-scale maps than from original data sources
- Combination of above methods can be used for topographic mapping as well

Usually the rule is that the more detailed the data sources (in larger scales) have to be used for compilation of smaller scale map (or less detail topographic map). For example, 1:8,000 scale aerial-photos can be used for updating 1:10,000 maps.

GIS, remote sensing, surveying, plotting (e.g. Survey Analyst), map publishing and label placement software are used for digital map compilation and publishing process that can include sequence of operations. These operations can be predefined within the map editor instructions and pre-programmed by a GIS Analyst.

For example, PLTS tools for ArcGIS can be used by the map editor to set up technological sub-steps of map compilation data flow. Its Mapping Agency Solution provides the tools to enable streamlining of map products through a consistent process (data control, map sheet management, tracking workflows, etc). The Mapping Agency Solution provides the following functionality:

- Generate digital and hard-copy map products
- Automate map sheet management
- Database driven validation and symbolization

- Rule-based map surround placement
- Symbol set management and editing
- Map footprint generation and management

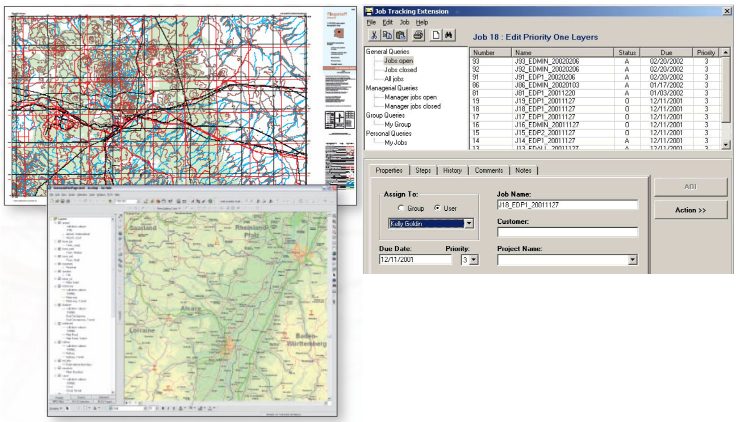


Figure: Production Line Tool Set (PLTS), ESRI

Module self-study questions:

- What are the visual design variables?
- Which are associated with nominal data?
- What are the three components of colour?
- What is the special significance of using a solid red for any elements?
- Other than oceans, what features are typically shown in blue patterns on a topographic map sheet?
- Why must contour lines never cross, merge or split?
- What is the primary contour interval used on 1:50,000 topographic maps of Lithuania and in what units are they expressed?
- What is a *contour interval*?

Required Readings:

- **The Cartographer's Palette: The Semiotics of Cartography,** The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/cartocom/cartocom_f.html
- **Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping,** Alessandro Cecconi, http://e-collection.ethbib.ethz.ch/ecol-pool/extdiss/extdiss_6.pdf
- **Topographic Mapping**, USGS, http://erg.usgs.gov/isb/pubs/booklets/topo/topo.html
- **Chapter 2: How maps inform,** Modeling Our World, Zeiler, M., ESRI Digital Library, 1999.
- **Canadian Topographic Maps Chapters,** Centre for Topographic Information, http://maps.nrcan.gc.ca/index_e.php

ESRI Virtual Campus Course:

- **Module 1: Big Picture Design** and **Module 2: Type Basics**, Cartographic Design Using ArcGIS 9

Assignment:

- **Assignment 6:** Topographic map production with GIS

References

- [1] Thematic Cartography and Geographic Visualization, Terry A. Slocum, Robert B McMaster, Fritz C. Kessler, Hugh H. Howard, 2nd ed., 2004.
- [2] Elements of Cartography, Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, Stephen C. Guptill, 6th ed., NY: John Wiley & Sons Inc., 1995.
- [3] The Geographer's Craft notes, Department of Geography, University of Colorado, http://www.colorado.edu/geography/gcraft/notes/notes.html

Terms used

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale
- Visual variable
- Model-based generalization
- Object-based generalization
- Cartographic generalization
- Controls of generalization
- Selection
- Simplification
- Collapse
- Aggregation
- Amalgamation
- Typification
- Exaggeration
- Displacement
- Smoothing
- Classification
- Measure
- Constrain
- National topographic system
- Scale series
- Topographic map index