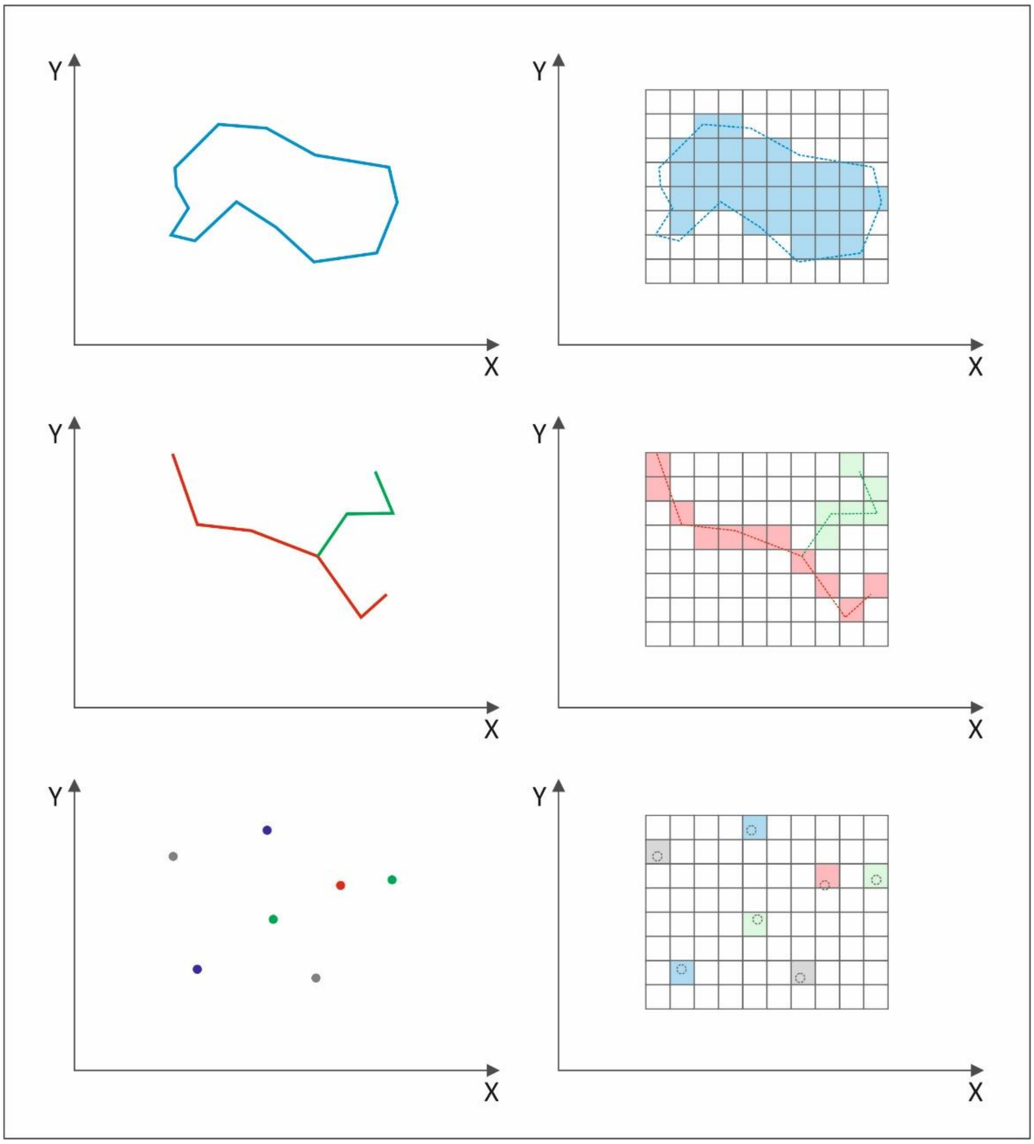
GEOGRAPHY:
a digital approach

Jacek Kozak



Geography: a digital approach

Jacek Kozak

The book was developed as part of *Projekt ZintegrUJ – Kompleksowy Program Rozwoju Uniwersytetu Jagiellońskiego*, co-financed by the European Union, the European Social Fund, and implemented on the basis of the grant agreement POWR.03.05.00-00-Z309/17-00

Kompleksowy Program Rozwoju Uniwersytetu Jagiellońskiego

Fundusze Europejskie Wiedza Edukacja Rozwój

Uniwersytet Jagielloński w Krakowie

Unia Europejska Europejski Fundusz Społeczny

POWR.03.05.00-00-Z309/17

The book is based on the 1st Polish edition, ISBN 978-83-64089-70-1, published by the Institute of Geography and Spatial Management, the Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Polska, available at: https://sigkit.geo.uj.edu.pl/documents/142531035/142888446/Geografia_podejscie_cyfrowe/f83966f2-0642-4b3a-bb58-bb9c830a0f3a

Translation by the author. Proof reading: Guy Torr

Kraków, 2022

ACKNOWLEDGMENTS

The book *Geography: a digital approach* constitutes the second textbook on geographic information, written at the Jagiellonian University's Institute of Geography and Spatial Management. The first handbook, *Introduction to Geographic Information Systems*, was published over 20 years ago. Its author, Prof. Wojciech Widacki, initiated research into the field of geographic information and founded the Geographic Information Systems Laboratory at the Jagiellonian University's then Institute of Geography. He was also the creator of the postgraduate studies course *UNIGIS Geographic Information Systems*. This first textbook summarized Wojciech Widacki's experience within a new field for Polish geography, but also drew upon his extremely rich knowledge of the theory of geography, in particular physical geography, and the philosophy of science. This is the greatest debt I owe to Wojciech Widacki in writing this book: the belief that issues related to geographic information are closely related to the theoretical foundations of geography. Thanks to Wojciech Widacki, I also know that for beginners in this field, the theoretical background is as equally important as a good knowledge of geographic information technology and skills in using the dedicated software. This book stems from this conviction and, first of all, it attempts to present the basic theoretical issues of geographic information, ones which will not become quickly outdated in this era of rapid technological change.

Geography: a digital approach summarizes over a dozen years of my academic experience, mainly related to the integrated course *Geoinformatics* which first-year students of geography attend at the Jagiellonian University's Institute of Geography and Spatial Management. To a large extent, this book is intended to assist students on this course. However, to write it, I have also used the experience related to running the specialization option *Geographic Information Systems* at the MSc level, post-graduate studies *UNIGIS Geographic Information Systems*, as well as classes in the philosophy of science for students of geography. Of course, I have not conducted these classes alone, and their preparation involved long, often stormy discussions with my colleagues from the Department of Geographic Information Systems, Cartography and Remote Sensing. Among those to whom this book and its author owe a lot, I would like to mention Mateusz Troll, with whom I have been working for more than 30 years, and Dominik Kaim, Natalia Kolecka, Małgorzata Luc, Krzysztof Ostafin, Aneta Szablowska-Midor, Piotr Trzepacz and Bartosz Załuski. Discussions with current and former PhD students on various topics relating to geographic information have also been important to me, in particular those with Monika Dobosz, Andrzej Kotarba, Said Nawar, Marcin Szwagrzyk and Elżbieta Ziółkowska. Many of my colleagues have prepared exercises for the *Geoinformatics* course edited by Mateusz Troll, and published on-line on the University e-learning platform, forming an excellent supplement to this book.

I would like to thank Ms Karolina Mostowik, a PhD student at the Jagiellonian University's Department of Hydrology of the Institute of Geography and Spatial Management, for sharing the figure and data used in Chapter 1.

TABLE OF CONTENTS

INTRODUCTION: WHAT IS THIS BOOK ABOUT?

This book deals with the most important concepts and problems relating to how reality is put in the computer,[1] or in other words, how various phenomena of the world around us are digitally represented. I have divided the book into eight main chapters, which partly retain the nature of a textbook on the theory of geographical information, and at times depart from a somewhat rigid academic lecture, becoming more an essay on geography. These eight chapters discuss the most important issues of digital modelling of real world phenomena, taking into account their spatial and temporal aspects, emphasizing the importance of this modelling for geography and for other sciences that deal with the complex reality that surrounds us.

If I were to point out some inspirations for this book, I would certainly mention William Bunge's *Theoretical Geography* (1962) and the well-known geography textbook by Peter Haggett: *Geography: A Modern Synthesis* (1983). Despite their many differences, they have one thing in common – they present a coherent vision and concept of the theoretical foundations of geography. It was very tempting for me to take up such a challenge and look at the foundations of geography through the lens of modern technologies enabling one to digitally model the world around us.

So this book is first and foremost a book about geography, and therefore it is worthwhile to begin with specifying what geography is, or at least roughly. Among many definitions of geography that have appeared in the history of this science, most emphasize **two aspects of this discipline**: the relationship between human beings and the environment in which they live, and the spatial approach to natural and anthropogenic phenomena (for instance, Turner, 2002). The first aspect imposes the selection of topics that geography is interested in. It is worth noting that the selection changes with time: purely nature-related problems have become less and less important in geography, while the significance of social, economic and cultural problems has been continuously growing, because of the increasing causative power of humans in relation to nature and the expansion of areas largely dominated by artificial environments. The importance of the human-environment relationship is confirmed in the discussions devoted to a new era in the history of the Earth: the Anthropocene, its various dimensions and consequences (Bińczyk, 2018). The second aspect is emphasized by the approach or method proper to geography: phenomena that are of interest to geography are considered not only from the point of view of their properties, but also – or perhaps above all – from the point of view of their location, geometry and spatial relations. The latter aspect is of old a distinct feature of geography, clearly articulated since the times of Immanuel Kant, whose views were adapted and developed by such geographers as Alfred Hettner and Richard Hartshorne (Turner, 2002), but also by their critics, who in the 1950s re-directed the attention of geographers to relational spatial approaches, creating the foundations for the development of modern spatial analysis and quantitative geography (Schaefer, 1953).

In this book, geography will be understood as the study of phenomena having various spatial properties located or occurring on the Earth's surface that are of importance to humans living on the Earth and shaping it.. Though atmospheric, hydrospheric and lithospheric phenomena important for humans occur at certain heights or depths, geography is a science about the Earth's surface and reduces the third dimension to the necessary minimum. When considering the question as to which phenomena may be the subject of geographical research, an important clue is that these phenomena take place at a scale appropriate to the perception by humans of the surrounding world and are most often limited to observable macroscopic objects and processes occurring in periods adequate to the length of a human life, both now and in the past (although geography also takes into account much longer processes, insofar as they are relevant to humans).

[1] This metaphor was used by David Gelernter in his book *Mirror Worlds: Or: The Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean* (Oxford University Press, 1992).

It is worth adding that the choice of phenomena considered by geographers is the subject of a long and unfinished debate, one of secondary importance for purposes of this book.

With such a definition of geography, it is not surprising that the methods of recording the locations and properties of various phenomena and the selection consequences of these methods given various problems are of fundamental importance, largely influencing the relevance of the research results and the conclusions geographers draw on different aspects of reality. This was clear from the very beginning of the digital modelling of real world phenomena, when the spread of quantitative approaches quickly led to the first successful combination of the computing power of the first computers and spatial analysis methods. This combination is known as *Canada Geographic Information System* – the first working Geographic Information System (GIS), designed in the 1960s under the leadership of Roger Tomlinson (Longley et al., 2011). The development of subsequent systems of this type required an in-depth discussion of the concepts relevant to the modelling of data used in these systems, as well as a broader theoretical reflection on geographic phenomena. The result of this discussion was the extension of the original, technologically-oriented concept of GIS as a computer system towards the theory-laden concept of *Geographic Information Science* that emerged in the early 1990s (Goodchild, 1992). Currently, both the theoretical and technological aspects related to geographic information are merged in the wide discipline of *Geographic Information Science and Technology* (GIS&T), well-defined and described in detail (DiBiase et al., 2006).[2] And this is exactly where geography meets the problems discussed in this book.

In Chapter 1, I will look at how we perceive and understand the reality around us, and how we move from perceiving the reality to interpreting and describing it. The basic categories of description are those that are then subject to our observations and measurements, leading further to data taking into account the location of phenomena (Chapter 2). Data models, that is formalized ways of recording selected aspects of the real world, are discussed in Chapter 3. The next two chapters discuss two important types of data: elevation data, giving geography a substitute for the third dimension (Chapter 4) and image data (Chapter 5), which – at least seemingly – present the world in the same way as we observe it. In Chapter 6 I deal with time, and more particularly with how the variability of Earth surface phenomena is represented. Chapter 7 focuses on various ways of modelling point, line and areal phenomena and the multitude of possibilities that these simple representations offer. Chapter 8 deals with extremely important issues of data quality – that is, the relationships between data, data models and reality. At the end of the book, I explain the reasons why some seemingly obvious topics related to geographic information have not been addressed (to find out what has been omitted you will have to read the entire book or simply check the last chapter). This book therefore deals primarily (but not exclusively) with the issues of two so-called *Knowledge Areas*, defined by DiBiase et al. (2006) within the framework of GIS&T as: *Conceptual Foundations* and *Data Modelling*.

For whom is this book? The simplest answer is that the book is for readers who want to understand the basics of modern geography, a science that makes extensive use of modern data acquisition technologies and has access to a wide variety of advanced data analysis and visualization tools. Among the readers, there can be both students of geography (or related sciences) and professional researchers who deal with various phenomena located somewhere on the Earth's surface. Everyone who reads this book should remember that the vision of geography presented in it may, and should, be the subject of critical interpretation and reflection – quite likely, with many shortfalls that can be identified and improved on.

[2] *Geographic Information Science and Technology Body of Knowledge* is being currently updated by the team led by John Wilson, University of Southern California, US, with results available at https://gistbok.ucgis.org/ (accessed: March 2020).

1. REALITY AND HOW WE PERCEIVE IT: TOWARDS REPRESENTING PHENOMENA

Keywords: scientific realism, phenomena, objects, processes, events, properties, conceptual modelling, first and second level objects, continuous fields, spatial scale, dimensionality.

The first chapter requires, above all, careful reflection on the world around us: what it is, how we perceive and have come to know it. Thus, the chapter goes beyond geography, touching upon issues that have for long been of interest to philosophers. For philosophy is concerned with both ontology, that is what the world is like, and epistemology, that is how we have come to know this world.

In a work devoted to ontology issues in the theory of geographic information Frank (2001) distinguishes several ontological levels leading from reality to the representation of phenomena. The first three levels are: [1] the physical reality, [2] observation of the physical reality by cognitive agents and [3] how agents form objects from their observations. These levels can be linked to a procedure that, in database theory, leads from reality to the data that describe it (Longley et al., 2011). Two stages of the procedure are **conceptual modelling** and **logical modelling**. Conceptual modelling is connected with perceiving, interpreting and verbalizing the phenomena of the real world around us – thus it is an attempt to capture and express the diversity of the real world. In turn, logical modelling translates these ideas into data models and their recording in a digital form. Somewhere in-between these two stages, observation and measurement can be distinguished as the methodical examination of quantitative and qualitative properties of real world phenomena, supported by various devices (Frank, 2001).

The first chapter of this book deals therefore with conceptual modelling, that is how the nature of the real world is perceived, discovered and expressed. Its logical consequence are the second and the third chapters, which discuss measurements and observations of the physical world and the methods of data storage using formalized models. This linkage of three ontological levels with the corresponding three chapters should, however, be treated to a large extent as somehow arbitrary – observation cannot be clearly separated from understanding the nature of the observed reality, and measurement and observation cannot be considered independently of the logic of data models. Nevertheless, the order of the first three chapters proposes a conceptualisation pattern that leads from the reality we perceive to abstract models by which various aspects of this reality are stored in computers.

In the considerations that follow it is worth adopting the position of **realism** and **scientific realism**. According to Thomasson (2001), a realist should accept that certain things (phenomena) exist and are as they are, regardless of any states of the human mind. However, this does not mean that, apart from things (phenomena) which are independent of the mind, there cannot be things which are dependent on the mind, subject to scientific inquiry. In turn, as Chakravartty (2017) explains, scientific realism is a positive epistemic attitude toward the content of our best theories and models, resulting in the belief in both the observable and unobservable aspects of the world described by science.

I will use the general term *phenomenon* to describe the manifestations of reality that surround us, which we perceive and are able to name, instead of the term *entity*, quite handy and often used in deliberations on the theory of geographic information, but rather non-intuitive. First of all, I will deal with the phenomena of the surrounding reality that are important to geography as a science.

What do we see?

Let us start with a simple example. As can happen on a mountain trip (especially when we have not read the weather forecast before going to the mountains), in the area we were crossing there was a cloudburst and an intense, though short-lived downpour. As it happens in such a situation, first of all we tried to find shelter – and if we were lucky enough to do so, we could observe various phenomena while waiting for the rain to stop. With a high intensity of rainfall, water began after a while to flow down the slopes, and if a stream was flowing nearby, we noticed how the flood started. The water was muddy – during the rain, it usually carries a lot of soil material washed out from the slopes. Puddles and mud appeared on the trail, which usually means that the rest of our journey will be a bit slower. If we sheltered under trees because we were hiking in a wooded area, then after some time we had to look for a better shelter because water started to drip down on us from the leaves or needles. What we saw during our trip is a typical meteorological and hydrological phenomenon: an intense, short rainfall, triggering various processes related to the flow of water downslope or its infiltration into the soil. To describe this phenomenon, we used a whole range of terms referring to various elements of the mountain landscape and the processes triggered by rainfall. What specific names and terms we used to describe a situation depended on our knowledge – we are not always able to choose the right ones for the situation in question. For example, the temporary possibility of sheltering under trees results from interception – that is, the process in which vegetation cover intercepts and retains some of the rainfall.

Perhaps we were a bit lucky and from the researchers conducting measurements in this area, we learned that in this mountain catchment with an area of 4.4 km^2, as a result of intense rainfall lasting two hours, a surface runoff was started and a rapid flooding in the streams flowing within the catchment area took place. The rain gauges recorded between 40 and 55 mm of precipitation, and on the main stream at the catchment closure, the maximum flow was recorded on the day of rainfall at 10.45 pm, equal to 1.03 m^3/s, with the water level reaching 82 cm.[3]

Though it may appear not overly complicated, this short description of the rainfall event contains an extremely large content load. First, the text refers to various **objects**. Although the term *object* seems obvious and understandable at first glance, it is worth defining its meaning in this book to avoid any misunderstandings: an object is a part of reality that can be separated from other parts (Bian, 2007), relatively persistent and material (Rettler, Bailey, 2017).[4] In the above scientific description of the rainfall, three objects are quite obvious: the catchment, the stream and the rain gauge. Objects have different **properties**: for example, the catchment is *a mountain one* and the stream is *main*. Objects and their properties are the basic elements used to describe the reality, and typically, one can add also the relations between them, receiving a triad of e*ntities*, *properties* and *relationships* (Peuquet, 1988). Moreover, the following processes take place in the catchment: precipitation and the resulting surface and subsurface runoff, infiltration, interception, runoff, erosion. Processes are related to the flows of matter (in this case water and, for example, the soil particles carried by it) or energy, although it is more accurate to say that processes are flows of matter or energy. The matter and energy is transported in a set of objects that can be considered as interconnected containers (Chorley, Kennedy, 1971), in which the properties of objects shape the flows of matter and energy, while, on the other hand, the flows interact with the objects (containers). In the case of the described catchment, its mountainous character results in faster transport of rainwater to the stream channel than in the case of a lowland catchment, due to the steeper slopes. But there are many more features of the

---

[3] These are actual data from measurements carried out by Karolina Mostowik of the Jagiellonian University's Institute of Geography and Spatial Management, as part of research into the water cycle in the Bieszczady Mountains. I would like to thank her sincerely for sharing this. The studied area is presented in Figure 1.1.

[4] The term *object* has many meanings and objects are defined in various ways. The work of Rettler and Bailey (2017) referred to in the text contains an extensive overview of this subject.

catchment that affect the rate and scale of the transformation of rainfall into runoff – such as soil permeability, vegetation cover, or the number of permanent and periodic stream channels that can drain the water.

These properties are of a different nature. For example, forest cover (the forest area share in any given area) can relate to the whole catchment area and can be expressed as one numerical value, for example a value of 75% would mean that ¾ of the catchment area is covered with forests. Soil cover permeability is usually determined at different points in the catchment area and may have various values at each point. Thus, permeability is a property of soil that varies in space.[5]

Permeability or forest cover are properties of the (land) surface of the Earth or any part of it, so referring these properties to a specific catchment is only an expression of our interest in that object, and is not done because such properties do not exist outside that specific catchment. Similarly, the number of permanent and periodic watercourses is a property of the entire catchment area – in this case the resulting numerical value characterizing one object, that is the catchment, depends on the number of other objects, that is watercourses, contained in the former. To some extent, such an occurrence of objects within another one determines also the forest cover, which depends on how many trees (objects) are within the catchment (another object). In summary, a catchment area has properties that are determined and constant for the whole object (for example, forest cover, total area) or properties that change depending on the exact position within the catchment, for instance the already mentioned permeability of soils, elevations or slope gradients – generally strongly varied in mountain catchments. Similar reasoning can be also applied to the main stream channel: it has properties that are defined for the whole object, such as length and average slope, and properties that vary in different parts of the object (for instance, a different width and depth). It looks a bit different in the case of the third of the above-mentioned objects, that is the rain gauge. Although, when examining this device, we can conclude that its properties are different in different places (e.g. colour, thickness), from the point of view of the function and purpose of rain gauges, these properties and their variability do not have any special significance. Properties that are important from the point of view of researchers who carry out research into the area may refer to the entire object – it will be, for example, the type or accuracy of the rain gauge.

In the description of the rainfall and the flood, the fourth object is somewhat hidden: this is the closure of the catchment. A catchment closure is a location to where water drains from the entire catchment area above. Due to the fact that each point located on the main watercourse may represent a catchment closure, it materializes itself only in the field when the researcher decides to place measuring devices there. The catchment closure can have different properties, for example one of them will be its elevation. For obvious reasons, a catchment closure is the lowest point of the catchment it closes.

Each of the above-mentioned four objects has specific spatial properties. The catchment boundaries, running through watersheds, and watercourse channels are located in specific places. Also a rain gauge and a catchment closure are located at a specific point. It is worth noting that the location of the catchment closure determines the boundaries of the catchment itself (Fig. 1.1). The values of object properties are also localized – e.g. a specific permeability may occur in one place of the catchment, and not in another, similarly, forest cover is a property assigned to the whole catchment area, and thus it characterizes a certain fragment of the Earth's surface. As already mentioned, objects are linked together through a variety of spatial relations. Examples of such relationships are the occurrences of some objects within the boundaries of others – for example, watercourses lie within the catchment and there is a rain gauge in the catchment area. Other

[5] Forest cover may be also assessed for smaller parts of the catchment, and, therefore, it may vary from place to place, in a similar way as permeability. On the other hand, permeability may have one single value (e.g., average), for the entire catchment.

examples of spatial relations are connections and intersections – for example, individual watercourses connect to each other to form a river network, one in small catchments not usually very complicated.

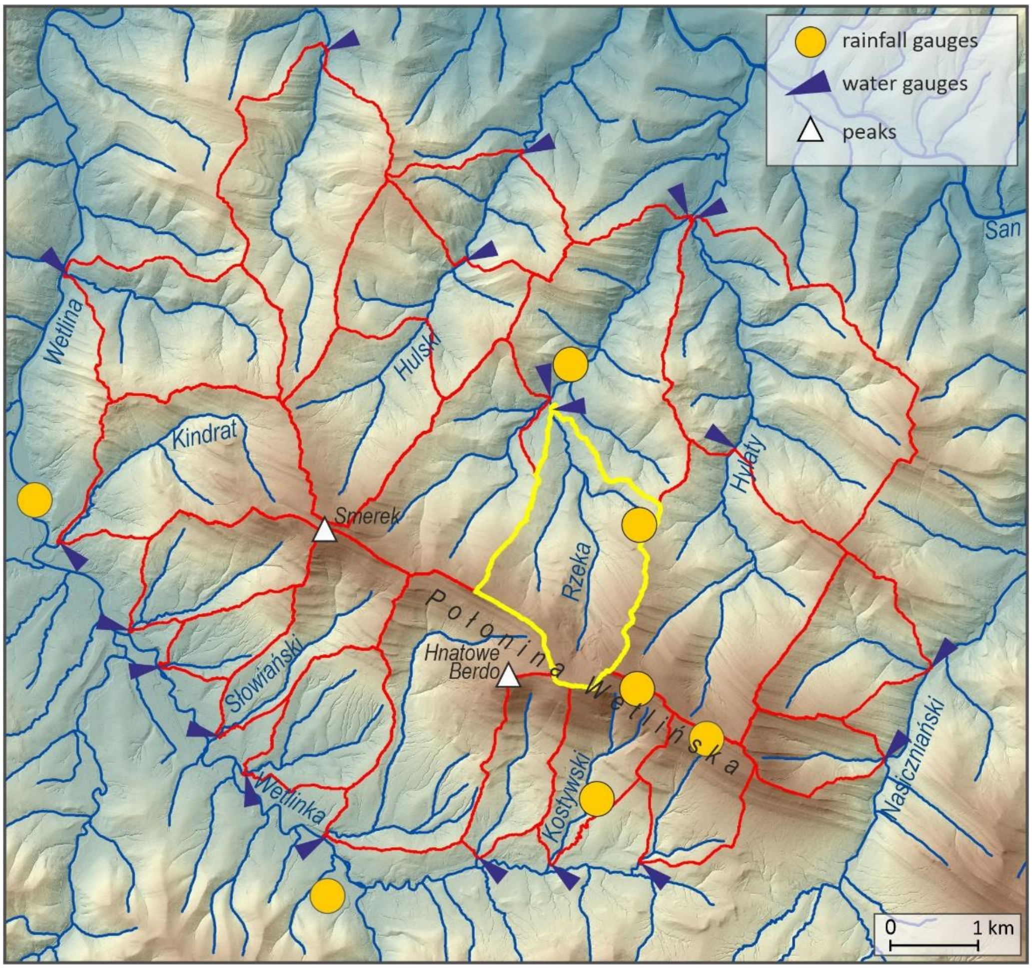


Fig. 1.1. Catchments in the Połonina Wetlińska range in the Bieszczady Mts.

The catchment, for which rainfall and runoff characteristics was given (Rzeka) is shown with a yellow outline. Catchment boundaries depend on the elevations and locations of catchment closures, quite frequently located in places where two streams join each other (the the case of Rzeka). Hydrological measurements are also carried out in catchment closures. A closure may be also located using other criteria, e.g., as is the case of Hylaty.

Courtesy: Karolina Mostowik, Department of Hydrology, the Jagiellonian University's Institute of Geography and Spatial Management.

The properties of objects change over time. The permeability of soil can change quickly (within days or even hours). In contrast, the slopes over the vast majority of the catchment are relatively constant (they do not change over hundreds, thousands or even millions of years). Processes take place over time – for example, flow refers to the instantaneous dynamics of the water flowing in a river or stream and varies from moment to moment. At this point, it is worth emphasizing one more term that appears in the description of the rainfall provided by the researchers, that is a flood. A flood is an **event**: an identifiable part of a process or processes that take place in specific sets of objects. In this context, a flood occurs when the runoff is significantly higher than the average, most often the result of intense rainfall or snow melt and has a relatively short duration (relatively short in this case means that the flood duration is short in relation to the duration of hydrological observations or measurements).

Intuitively, events are different from the objects that we perceive as material, unchanging, or changing very slowly, while events appear and disappear over time and are not directly material in their nature. Objects last, and events just happen (Casati, Varzi, 2017). However, a lot depends on the time scale in which the observer perceives the surrounding reality. From our point of view, a flood that lasts several hours in a small river is an event, just like a road accident that takes place within seconds. However, for a hypothetical creature living for millions of years, an event could be a lake that lasts several thousand years, an object for that creature as ephemeral as a puddle appearing after rain seems to us. An interesting difference between objects and events is also that objects can change their position and events cannot – the latter are, in a sense, assigned to a specific location (Casati, Varzi, 2017).

Can we always name correctly what we see?

The initial (but at the same time essential) stage of interpreting our perceptions of reality is the naming and defining of phenomena. When we look outside the window, we notice various phenomena that, in the vast majority of cases, we can name or describe. However, the more they differ from common experience, the more difficult it is to correctly identify and name them, and in many cases expert knowledge is required, supported by knowledge of definitions and various terminological nuances.

Let us consider the following example: even without biological expert knowledge we have in principle no problem to correctly identify and name an object of the *tree* type, distinguishing this object from others (for example, shrubs), although the diversity of trees in our climate is quite large and each of us can see different examples of such objects practically every day. However, many people could have a problem with bamboos, a plant which in Southeast Asia can reach a height of several metres with what looks like a trunk and parts of the plant that reminds one of a tree crown. Bamboos, however, are grasses, and although the term *bamboo forest* is used, the forest is not composed of trees. An even more difficult problem is posed by the term *forest* itself. Again, it would seem that in the vast majority of cases we know when we are in a forest and when we are not. Nevertheless, science and practice distinguish various ways to define this type of object (Box 1.1).

Box 1.1. Forest definitions

Lund (1999) distinguished over 100 forest definitions used in different countries, which generally apply a combination of three criteria: vegetation that grows in the area considered to be a forest, the human activities in forests, and finally, the minimum area of a forest patch. Chazdon et al. (2016) has discussed different aspects of the whole range of forest definitions used in different countries, highlighting historical changes in the functions that forests have had for humans and how these changes affect the ways in which forests are defined. Since the 18th century, the main function of forests has been to provide timber, and therefore the aspects related to timber production are emphasized in numerous forest definitions. Currently, functions related to nature protection and carbon sequestration have become important, which means that forest definitions increasingly include aspects relating to biodiversity and biomass accumulated in forest ecosystems, which have not been taken into account before. In the definition accepted in the global forest assessments the *Food and Agriculture Organization of the United Nations* (FAO) defines forests as *land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use* (Terms and Definitions FRA 2020, 2018). It should be noted that the FAO definition refers both to how the forest should look (it has to be an area with a sufficiently dense tree cover, with trees of a certain height) and to how the forest is to be used, and also determining its minimum size. In many countries, similar definitions are adopted, but often with slight modifications to the criteria. The definition of forests used in Poland is quite similar to the FAO definition, although the Polish definition uses a lower forest area threshold (0.1 ha) than in the FAO definition.[6] According to the Polish definition, a forest can also be an area temporarily devoid of trees, as long as it is used according to forest management principles. This seemingly paradoxical statement (a forest without trees) is perfectly logical, because if trees were cut down in some place, and then new trees were planted immediately, this area, from the economic point of view, is still a forest (even when temporarily there are no trees at all). The only difference is that instead of old trees, young ones are growing or will soon grow in this area. In turn, not every area covered with trees in Poland is a forest – for example, urban parks are not forests, because the trees growing there have other functions than the trees growing in forests.

Continuous and discrete approaches to reality

From the above description of a flood in a catchment, as well as considerations on forests and their definitions, as well as taking into account the range of spatial scales in which geographers operate, it appears that geography deals with objects of two types, or levels. Let me call them here first and second level objects. **First level objects** are fragments of the Earth surface (both land and water) of various sizes – from very small, measuring square metres, to very large, millions of square kilometres – including the entire Earth. By *the surface of the Earth* I mean the outer layers of the hydrosphere and the lithosphere, with the lowest layer of the atmosphere adjacent to them, where life (biosphere) develops and humans live and interact. This thin sphere (or shell), with a negligible thickness compared to the size of the entire Earth and irregularities resulting from the variation in the topography, is the main object of geographical research.

First level objects and their properties can be described in two different ways. Let us return at this point to the example of a mountain catchment in which there was sudden rainfall and consider one of its important properties, that is the shape of the land surface, or simply elevation, which affects the water cycle and runoff formation (elevation as a property in the extreme case refers to the entire Earth surface, including also the ocean floor). The first way of description is as follows: we measure and assign some values characterizing the features of the surface of the whole catchment, for example those which are important from the point of view of how rainfall transforms itself into runoff. Examples can be an amplitude of elevations, mean elevation, mean slope or valley density. Note that the value of each of these properties applies to the entire catchment, strictly defined in a spatial sense and treated as one entity. The second way of describing the catchment surface consists of determining the value of elevation at selected points in the catchment. Unlike the first method, with this approach, we do not need to define the boundaries of the catchment in advance, because measuring elevation at a given point will give the same result regardless of whether we know where the catchment boundary is or not. In the case of the first method, the specific values of the selected properties of the catchment depend on the prior accurate determination of its boundaries. For instance, the change in locating the catchment closure alters the values of the properties describing the catchment.

The described methods are adequate for other fragments of the Earth surface, both larger and smaller than the catchment area. Looking more carefully at the catchment elevation – for example, along a selected profile (Fig. 1.2) – we can try to distinguish smaller fragments that correspond to the characteristic relief forms, such as ridges, plateaux, slopes, valley floors, hollows. By naming and defining each relief form, we distinguish it from its surroundings and define its boundaries. Each such relief form can then be described as a whole using its selected properties, resulting in a more detailed picture of the catchment surface than in the case of the first method, in which the selected property of the whole catchment was described using a single numerical

[6] Forest Act dated 28 September 1991.

value. On the other hand, since each of the delimited relief forms can be characterized by specifying the value of a selected property (e.g., elevation) in different places, dividing a larger first level object into smaller ones does not exclude the use of the second method in relation to smaller first level objects.

These two approaches to describing first level objects relate to two fundamentally different ways of understanding and describing reality. With the first approach, a discrete one, the diversity of the real world is reduced to the diversity of objects in it, of a discrete nature, that is **discrete objects** – objects that have boundaries separating them from other objects and are countable. For instance, in the Polish Carpathians there are several mountain ranges (Tatra Mts., Silesian Beskid, Pieniny Mts., …) and the topography of the Carpathians can be described using the properties of these ranges, each treated as a discrete object. However, the boundaries between them are not always so sharp that their definition remains crystal clear. For example, the northern boundary of the Tatra Mountains is considered to be explicit, yet this is not the case of the northern boundary of the Gorce Mountains. The second approach to describing reality, a continuous one, is a slightly more abstract. It does not distinguish the boundaries of the first level objects and focuses solely on the very properties of the Earth surface. Values of these properties can be determined (for example, measured) at any location. Properties, for which values can be determined by such measurements, form **continuous fields**. In this way, we can describe the topography of the Carpathian Mountains, providing elevation values anywhere in these mountains, without paying attention to whether these locations are in the Beskid Mały, Pieniny or Tatra Mountains themselves.

In addition to the first level objects, that is freely defined fragments of the Earth surface, geography also deals with objects so small that their size and internal structure are negligible at scales appropriate for geographical considerations. However, the distribution of these objects and some of their properties are extremely important for these considerations. I define these objects as **second level objects**. Examples of such objects include trees, wells, buildings, or springs. An important demarcation feature between objects of the first and second levels is that the former, unlike the latter, can be fragments of the Earth's surface of any size, with the only limitation being the physical size of the Earth itself (or of its land surface). The first level object is, for example, a forest complex, which can be as small as a grove of several hectares, or as vast as the entire Siberian taiga. Another example of first level objects are countries that can be of the size of San Marino, but also of Canada or Brazil. As for the second level objects, although it is possible to imagine larger or smaller objects of this type, there are physical or technical barriers that limit their size. In this respect, events are similar to the second level objects as they are limited in terms of their spatial extent (and in addition, their duration is limited). The discrete approach is typically used in geographic research to describe second level objects and events.

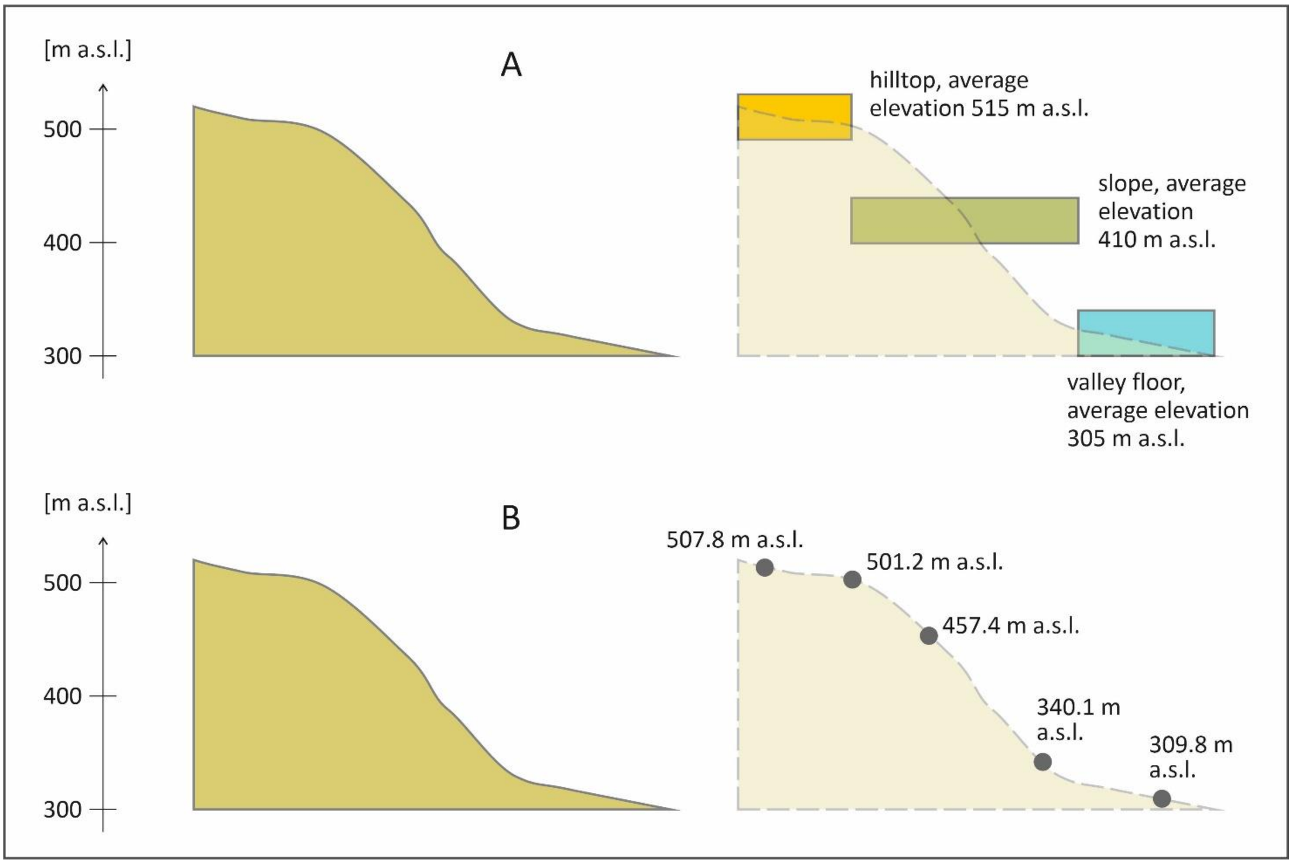


Fig. 1.2. Description of the topography along the hypsometric profile.

The profile is shown on the left. A – discrete description, relying on the division of the profile into distinct relief forms, with specific properties; B – continuous description, through a specification of values of a selected property at selected locations.

Discrete and continuous approaches to describing reality are commonly used in everyday life. Going from Jabłonka through Lipnica Wielka to Babia Góra along the green trail, at some point, at the northern end of the village of Lipnica Wielka, we leave the slightly undulating terrain gradually rising to the north and start the arduous climb up the slope of Babia Góra. Somewhere in this area there is a border between the region called Działy Orawskie (or the Orawsko-Jordanowskie Foothills) and the southern slope of Babia Góra, included in region of the Beskid Żywiecki Mts. It can be said that at the point where the steep climb starts, one object (Działy Orawskie) ends, and another one (Beskid Żywiecki) begins, the culmination of which – Babia Góra – is our destination. A tourist taking the same trail, yet unaware of the intricacies of the regional divisions in the Carpathians, who does not know the names of the regions where she is wandering, will easily notice in the same area that the path begins to climb steeper and requires much more effort than before. This tourist will therefore notice a change in the nature of a continuous property of this part of the Earth surface during her journey, without noticing that the boundary of two discrete objects is crossed.

Discrete and continuous approaches also apply in the context of the forest definitions discussed previously. The characteristics of the trees themselves (e.g. height) and of their occurrence in a given area (specific density, area covered) are used to define forests. Different configurations of these characteristics can create a problem in determining whether a given area is a forest or not. For example, in the mountains, at a certain elevation, trees gradually dwarf and thin out, and when ascending, the forest ends – however, determining the line at which this change occurs (that is the so-called upper timberline) is often difficult and may require the adoption of additional criteria (Fig. 1.3). The upper timberline and the problems associated with its designation are therefore an example of the complementarity of discrete and continuous approaches to the description of reality. The distribution of trees – second level objects – allows one to determine the boundaries of forest complexes and to separate forests from non-forested areas, thus distinguishing discrete first level objects within the studied fragment of the Earth's surface. On the other hand, we can also describe the studied fragment of the Earth's surface using a continuous approach, taking into account only one specific property – the number of trees per unit area (tree density), without considering where a particular value of this property determines the fact we are dealing with a forest or not. I will return to this issue in Chapter 7.

These two approaches, discrete and continuous, are a different but also complementary way of perceiving and representing reality. The first focuses on the qualitative aspects of reality where the proper naming and definition of discrete objects is most important. The second approach focuses on measurement and quantitative aspects of reality. It should be added that the discrete approach does not exclude quantitative measurement and description, similarly as the continuous approach does not exclude a qualitative one. Several papers (e.g. Bian, 2007) treat these two approaches as equal, alternative ways of perceiving and describing reality: a mixture of discrete objects and continuous fields. I believe, however, that such a distinction makes sense in geography only in relation to the Earth's surface and first level objects. Second level objects, without exception, are treated in geography as atomic discrete objects and the continuous approach is of little relevance in dealing with them.

Dimensionality of objects

Geography is the science that deals with the Earth's surface at a spatial scale appropriate to the human perception of various terrestrial phenomena. In fact, it is not a single scale, but an entire scale range, due to the different ways in which humans perceive the Earth surface. This range is often described with a series of rather vague, but self-explanatory terms: local – regional – continental – global scale. The spatial scale is fundamental for representing reality and conceptual modelling, and its significance can be illustrated by a simple example. At a scale corresponding to the perception of the Earth's surface with our senses, and when walking on our own feet (that is at a local scale), Babia Góra – the highest mountain in the Beskidy in Poland – is a vast mountain range, very diverse in terms of its properties (e.g., topography, vegetation, soils). One can spend months at Babia Góra studying its biodiversity and geodiversity. However, when viewed from an altitude of several hundred kilometres, from the deck of a spacecraft orbiting the Earth (continental or even global scale), Babia Góra reduces itself practically to a point – it is a bit more than a dot, a slight convexity of the Earth surface, at the intersection of geographical coordinates equal to 49°34'23'' northern latitude and 19°31'49'' eastern longitude, an elevation equal to 1725 m. From an altitude of several hundred kilometres, or in other words at the continental and global scale, all the diversity of the Babia Góra range practically disappears.

Thus, the scale at which we perceive and describe the Earth determines the **dimensionality** of objects, that is, a measure that determines the number of dimensions of an object that are important for a given scale range. Of course, all objects that we deal with in the real world are three-dimensional, however, the way geography approaches the reality, focusing on the Earth's surface, in most cases excludes the third dimension – depth, thickness or height – from the consideration, resulting in the reduction of real three-dimensional objects to two-dimensional ones (also referred to as areal or 2D objects). However, certain types of objects can be reduced to one-dimensional objects (also known as linear or 1D objects) and zero-dimensional (also known as point or 0D objects). Most often, such a reduction is legitimate for various second level objects, but it may also apply to first level objects in a case when a discrete approach is used to describe them. A good example of a first level object that can be reduced to a point object is a city. A large city often extends over tens or even hundreds of kilometres – and in the future, such mega-cities will likely become more and more numerous. Nevertheless, taking into account the impact of the city as a hub for various functions at a global scale, even very large cities can be reduced to a point (Fig. 1.4).

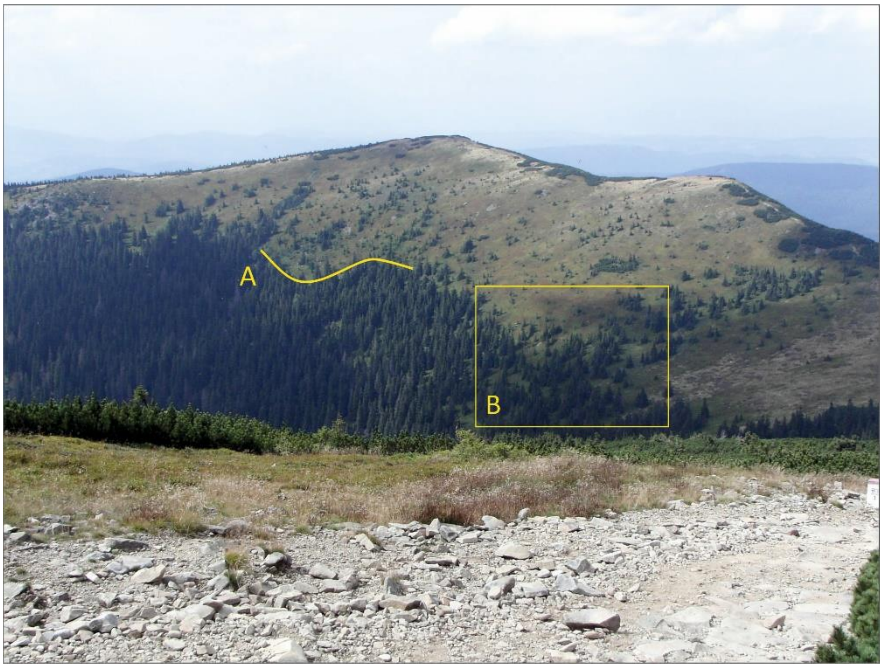


Fig. 1.3. A view of Cyl Mt. from the trail leading from the top of Babia Góra to the Brona Pass.

In the place marked with A, the upper timberline is relatively clear (shown with a yellow line). Within the part of the slope marked with B, the distribution of trees offers various interpretation possibilities as to the course of the upper timberline. The photo was taken in 2008, and it is worth adding that due to the anthropogenic origin of the upper timberline in this area, changes in the distribution of forests on the slopes of Cyl occur relatively quickly and currently the course of the upper timberline may be slightly different (photo credit: Jacek Kozak).

In her well-known work with an inspiring title *People manipulate objects, but cultivate fields*, Couclelis (1992) associates the scale rather with the previously presented discrete and continuous approaches instead of linking it with object dimensionality. Moreover, with respect to the discrete approach, she claims that the scale allows one to distinguish between the first- and second level objects.[7] Among four scale categories (A, B, C, D), scale A and B concern, according to Couclelis, second level objects, either smaller (A) or larger than human beings (B). Scale C is related to small parts of the Earth's surface, which can be observed from a certain view point, while scale D refers to areas that are not directly accessible to human perception. Discrete and continuous approaches can coexist at the C and D scales.

Conceptual modelling and ontologies

Various objects, processes and events, their properties and relations between them, with clear criteria for distinguishing them (defining criteria), constitute the foundation of conceptual modelling – the first step on the way from the real world to its digital representation (Peuquet, 1988; Longley et al., 2011), and the first stage of designing and building a database. This stage inevitably involves the simplification of reality in such a way that its infinite complexity can be reduced into a finite system of concepts relating to various phenomena and their properties that are relevant from a specific point of view. Constructing concepts and

[7] H. Couclelis refers to the paper of F. Zubin, presented in 1989 at *NCGIA Initiative 2 Specialist Meeting*, Santa Barbara.

their refinement, leading to dictionaries as unambiguous as possible, are referred to as the building of **ontologies** (Smith, Mark, 2001). Following Couclelis (2010), ontologies define the structure and basic properties of the simplified worlds that we are trying to represent. They also explain the meanings of the terms we use.

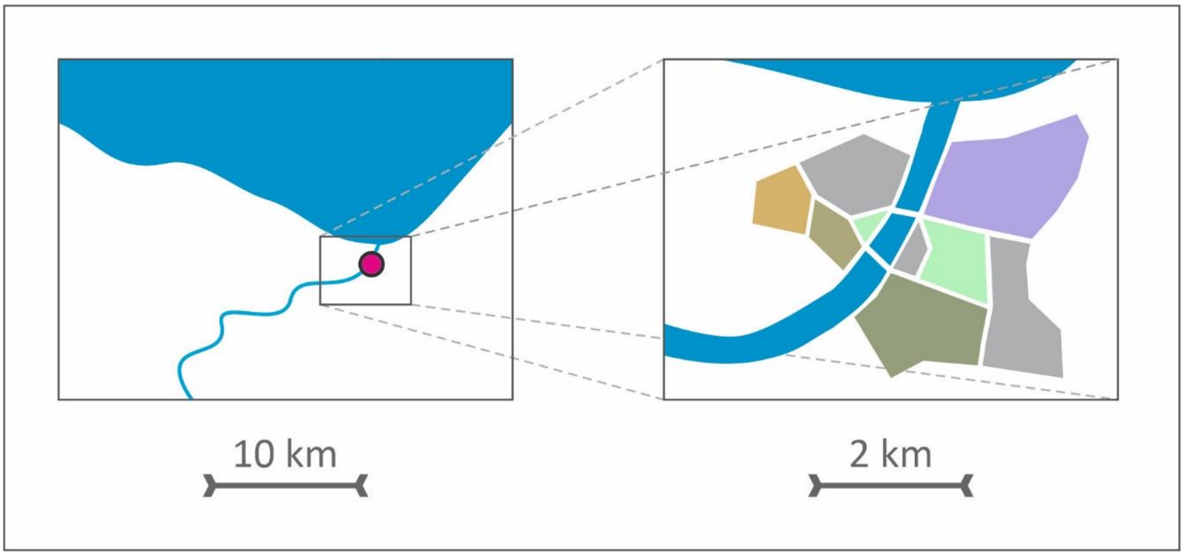


Fig. 1.4. Scale and object dimensionality.

A city as a point object (0D object) and a river as a linear object (1D object) – left; the city and the river as areal objects (2D objects) – right.

Defining geography, once again

In the introduction I have presented a preliminary definition of geography. At this point, after discussing the nature of the phenomena considered in geography, I wish to return to this definition, referring once again to what exactly geography focuses on. The subject of interest and research of geography are various phenomena: objects and their properties, processes and events, distributed in space and time, occurring on the Earth's surface and important for humans. Geography studies the spatial patterns of properties of first level objects using either a discrete or continuous approach, and the spatial arrangement of the second level objects and events, taking into account the discrete approach. I will return to this definition at the end of Chapter 6.

2. MEASUREMENTS AND GEOGRAPHIC DATA

Keywords: measurement, observation, sampling, attribute, time, space, data model, data, information, location, areal unit, spatial reference system

In the previous chapter, I have presented a general way of describing reality trying to organise the foundations of conceptual modelling with regard to phenomena that are important for geography (objects, events, processes). I have also pointed out the importance of two different approaches to describe geographical phenomena: discrete and continuous. In this chapter, I will examine how different methods of observation and measurement lead from the perception and conceptualisation of reality to formalized ways of representing and recording its various aspects, while at the same time taking into account the properties of objects and events as well as their location in space and, if necessary, also in time.

Properties, space and time

Sinton (1978) has stated that in order to turn an accidental observation of a phenomenon into useful information, it is necessary to provide a precise account of three aspects of this observation: the properties of the observed phenomenon, its location and the time of observation. In Sinton's original work, these three aspects are *theme*, *location* as well as *time*, later on, however, I will use in the same context A (*attribute*),[8] S (*space*) and T (*time*). Sinton noted that each phenomenon is related to a certain way of measuring and recording each of these three aspects. This also applies to the first- and second level objects, events and processes described in Chapter 1. From a geographic point of view, space is a key aspect of the three, since phenomena important to geography or other spatial sciences must be localized (or localizable) in space. Traditionally, in geography, the registration of phenomena over time was less important, although the dynamic changes in the environment at a global and local scales in the last few decades have demanded an increasingly accurate representation of time and changes over time (see also Chapter 6).

At this point, it is necessary to reflect on what actually the observation or measurement (registration) of phenomena and their properties is. **Observation** is a cognitive process carried out using the observer's senses, in particular the sense of sight, which allow one to determine some properties of the phenomenon (Bogen, 2017). Observation can be supported by various tools that facilitate its process. On the other hand, **measurement** is an empirical assessment of certain objective properties or relationships, an activity during which there is an interaction with a phenomenon in order to determine the properties of this phenomenon and to record them (Tal, 2017). The measurement can be an observation made by a human being with or without supporting tools, the result of which is recorded in some way, or the determination of selected properties of a phenomenon performed only with the use of automatic measuring devices, without the participation of an informed observer, but in a way that ensures some form of recording for the measurement results.

What we know about the water level in a river discussed in Chapter 1 results from the measurement of the properties of a certain object (river, stream) and the process taking place within that object (runoff). This measurement is carried out at a well-defined location – for instance, it may be the catchment closure mentioned before. In order to capture the dynamics of the runoff, the measurement must be carried out over time in such a way that the planned frequency of measurements (that is the distribution of individual measurements over time) allows for a realistic reproduction of the process. For example, during sudden

[8]In this book, two terms: *property* and *attribute* will be used interchangeably.

flooding, it may be necessary to receive water level readings at intervals of one minute or even less. This description of the water level measurements can also be reduced to the notation [x, y, H_1, H_2, …, H_n], where H_1, H_2, …, H_n are water levels measured in time from t_1 to t_n, and x, y are spatial coordinates of the place where the water level is measured. Therefore, it is worth paying attention to the fact that such a measurement (and also a simple observation) relates to three aspects (*A*, *T*, *S*) of any phenomenon, and that each of these three aspects is treated differently during a specific measurement. Water level measurement means that the attribute (water level) is **measured**, location in space is **constant**, by the specific and time-invariant positioning of the measuring device, while time is **controlled**, by scheduling the frequency of readings of the measuring device (Fig. 2.1).

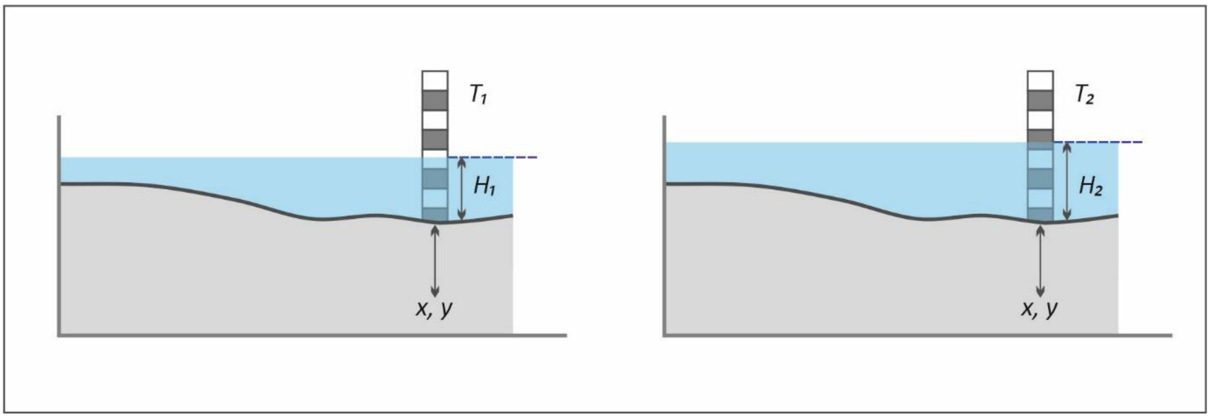


Fig. 2.1. The measurement of water level.

The measurement of the property (water level) is made at a fixed location, with a selected and controlled frequency.

How can one relate Sinton's concept to the relief of the catchment, mentioned earlier in Chapter 1? In this case, one can measure the elevation values in different places. If such a measurement is repeatedly carried out in many places, it will give us some idea about the topography of the catchment, and the more places there are, the more accurate it will be. The result of an individual measurement can be recorded in the same way as described above for the water level measurements, as [x_i, y_i, Z_i], where Z_i is an elevation value in a location with x_i, y_i coordinates, and *i* denotes subsequent measurements in other locations. As in the previous example regarding the water level in the river, an attribute – a certain property of the catchment (elevation) – is measured, but this time it is the location in space that is controlled (by appropriately planned selection of measurement sites), while time is fixed – measurements are made in a certain period, the precise definition of which (date, length of the period) is of little importance in this case, because unless the catchment is subject to very dynamic erosional and denudational processes, measurements can be carried out at different times, even years apart, without significantly affecting the output results.

Measurement and sampling

As in the case of water level and elevation measurements (and in fact in the case of any measurement whatsoever), the greater the frequency of measurements or the greater their density in space, the better we understand the properties of the phenomenon under study. In other words, the measurement of phenomena on the Earth's surface involves the **sampling** of space or time, which leads to the determination of attributes of the phenomena in a finite number of selected locations or moments, while ignoring the rest (that is omitting infinitely many other locations or moments). The finite number of selected locations or moments is a **sample**, selected from the entire set, that is the **population**.

In measurements leading to the possibly most perfect representations of real phenomena occurring on the Earth's surface, purposive or systematic sample schemes are very often used. A random sample, typical for example in sociological or biological research, is used much less frequently. If one wants to accurately reproduce the course of the river from its source to the mouth by measuring the location of the centre points along the river bed, one will not sample the measurement points randomly, but rather deliberately, in places characteristic for the course of this river. This could be, for example, a sharp bend in a river or a point where a tributary joins (Fig. 2.2). A similar rule of purposive sampling can be applied to surface topography measurements, during which the elevations of characteristic points or lines (e.g., peaks, ridges or valley axes) are usually determined. In these two cases, a sensible solution is also to use a systematic sample, that is measurements at points spaced at specific intervals along the river course or in the area which the surface topography is studied. These intervals should be small, taking into account the spatial variations of object geometry or properties (Fig. 2.2).

The discussed examples of water level or elevation measurements have two important features in common. First, each of the individual measurements (that is, in the discussed case, each individual recording of water level or elevation) is carried out in a specific location. More precisely, this is a point measurement. Next, selected properties are point-measurable ones – the water level can be measured at any selected point of the river bed, and the elevation can be measured at any selected point in the area of interest. Second, each measurement produces a numerical value. It may be the height of the water level in the river related to the accepted zero level marked on a gauge or absolute height in the case of topography and elevations. However, not every measurement has to have these two features. Measurements can refer to properties that are not point-measurable but, for example, relate to an area. Moreover, measurements (and in particular observations) can also give a qualitative output. These variants of measurements will be discussed below.

Areal units

The simplest example of an important property of the land surface that is not point-measurable and always refers to an area is population density. Population density is the number of people living in some area, per unit area. Even now, it is most often calculated during a census carried out in administrative units. Thus, it can be said that administrative units of various ranks (for example communes or poviats) for which data on the number of inhabitants are collected, are the basic areal units of the measurement that allows one to calculate population density. However, this assessment is possible not only for these specific areal units, but also for their various combinations. One can also make some mathematical assumptions about the spatial variation of population density over a certain area composed of a set of areal units to calculate the value of the population density at any location of this area. In this way, although population density cannot be measured at a point, it can be estimated at that point on the basis of measurements carried out in various areal units (see also Box 2.1).

This way of measuring population density and the use of areal units is common in geography as well as in other spatial disciplines. In essence, estimating density relies on determining the number of second level objects (e.g., people, trees, buildings) or events (e.g., accidents, crimes) occurring within freely designated first level objects, such as administrative units or geographical regions. However, this is not the only method to use areal units in measurements. For example, for any areal unit, one can specify how much of its area has a specific property. In this way, for instance, forest cover is determined as a percentage of the areal unit covered by forests. As with population density, there is no way to measure forest cover at any point – it is always determined in relation to some areal unit chosen for this purpose. Various possibilities, but also pitfalls of using areal units, including the problem of second level objects and events density, will be discussed in more detail in Chapter 7.

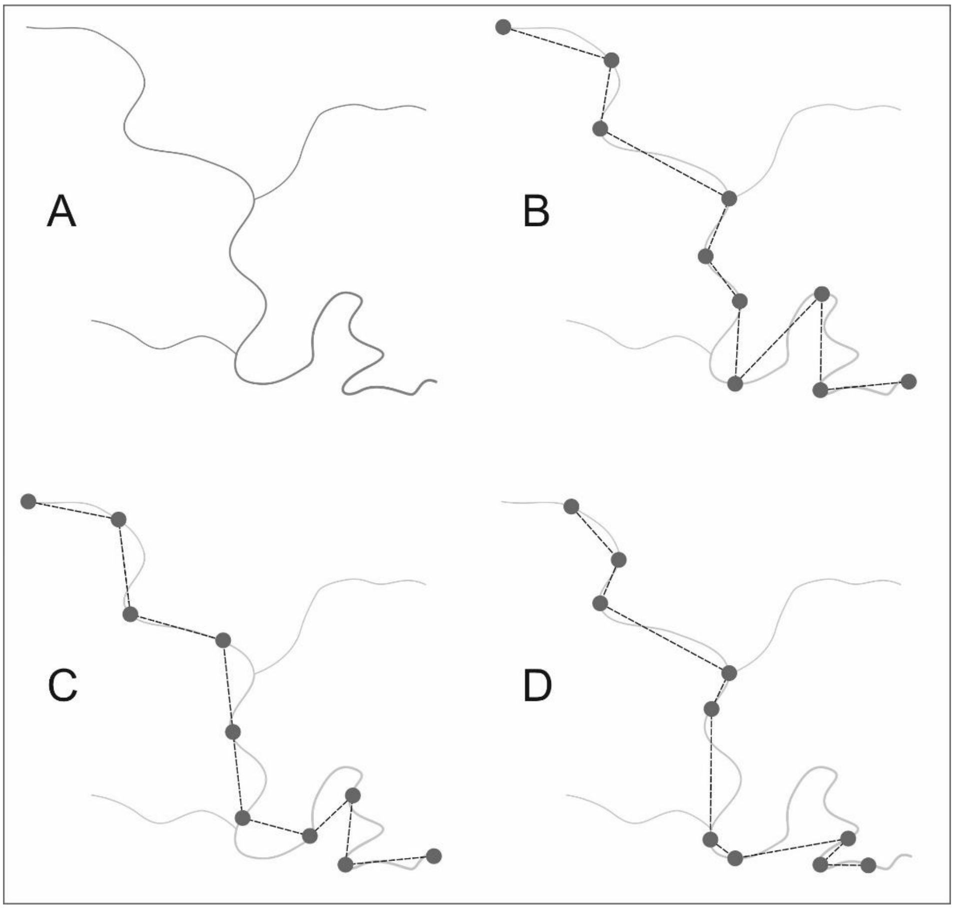


Fig. 2.2. River network sampling.

A – river network; B – measurements of the main river course with the purposive sample (measurement points located in characteristic points of the main river course); C – systematic sample (measurement points located along the main river course with regular intervals); D – random sample. Each sample has 10 points, and the purposive sample allows the best approximation of the main river course.

Box 2.1. *Gridded Population of the World*

Gridded Population of the World (GPW) is a database on the Earth's population, created and made accessible by the *Socioeconomic Data and Applications Center* (SEDAC), one of the centres of *Earth Observing System Data and Information System* (EOSDIS) network, managed by the *National Aeronautics and Space Administration* (NASA).[9] To build GPW datasets, first, population data from administrative units of all the countries of the world and the boundaries of administrative units were collected. For some countries, for example, such as Estonia or Poland, it is easy to obtain population data collected for a large number of small administrative units (in Poland these are about 2.5 thousand municipalities, in Estonia the number of administrative units used in GPW exceeds 4.5 thousand), but in various countries the value is usually much lower (e.g., for Angola 161 administrative units were used) and therefore they are much larger. Then, data from administrative units have been converted to a grid of identical areal units. These areal units can be

[9] Accessible at: https://sedac.ciesin.columbia.edu/data/collection/gpw-v4, accessed March 2020.

roughly defined as squares, with GPW employing several grids with square sizes ranging from 1 x 1 km to about 100 x 100 km. In this way, a reliable population distribution in comparably uniform and regular areal units was received for the entire land area of the Earth.

Measurement scales

Although the measurement outputs most often include numerical values, the attributes of phenomena can be described both quantitatively and qualitatively, using various **measurement scales**. A qualitative description, in other words a qualitative measurement scale or a nominal scale, are names referring to the properties of phenomena – events or objects. For example, in the case of a small piece of land, the name *pasture* describes its specific form of land use, the name *Kowalski* – the surname of its owner, and the name *fertile* describes its soil properties important from the point of view of how this piece of land is used. Quantitative measurement scales are numerical values that characterize selected properties of objects or events. A pasture may be located at an altitude of 525 m above sea level, the average annual temperature measured at this location in the period 2005-2010 may reach 7.2°C, and the number of sheep using this pasture per year may equal 37.

The transition from a qualitative to a quantitative measurement scale is often vague. The term *fertile* may be supplemented with the terms *moderately fertile* and *poor*, which, although apparently qualitative, can also be used to organize information about different parts of the land (for example, different pastures) in terms of their increasing agricultural suitability and availability of fodder for animals. In this way, the qualitative measurement scale takes on some features of a quantitative scale in which the ascending or descending ordering of values is an obvious possibility.

Quantitative measurement scales can vary widely. Ordinal scales are similar to the qualitative measurement scale described above allowing for the ordering of phenomena, using numerical values that have only an ordinal character.[10] The terms *fertile*, *moderately fertile* and *poor* may be replaced in such a measurement scale by, for example, the values 1, 2 and 3. Quantitative measurement scales can use both integers and rational numbers, positive or negative, from bounded intervals or not. The choice of the numerical range depends on many factors – these include, among others, the nature of the phenomenon and its properties, measurement accuracy, adopted units and methods of measurement. For example, the forest cover of an area is described by non-negative rational numbers from the interval <0, 1> (or <0, 100%>). Slope gradients measured in degrees are non-negative values from the interval <0, 90>, frequently rounded to integers. Cloudiness measured at meteorological stations is expressed with integers from the interval <0, 8>, where 0 denotes no clouds, and 8 – full cloud cover. Absolute heights are measured in metres, in relation to the sea level. Typically, its values are given as integers, although decimal values are also used for accurate measurements. Due to the topography of the Earth's land surface, the absolute heights of some areas (depressions) are expressed using negative numbers. Obviously, a completely different range of values will be used if the unit of measurement are feet instead of metres as, for instance, in the United Kingdom (Box 2.2).

[10] Some researchers do not include the nominal and ordinal scales to measurement scales, reserving the concept of a *measurement scale* solely to quantitative phenomena (Tal 2017).

Box 2.2. *Munros*.

In Scotland, where only two peaks exceed 1,300 metres above sea level (Ben Nevis, 1,345 m a.s.l., Ben Macdui, 1,309 m a.s.l.), that is less than many peaks in the Beskidy Mts. (e.g. Polica, 1,369 m a.s.l., Romanka, 1,366 m a.s.l.), all peaks exceeding 3,000 feet (914 m) are proudly called *Munros*. *Munros* are peaks registered on the list of the *Scottish Mountaineering Club*, with its first version given at the end of the 19th century by Sir Hugh Munro. The current list of *Munros* names 282 peaks. In the UK outside Scotland, and also in Ireland there are 34 peaks, which fulfil the criteria of *Munros* – these are so-called *Furths*. Of course, for climbers much more prestigious than *Munros* is the Crown of the Himalaya and the Karakoram – a list of 14 of the Earth's peaks over 8,000 m high, 10 of which are located in the Himalayas and four in the Karakoram. If round heights in feet were taken as a reference, the list of "thousanders" would depend on whether it is 26, 27, 28, or 29,000 feet. Above 29,000 feet is only Mount Everest, with only three peaks exceeding 28,000 feet. In contrast, there are already 18 peaks higher than 26,000 feet, four more than listed by the Crown of the Himalaya and the Karakoram.[11]

Geographic data and geographic information

Chapter 1 presents a general description of various types of phenomena (objects, events, processes and their properties) corresponding to the stage of conceptual modelling – the initial stage when creating a digital representation of reality. From what has been already discussed in Chapter 2 it is clear that the way we observe, measure and record the properties of various geographical phenomena determines the schemes and principles of any detailed description of these phenomena, that is the adoption of an appropriate **data model**. An example of a data model can be a geo-atom, proposed in a study devoted to basic concepts of representing geographical phenomena (Goodchild et al., 2007). A geo-atom is the simplest formal notation linking the properties of a phenomenon and its location, expressed as follows:

$$x, Z, z(x) \qquad [2.1]$$

where *x* denotes location in space-time, *Z* is a specific property, and *z(x)* is its measured (observed) value. Another example of a data model may be the previously discussed water level record at the water gauge station:

$$x, y, t_i, H, h_i \qquad [2.2]$$

where a specific property (water level, *H*) is measured at a location with coordinates *x*, *y*, in time t_i, and a measured value of h_i. This model takes into account both the value of the specific measured property, the location of the measurement site and the moment at which the measurement is done. Contrary to the notation proposed by Goodchild et al. (2007), location in space is determined by a pair of coordinates, while location in time with just one.

If data model is a formalized, schematic notation of the results of the measurements or observations of selected properties of the real world, then **data** is nothing more than text, numbers or other symbols representing specific properties of real world phenomena, neutral and without context (Longley et al., 2011), recorded and stored according to a certain data model. If a particular data model takes into account the location of the phenomenon to which the data relate, and the phenomenon occurs somewhere on Earth, it

[11] Accessible at https://www.smc.org.uk/hills/hill-lists, accessed: March 2020 and https://en.wikipedia.org/wiki/List_of_mountains_by_elevation, accessed: May 2021.

is a **geographic data model**. **Information** differs from data: it refers to data that have been processed for a specific purpose, data that attract someone's attention or someone is aware of (Longley et al., 2011). If data are localized, that is data are geographic, then the information based on these data is also localized and therefore is referred to as **geographic information**.

In order to accurately distinguish the meaning of these two terms, data and information, imagine that a certain measurement system has been collecting data on the water level in the closure of a certain catchment every hour, for ten years, on a hard disk. Thus, there are already tens of thousands of individual data records (measurement results written and stored in a specific way). Suppose someone is interested in the flood risk assessment in the catchment and wants to find out how often the water level in the river exceeds 2 m. She may check the data and use simple operations to extract the desired **information**: how frequently the water level exceeds 2 m with reference to the total number of measurements. Data processed for a specific purpose and in a specific context become information. However, you can also imagine that someone accidentally threw the disk away with the saved data into a rubbish bin, and then the rubbish went to a landfill. The hard disk still keeps the collected data – but the chance that someone will extract any useful information from this data decreases with each passing day and with each successive batch of garbage delivered to the landfill, burying the unlucky disk.

Recently, the term *geodata* is used instead of *geographic data*. By analogy, *geoinformation* is sometimes used instead of *geographic information*. Additionally, *geospatial* has become a frequently used synonym for the term *geographic*.[12] To some extent the term *spatial* is even more handy than *geographic* because it emphasizes the spatial location without implying that data or information must have geographic relevance, and such a misinterpretation often occurs when using the terms *geographic data* or *geographic information*. However, tradition and a sense of loyalty to my own discipline make me choose the terms *geographic data* or *geographic information* whenever I refer to localized data or information. However, I would like to emphasize once again that the term *geographic* only means that **the data or information refer to a certain location on the Earth**. The terms *geographic data* and *geographic information* should not be interpreted therefore in the context of their potential importance for geography as a scientific discipline – for example, an archaeologist, biologist or economist may be also interested in geographic data. On the other hand, data and information taking into account the location of phenomena almost always arouse the interest of geographers. Space is so important to geography that this science most often uses location-based data, that is geographic data.

Spatial reference systems

Until now, measuring and expressing a location has not been discussed, but at this point it is worth considering in more detail what actually recording a location in space means and here in relation to the Earth. How do we measure a location of any phenomenon on the Earth's surface, georeferencing in this way data or information? The simplest answer to this question is that to determine a location, one of many known Earth spatial reference systems can be used. The in-depth discussion of spatial reference systems is far beyond the scope of this book, therefore below only the most important information is provided. The interested reader should refer to numerous basic or specialized studies in the field of geodesy and cartography, as spatial reference systems are the domain of these two scientific disciplines. For hundreds or even thousands of years cartography, and later geodesy, have been perfecting the methods and tools that allow one to determine location on the surface of the Earth with an incredibly high degree of accuracy.

[12] Some researchers think that *geographic* and *spatial* combined with either data or information do not have exactly the same meaning. Data or information can be localized, yet with no reference to the Earth (e.g., data collected by satellites observing the surface of Mars) – in such a case *spatial* is the correct term.

The Earth is a rather complicated three-dimensional solid, which, however, is relatively accurately approximated by the model of a sphere with a certain radius, and even more precisely approximated by the model of an ellipsoid, with defined short and long axes, and polar flattening resulting from the difference in their length. It is worth noting, however, that in the small scales used to represent the entire Earth, the difference between the ellipsoid and the sphere is practically invisible (Fig. 2.3).

For the spherical model of the Earth, an unambiguous spatial reference system is a geographic reference system, with geographic coordinates (Fig. 2.4), longitude (λ) and latitude (φ) – used to define locations already in ancient times (Pasławski, 2006). In the case of an ellipsoidal model, the equivalent coordinates are geodetic latitude and longitude. For both the sphere and the ellipsoid models, one can apply various transformations (that is map projections) that convert geographic or geodetic coordinates into coordinates of a Cartesian coordinate system (planar coordinates), much more practical than geographic or geodetic coordinates, if geographic data is to be presented on a flat screen or on a sheet of paper. It is extremely important to remember that regardless of the mathematical method of this conversion, it is impossible to transfer locations from the curved surface of the Earth to a plane without distorting the distances and directions between these locations. Therefore, the conversion from geographic or geodetic coordinates to a Cartesian coordinate system always involves the distortion of angles, areas or distances. However, one can choose a projection and a coordinate system to minimize certain types of distortion or to minimize distortions for a specific area. This important aspect of cartography and geodesy will be here omitted, and the interested reader should refer to textbooks on these two disciplines.

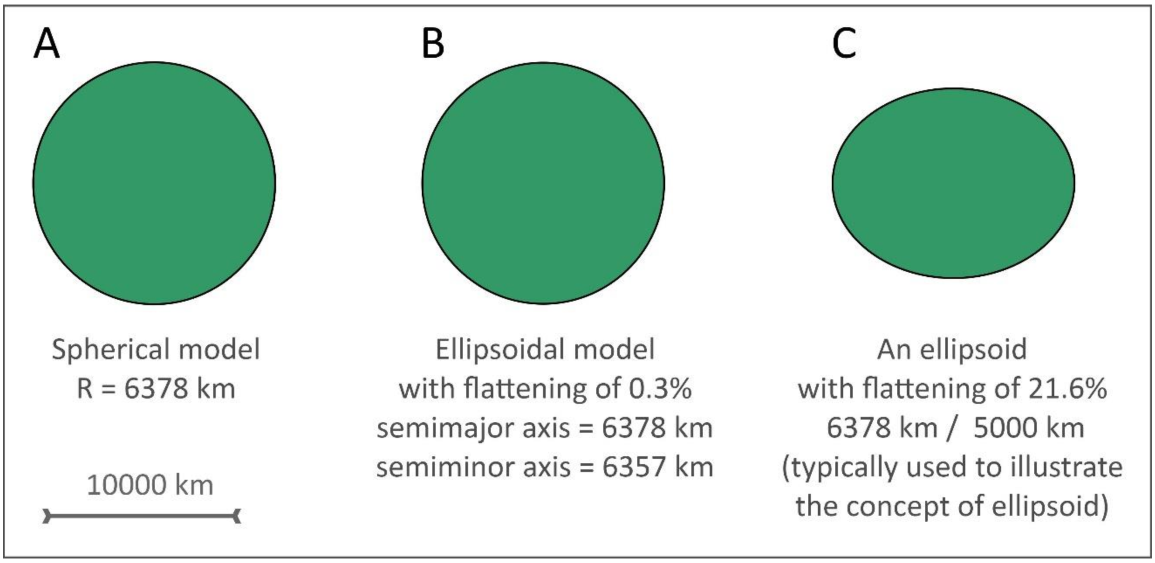


Fig. 2.3. Models of the Earth.

Left – the sphere model with a radius of 6378 km, centre – the ellipsoid model with a flattening of 0.3%. At this scale, it is impossible to note a visual difference between these two models. On the right hand side, for comparison, the ellipsoid model with a flattening above 20%, clearly visible, but is several dozen times greater than the actual flattening of the Earth. Models of this type are often used to illustrate the concept of the Earth's ellipsoid.

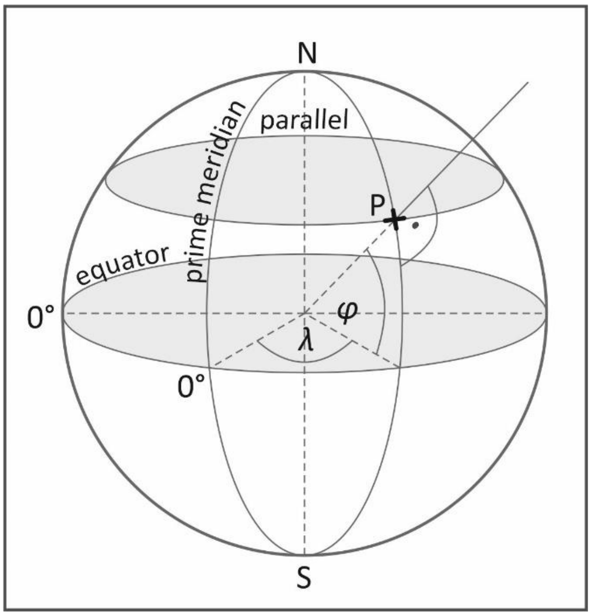


Fig. 2.4. Geographical coordinates.

For any point P other than one of the Earth's poles, longitude (λ) is the angle between the plane of the prime meridian (the prime meridian is defined using a certain convention) and the plane of the meridian passing through P, while latitude (φ) is the angle formed by the radius of a sphere passing through P with the plane of the equator. The longitude of the North Pole (N) and the South Pole (S) is undefined.

Adapted from: Pasławski, 2006

Countries of the world define their own spatial reference systems in various ways. In Poland, the national spatial reference system defines the PL-1992 coordinate system, which is used for geographic data developed in spatial scales between 1 : 10 000 and 1 : 250 000 (*Rozporządzenie Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych*), in this way PL-1992 is most frequently applied to geographic data elaborated in Poland. The PL-1992 coordinate system uses GRS 80 ellipsoid (GRS stands for *Geodetic Reference System*), and the points of the ellipsoid surface are projected to the plane using the Gauss-Krüger projection. In the Gauss-Krüger projection, the surface of the ellipsoid is projected onto the surface of a cylinder wrapped around the ellipsoid and adjacent to a selected meridian, known as the central meridian of the projection. In this case, we are talking about a cylindrical transverse projection, often taken to be equivalent to the well-known Transverse Mercator projection. In a cylindrical transverse projection, the natural axes of the Cartesian coordinate system are the equator (OX axis) and the central meridian (OY axis). The location of a point projected from the surface of the ellipsoid onto the plane is determined by its distance from the equator, measured in parallel to the central meridian, in the south-north direction (*y* coordinate) and its distance from the central meridian measured in parallel to the equator and in a west-east direction (*x* coordinate).[13] These distances are specified in an accepted unit of length, for example metres. It is worth adding that, to minimise distortions, transverse cylindrical projections are most often used for relatively narrow belts located along a central meridian. In order to extend the projection to large fragments of the Earth's surface, a zonal system is adopted with pre-defined central meridians and a specific zone width. In this way, any specific location on Earth is mapped in one of the projection zones (see Box 2.3). The PL-1992 coordinate system includes also a vertical datum to measure heights or elevations – the mean sea level in a strictly defined location. In the case of PL-1992, the chosen location is the mareograph in Amsterdam which

[13] In Poland official geodetic and cartographic sources and materials use the inverse notation of both axes: coordinate *y* refers to the west – east direction and coordinate *x* refers to the south – north direction, as this notation has been traditionally used in geodesy. In this textbook, however, I have chosen the notation used in GIS.

allows one to determine the mean sea level for the North Sea (*Rozporządzenie Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych*).

In principle, the information given above sufficiently defines the PL-1992 coordinate system, however, to gain its full overview, some modifications applied to this system for practical reasons should be discussed. First, the Gauss-Krüger projection is secant and not tangent, that is, the cylinder does not adhere precisely to the surface of the ellipsoid, but slightly cuts under its surface. The image of the central meridian in this projection is therefore shortened in relation to the original by 0.07%. Such a projection structure reduces the distortions over the territory of Poland compared to the projection in which the cylinder is tangent at the central meridian. Second, the origin of the coordinate system was shifted 5,300,000 meters north (so-called false northing) and 500,000 meters west (so-called false easting, Fig. 2.5). This shift of the origin of the coordinate system[14] allows one to express the coordinates of any location in Poland using six digits, with an accuracy of 1 m. Poland recalls a fairly regular trapezoid, 700 km x 700 km (that is, 700,000 m x 700,000 m), so if the origin of the coordinate system is slightly southwest of the south-west corner (Turoszów region), the furthest distances of extreme points in the north-west, north, north-east, east and south-east, both east-west and north-south will be expressed with values from the range (0, 999999). Such a modification simplifies the input and storage of coordinates that are always non-negative and consist of six digits. Note that if the origin of the system remained at the intersection of the equator and the central meridian, then the y coordinates would have to be expressed with seven digits (the distance of any location in Poland from the equator exceeds 5,000,000 m), while the x coordinates would be negative for any location to the west, and positive to the east of the central meridian.

Box 2.3. UTM coordinate system

The transverse cylindrical projection in the zonal variant is applied in the widely used global coordinate system called Universal Transverse Mercator (UTM). In this system, the world is divided into meridional zones 6 degrees wide, each with its own central meridian. Poland is covered by zones 33 and 34, although the extreme eastern parts of the country near Hrubieszów are within the boundaries of zone 35 (its western edge has the longitude 24°E, crossing Poland slightly east of Hrubieszów). The central meridians of zones 33 and 34 have longitudes of 15°E and 21°E.

We are not always able to give the geographic or Cartesian coordinates of the place where we are or which we are interested in for various reasons. For example, in the event of a road accident and the need to call for help, we do not need to enter the coordinates – usually it is enough to provide the address (for example, the accident happened at Mogilska street in Kraków, near number 78). An address is an example of indirectly locating phenomena on the Earth surface – either the person concerned simply knows where to look for such an address (for example, she knows the city and entering the address is sufficient to find the place she is looking for), or she can convert the address into the coordinates of any selected spatial reference system and use these coordinates to locate the required place. The procedure of assigning coordinates (for example, geodetic coordinates or Cartesian coordinates) to addresses, but also to characteristic places or objects which can be used to indirectly locate a specific place is referred to as **geocoding**.

[14] The origin of the PL-1992 coordinate system is located in Walchsee in Austria, close to the border with Germany, and approximately 20 km north of the very well known Alpine ski resort, Kitzbühel.

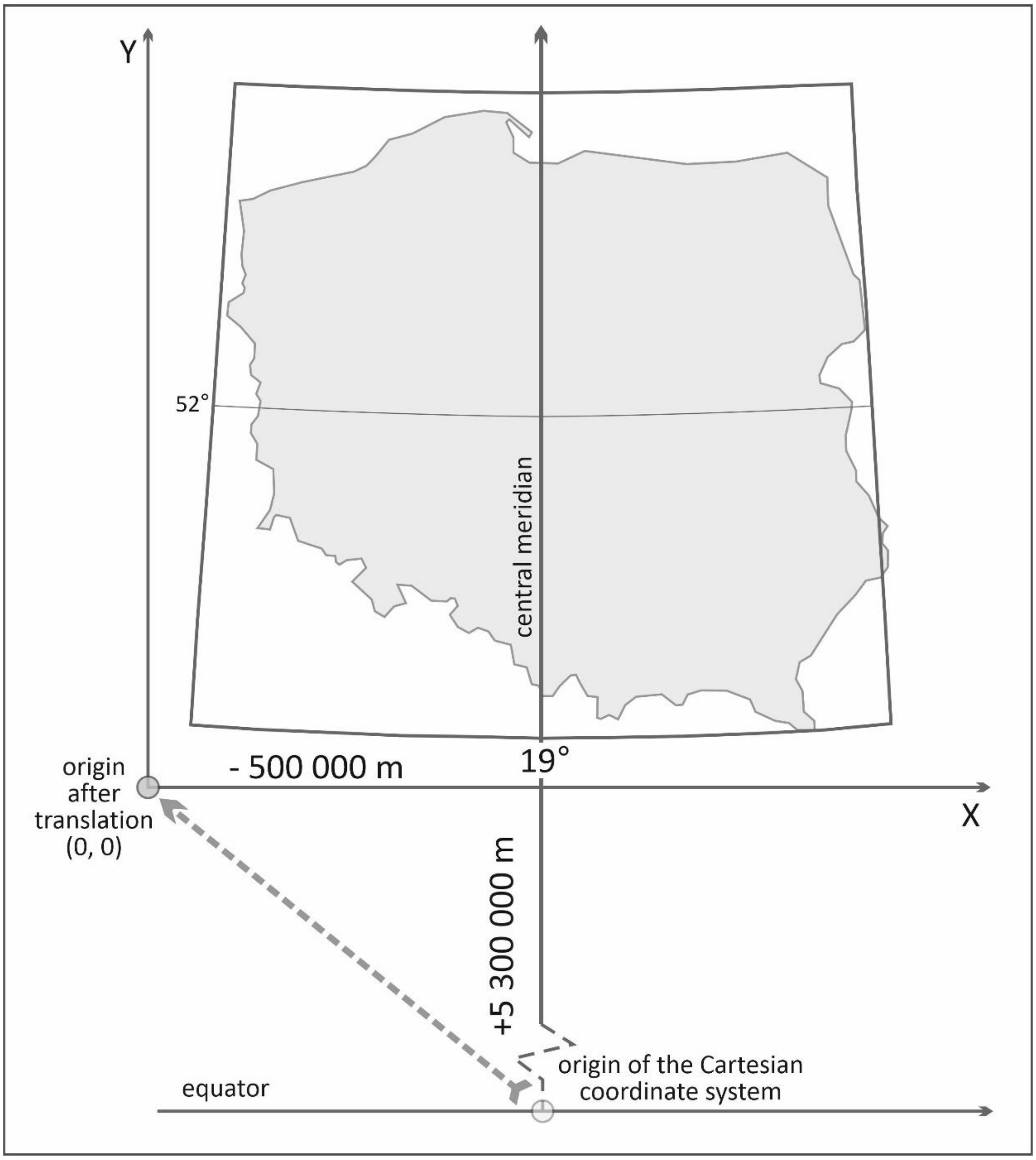


Fig. 2.5. PL-1992 coordinate system.

The origin of the system is the intersection of the central meridian and the equator. For practical reasons, it is shifted by 500,000 m to the west and 5,300,000 m to the north, which allows one to express the coordinates of any location in Poland using 6 digits, with an accuracy of 1 m (there are no negative coordinates and no coordinates greater than 1,000,000 m). The figure shows the axes of the PL-1992 system as in the notation used in GIS.

Adapted from: Kadaj, 2002

Postal addresses are one of the simple systems of locating phenomena on the Earth surface indirectly, practical and easy to grasp. However, postal addresses have also disadvantages – they allow one to easily and accurately locate phenomena in urbanized areas, but fail outside them – locating anything in the desert with an address does not make sense, even if it is an area as small as the Błędowska Desert in southern Poland. There are also chaotically expanding urban areas in developing countries where the address system either does not work at all or only to a limited degree. This does not mean, however, that indirect systems are useless outside well-developed areas – an interesting universal system using indirect spatial referencing is *what3words*, proposed some time ago (Box 2.4).

Box 2.4. *what3words* system

what3words[15] is a system of dividing the entire Earth surface into squares 3 x 3 m, with the square position clearly defined in the geodetic coordinate system. Each square of the global grid is identified by a unique string of three words (hence the name of the system) that are easy to remember. For example, the statue of Adam Mickiewicz in the Market Square in Kraków lies in a square *pieczarki.żelazko.lotka* (*mushrooms.iron.quill*). In the intention of its creators, *what3words* should replace postal addresses, because it works also in completely uninhabited areas, that for obvious reasons have no address system. Knowing the three words of a given square, we can unambiguously locate it, because these three words are *tied* to the exact geodetic longitude and latitude of the square's centre.

Although I have used the terms *location* and *place* basically interchangeably, it is worth noting that in geography, the concept of location differs from the concept of place. Although we colloquially say that if something is located, it means that something is *in some place*, *place* in geography (especially in human geography) has a much wider meaning – it is defined by the subjective relations of a conscious mind, for instance of a specific person, with a certain part of the world the person is perceiving. In other words, *place* is a location full with meanings that cannot be reduced simply to the coordinates of any spatial reference system. These meanings may result, for example, from someone's attachment to a specific place, familiarity with a place, or aesthetic impressions about it. In this book, I will not deal with this meaning of the term *place* – so wherever this term appears, it should be treated as a colloquial synonym for the term *location*.

[15] Accessible at https://what3words.com, accessed: March 2020.

3. GEOGRAPHIC DATA MODELS

Keywords: raster geographic data model (raster), cell (pixel), spatial resolution, vector geographic data model, topology in vector models, layers, relational data model, geographic database (geodatabase)

Now that the general definition of a geographic data model has been presented (Chapter 2), let us look at how this general definition is practically applied to different phenomena, for example first- or second level objects and their properties. I will start with an example: imagine that you are interested in identifying and recording (that is measuring) the land cover[16] of a certain area, and let us assume that this area, 10 by 10 km, is located somewhere in northern Poland, in the Mazury district, with its characteristic landscape of lakes and forests. The land cover entries for this area are geographic data, and to receive them one can use two different methods.

The first method (Fig. 3.1; the figure presents a small section of the hypothetical study area) relies on defining the correct land cover type for a large number of locations (points) within the analysed area. As long as these locations are distributed sufficiently densely, this method will give us a fairly accurate representation of the land cover of the surveyed area. Then land cover category for these locations can be recorded using, for example, the notation of the geo-atom (Goodchild et al., 2007) presented in Chapter 2.

$$\begin{aligned}&\textit{location of point 1, land cover type,}\\&\textit{location of point 2, land cover type,}\\&\ldots\\&\textit{location of point n, land cover type}\end{aligned} \qquad [3.1]$$

The other method (Fig. 3.1) takes advantage of the fact that homogeneous land cover types usually occupy certain areas and form a mosaic of irregularly shaped geometric figures. In the case under consideration these are patches of forest and lakes. We can try to define their boundaries by separating different land cover types, and then to record the location of these boundaries as accurately as possible, sampling a sufficiently large number of points along the boundaries between the land cover types. For each homogeneous land cover patch, its boundary will form a closed perimeter consisting of a certain number of points. In this way, the land cover of the studied area can be written as follows:

$$\begin{aligned}&\textit{boundary of patch 1 (points } p_{11}, p_{12}, \ldots, p_{1j}\textit{), land cover type,}\\&\textit{boundary of patch 2 (points } p_{21}, p_{22}, \ldots, p_{2k}\textit{), land cover type,}\\&\ldots\\&\textit{boundary of patch n (points } p_{n1}, p_{n2}, \ldots, p_{nm}\textit{), land cover type}\end{aligned} \qquad [3.2]$$

Considering the first method, we are interested in the properties of a section of the Earth's surface (i.e. a first level object) in specific locations – although these locations themselves do not have any special meaning. In the case of the second method, we are interested in what types of discrete objects result from the spatial patterns of a certain property and what their spatial distribution is. This first approach to the data model is known as the field-based data model, the other – as the object-based data model. Although the similarity of these terms to the previously discussed continuous and discrete approaches suggests that the field-based model is associated with continuous phenomena and the object-based model is related to discrete

[16] A land cover type results from the physico-chemical properties of the material covering the Earth's surface (e.g., grass, ice, rock). In defining land cover, similar terms as in defining land use are often used. Land use, however, describes how humans use the Earth's surface, contributing to a variety of land covers.

phenomena, this is not always the case – the relationships between these approaches and models are slightly more complicated (Galton, 2001).

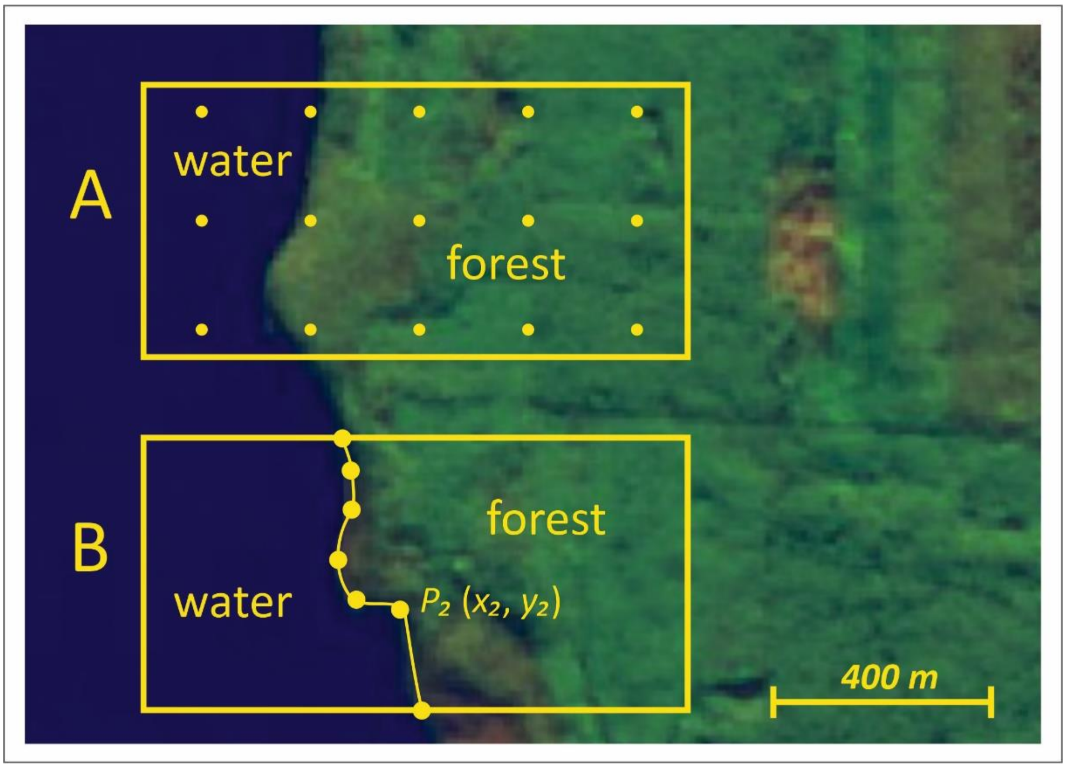


Fig. 3.1. Measurement methods for a section of the study area, 0.5 x 1 km.

A – points of a regular grid, for which land cover has been defined during measurements; B – points delimiting the boundary between a lake and a forest patch, for which the coordinates (e.g., x_2, y_2) have been recorded during the measurement.

The source for the background image comes from Sentinel Hub Playground, https://apps.sentinel-hub.com/sentinel-playground.

The first method seems to be quite time-consuming, regardless of the land cover patterns in the studied area. If one wants to obtain a relatively accurate representation of the land cover of this area, it will require measuring at least several dozen points per square kilometre, with a total of several thousand individual measurements for the entire analysed area. The more accurate a representation expected, the more measurements are needed. Even with a relatively small variation in the land cover, measurements are needed everywhere, because in this case the measurement does not only allow one to assess a land cover type in a given location, but also allows one to record it in some form. So if a measurement is not carried out at a certain point, then there will be no record for that site, which means no land cover data.

The second method can be very time-consuming when the land cover pattern displays a large spatial complexity – in this case it will be necessary to determine and record the location of many boundaries between the types of land cover. On the other hand, if the land cover of the study area consists of large homogeneous patches, the number of boundaries decreases, and the method can result in significant time savings when compared to the first. However, the attentive reader will notice something strange at this point: how do we know whether the method will be time consuming or not, so how do we know *a priori* what the land cover pattern is when we have yet to study it? In order to solve this problem, it should be noted again that the measurement is both [1] determining the land cover in a given location and [2] recording the measurement result. The first can be done by simply walking around the study area, both in the case of the first and the second method, because we can see exactly what land cover patterns we are dealing with. This means that a simple observation is the introduction to the measurement, as it allows us to decide which method will be more convenient or effective to record the phenomenon and its properties. In the case of the first method, the recording of the measurement result requires a large number of points, regardless of the complexity of the land cover pattern, and in the case of the second method, the recording density may be adjusted to reflect the degree of the complexity of the land cover pattern. Therefore a preliminary assessment is sufficient to make a rational decision on the choice of the first or the second method.

Taking into account the scheme proposed by Sinton (1978), with the first method attribute (land cover) is measured, time is fixed and location is controlled: thus the attribute is measured for pre-defined locations (measurement points). In the second method, attribute (land cover) is controlled, time is fixed, and position (location) of points with a specific combination of attributes is measured (points that define the boundary between two land cover types, for example forest and water).

The first method is most often used for regularly spaced locations (points), forming rectangular (usually square) grids with a different mesh size. This type of measurement can be automated – for land cover, the measurement is made using satellite images that are acquired when a sensor placed on board a satellite measures and records, in a regular measurement grid, the amount of electromagnetic radiation reflected from the Earth's surface in various bands of the electromagnetic spectrum that is directly dependent on the land cover itself (more on this in Chapter 5). The second method can be used in field research. For example, while walking along a boundary between different land cover types, one can regularly record the position of selected points on that boundary with a satellite navigation receiver. Both methods have been and are used in practice in land cover research. Good examples are international land cover and land use projects in Europe (box 3.1).

The first method, as long as a regular rectangular grid of measurement points is used, allows one to record the measurements in the form of a **raster geographic data model**. The other method records the measurement using a **vector geographic data model**. After the conceptual modelling stage, the selection and definition of a data model is the next step on the way from reality to its digital representation. As mentioned in Chapter 1, this step is known as logical modelling (Longley et al., 2011).

Box 3.1. European land use and land cover databases.

In Europe, various approaches to the study of land use and land cover are applied. The *Corine Land Cover* (CLC) project[17] aims to build comprehensive land use and land cover databases for most European countries. The databases store the boundaries of relatively homogenous land use and land cover patches. CLC started in the 1990s, and since then, several editions of the CLC database have been completed (the most recent being the fifth, in 2018). Therefore CLC allows one to study land use and land cover changes in Europe. Another approach is used in the European *Land Use/Cover Area frame Survey* (LUCAS) project.[18] In this project, a regular grid of points with 2 km spacing was defined in Europe. For these points, every three years (since 2006, earlier work being a pilot study) land use and land cover is assessed. This method allows one to

[17] Accessible at https://land.copernicus.eu/pan-european/corine-land-cover, accessed: April 2020.
[18] Accessible at https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey#The_LUCAS_survey, accessed: April 2020.

receive an overview of the state of land use and land cover at the time of a given survey, as well as changes in between the subsequent surveys.

Raster geographic data model

Ideally, the raster geographic data model (I will also use the abbreviated form of *raster model*) is based on measurements carried out in a regular grid of measurement points. At each such point, one can define and write down a property of a phenomenon or, for example, the mere fact of its occurrence (or its absence). Because a distribution of points forming a regular square grid[19] is in general very dense, it is basically a matter of convention as to whether the measurement is done for example at the centre of a square and represents the whole square, or whether the measurement is carried out for an entire square and the implicit measurement location is the geometric centre of the square. The smaller the squares, the less significant this distinction is.

The raster model is a case of tessellation, that is a disjoint and complete division of a plane into geometric figures, in which each point of the plane belongs to one, and only one, figure. The raster model is a square tessellation in which each figure that makes up the tessellation is an identical square. These squares, arranged in rows and columns, form a raster matrix and are referred to as cells or pixels.[20] In a raster matrix, the location of each cell is uniquely determined by specifying the row and the column of the matrix in which the cell is located. Since each cell of the raster matrix is also assigned a specific value of the selected property, this value together with the location of the cell all create a complete geographic data record (Fig. 3.2). The effectiveness of the raster geographic data model is determined by the fact that the notation of a raster matrix can be radically simplified: it is enough to adopt a specific order of cells in the raster matrix to reduce the notation only to the values of the selected property. In this case, the location of a cell in a particular row and column of the raster matrix is determined by the position that the value for this cell occupies in the value string representing the entire matrix. For example, in a raster matrix with a size of 6 columns by 5 rows, the 16th value will fall in the third row and the fourth column, provided the order of writing is left to right and top down (Fig. 3.2).

Note that the notation of the raster matrix shown in Fig. 3.2 can be simplified even more – it is enough to adopt a convention in which instead of values written in matrix cells, a pair of numbers is used, in which the first number specifies the value in the cell, and the second – the number of cells in which the value appears one after another given the accepted writing order. If the same values are found frequently in adjacent cells, the number of characters that can be used to write raster matrix values is fewer than the number of matrix cells (and therefore fewer than the number of original values). For a record such as that shown in Fig. 3.2, the sequence of numbers:

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

can be read as: value 5 once, 6 once, 4 once, 3 four times … 6 twice, 4 once. This notation is one of the possible lossless compressions often used to effectively write raster matrices, called run-length encoding (RLE). In the case of the matrix shown in Figure 3.2, the entire matrix can be written using 28 characters instead of 30 in the non-compressed, original form.

[19] As mentioned earlier, raster models allow also for rectangular grids. However, as the square grids are most common in raster models, this chapter will entirely focus on this type of raster models.

[20] These two terms will be used interchangeably: when image data will be considered, I will use the term *pixel*, while for other raster model data I will use the term *cell*.

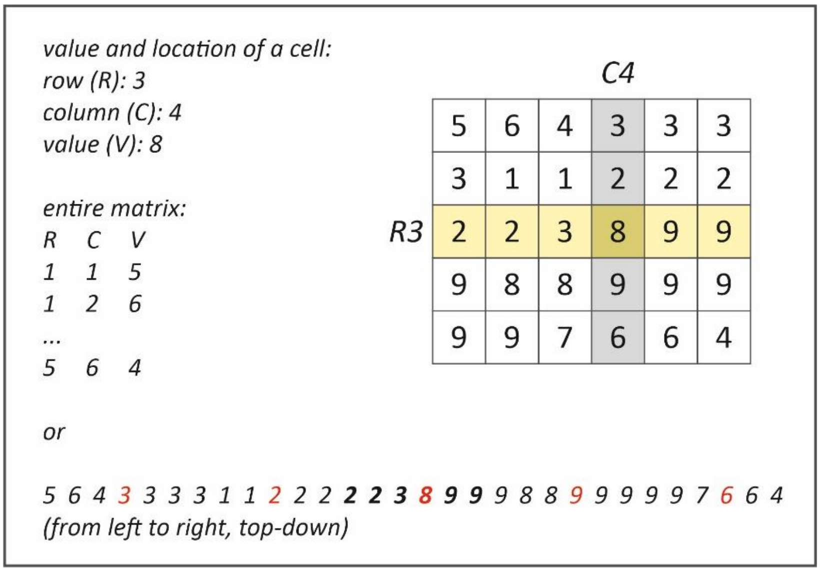


Fig. 3.2. A raster matrix.

The raster matrix is made up of five rows and six columns, and 30 cells in total with the values of a certain property. A single cell is located if its row and column numbers are given, yet with a defined sequence of recorded values, row and column numbers can be skipped, leaving only the values of the property (shown at the bottom of the figure). Row 3 of the matrix is shown in bold, column 4 – with red characters, and the cell in the third row and the fourth column (value 8) is bold and red.

Locating a raster cell using a row and column system allows one to locate a cell only within a given raster matrix, but it does not allow one to locate a raster cell on the surface of the Earth, because the row and column system is not directly related to any spatial reference system (see Chapter 2). But creating such a relationship for a raster matrix is not difficult: one needs only to indicate the accurate coordinates of a corner of the raster matrix in any coordinate system. Such a procedure is referred to as georeferencing. Then, if a cell size is known (it is constant for a given raster matrix, with all cells identical), and assuming that columns are oriented in the north-south direction, the coordinates of all other corners of the raster matrix as well as the centres or corners of any cell of the matrix can be easily determined in the chosen spatial reference system (Fig. 3.3). In such a case, each raster cell can be related to the corresponding fragment of the Earth's surface, thus obtaining a clear relationship between the property of a specific location and how this property is recorded in the raster model.

As can be seen in Figure 3.3, the referencing of a raster matrix to the coordinates of any spatial reference system defines the most important feature of a raster geographic data model – cell size, often referred to as the **spatial resolution.**[21] The size of the cell allows one to define the scale of the analysis or visualization at which the data recorded using the raster geographic data model will be useful. For example, if we are interested in the elevation of Poland, the size of a cell equal to 100 m x 100 m will be sufficient for a country-wide analysis. However, if we plan to conduct surface analysis in the smallest national park in Poland – Ojców National Park, with an area of just over 21 km^2 – then such a cell size will not allow us to represent the various

[21] Typically, spatial resolution is given as the size of a raster cell or the length of its side, e.g., 10 m or 100 m spatial resolution. Though customary, this is not fully correct: because spatial resolution is an inverse of the size of a cell, the correct term is a spatial resolution of 1/10 m, 1/100 m. **Spatial resolution decreases with increasing cell size, and increases with decreasing cell size**, and this relations is correct if a fraction is used (1/10 m is greater than 1/100 m).

rock forms for which this small park is well-known, such as the famous Hercules's Club. In this case, one should look for raster matrices with cells smaller than 5 x 5 m.

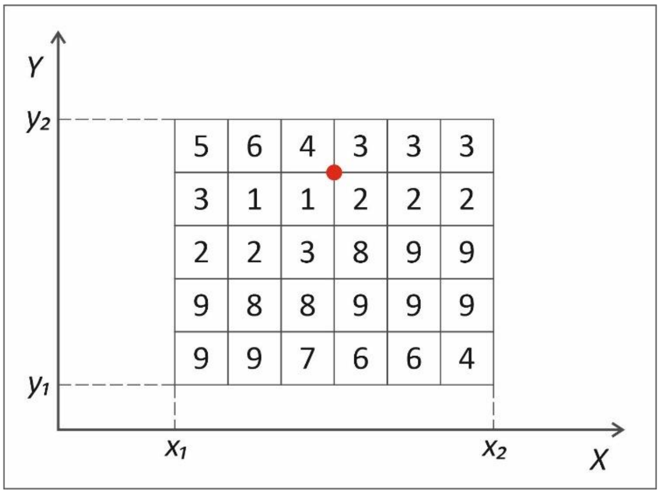


Fig. 3.3. Raster matrix related to a Cartesian coordinate system.

The coordinates of the lower left corner of the matrix are known. Assuming the size of matrix cells is known (s_x and s_y, typically $s_x = s_y$), coordinates of the point marked with a red dot are [$x_1 + 3 \cdot s_x$; $y_1 + 4 \cdot s_y$]. The cell size can be calculated if the coordinates of two opposite corners, the number of columns and the number of rows of the matrix are known, for instance $s_x = (x_2 – x_1) / 6$, while $s_y = (y_2 – y_1) / 5$.

In a raster matrix one cell corresponds to one value. The spatial patterns of properties or the distribution of phenomena over an area is thus represented by various values recorded in all cells of the raster matrix. If someone wants to use a raster model to record a different property of the phenomenon under study or a property of another phenomenon for the same area, another raster matrix is needed. Each such raster matrix is a **thematic layer**, the collection of which allows one to record any number of properties of phenomena of any kind for a selected area (I will also use the term **raster layer**, as opposed to the vector layer discussed later on). Using the notation proposed by Goodchild et al. (2007), various properties of a phenomenon or various phenomena in one location x, y can be recorded and expressed in the following way:

$$x, y: J_1, K_1, Z_1, z_1(x, y);$$
$$x, y: J_2, K_2, Z_2, z_2(x, y);$$
$$\ldots;$$
$$x, y: J_n, K_n, Z_n, z_n(x, y) \quad [3.3]$$

where *x, y* denotes location, falling in a cell of a raster matrix in the row *J* and column *K*, Z_1, … Z_n are various properties, $z_1(x, y)$, … $z_n(x, y)$ are values of these properties recorded in raster cells corresponding to the location *x, y*. Note that the same *x*, *y* location may have different row and column numbers in different raster matrices, because the position of the corners and the spatial resolution of each raster matrix may be different.

In Chapter 2 I have noted that various properties are measured using various measurement scales. The choice of the measurement scale has an impact on how values in a raster matrix are recorded, and several rules are applied in this context. First, numerical values are encoded using the binary code, with a finite number of bits, that is units of information that reflect a choice between two values: 0 and 1. For instance, in the binary code values 0, 1, 2 and 3 are written as 0, 1, 10, 11. Using one bit it is possible to encode only two values: 0 or 1, but to encode 2 or 3 already two bits are necessary – that is, two values 0 or 1. Second, I have already mentioned that one cell of a raster matrix stores only one value. Third, each cell in a matrix uses the same pre-defined number of bits to write a numerical value – which means that if a raster matrix will store values of a certain type, it has to have a sufficient number of bits to write any potential value that can be recorded in the matrix. For instance, the cloud cover observations carried out in Poland at meteorological stations use a 0-8 scale that requires 4 bits to store any value in the range, in the following way:[22]

0: 0000	3: 0011	6: 0110
1: 0001	4: 0100	7: 0111
2: 0010	5: 0101	8: 1000

To record elevations in Poland with an accuracy of 1 m one may use a 16-bit notation, signed (that is each value is written using 16 bits, including its sign, which is necessary as some areas in Poland lie below sea level). Real numbers, on the other hand, can be approximated using the IEEE-754 standard for encoding floating point numbers with various precisions, e.g., single (32-bit encoding) or double (64-bit encoding).

The decision on how values are encoded in a raster matrix, for a given measurement scale and range of measured values, is important to save storage space. Raster matrices sometimes have millions of cells (Box 3.2). Too many bits allocated per each cell may therefore mean megabytes of completely useless storage and occupied memory space for non-existing data. On the other hand, allocating too few bits per each cell in a raster matrix may lead to data recording errors (too high values will be truncated, signs will be lost, the decimal part of a number will be truncated or lost).

Box 3.2. Spatial resolution of ETOPO1 data.

ETOPO1 is a raster data set including global bathymetry and elevation data for the entire Earth (*ETOPO1 Global Relief Model*).[23] Its cell size has a side of approximately 2 km (to be more precise, 1 arc minute, which roughly corresponds to a cell size of 1.85 x 1.85 km at the equator). The size of the raster matrix for the entire globe is 21,600 columns and 10,800 rows, that is 233 280 000 cells in total. Elevations or depths are stored as negative or positive integers, with 1 m precision, using signed 16-bit encoding. In such a case, the volume of the entire raster matrix is approximately 0.5 GB. If the data are stored as floating point numbers, with double precision, the raster matrix would be 4 times larger in volume (2 GB). The benefit of storing elevations and depths with higher precision is illusory, because contemporary methods of measuring elevations and depths at the global scale do not allow one to reach accuracies better than several metres.

A nominal scale (that is names, see Chapter 2) is not directly used to store data using a raster model. Instead, a simple system is used with a raster matrix storing numerical values, each representing a specific name that defines a certain qualitative category recorded in a particular cell. In the *Corine Land Cover* database

[22] Converting a value written in a binary code to the decimal notation is as follows: each bit, starting from the right one, is multiplied by 2 raised to the power of sequential integers, starting with 0, and then all components are summed up. For 7 (that is 0111), the calculation looks like this: $1 \times 2^0 + 1 \times 2^1 + 1 \times 2^2 + 0 \times 2^3 = 1 + 2 + 4 + 0 = 7$.

[23] Accessible at https://www.ngdc.noaa.gov/mgg/global/, accessed: May 2020.

discussed above, a specific numerical value represents a land use or land cover category which has both a name and a numerical code (e.g., class 312 is *Coniferous forest*).

The above considerations related to raster matrices and how they are stored require one more important remark. As said before, a raster matrix is just a sequence of numerical values. Therefore, to properly build a raster matrix one needs to have additional information specifying the number of rows and columns of the matrix. Without this information the cell values shown in Figure 3.2 could be considered, for instance, as values of a raster matrix with 6 rows and 5 columns or as a raster matrix with 10 rows and 3 columns. Any of these two cases requires 30 numerical values to fill all the cells of the raster. The information about the number of rows and columns of a raster matrix is an example of an essential **metadata**[24] for a raster geographic data model. Other important metadata for a raster model are raster matrix corner coordinates in a given spatial reference system. From the point of view of computer storage, a raster geographic data model is a set of ordered files (sometimes stored in a separate folder) or a complex file consisting of both raster matrix data and metadata (Box 3.3).

Box 3.3. GeoTIFF.

GeoTIFF is an example of the raster geographic data model, elaborated in the 1990s based on the raster data model TIFF (*Tagged Image File Format*) which was known and used earlier in computer graphics. To receive geographic data, the TIFF format has been extended with metadata that allow for georeferencing raster data. GeoTIFF is an open data format, initially perfected by the scientific community using geographic data in a raster format. Since 2019, GeoTIFF is one of many standards developed by the Open Geospatial Consortium.[25]

Vector geographic data model

The second method outlined at the beginning of this chapter results in a vector geographic data model (I will use also the abbreviated form of *vector model*). Principally, it differs from the first one in that it focuses rather on the locations and shapes of objects and not on measuring their properties (although these properties can be measured and recorded). The land cover patches discussed earlier have boundaries which can be delimited during measurements and stored as a sequence of points with known coordinates x_i, y_i (in a given spatial reference system):

$$\text{A}: x_1, y_1; x_2, y_2; \ldots ; x_n, y_n \qquad [3.4]$$

where A is a specific object from a certain population of objects, for instance a forest patch or a lake. This notation allows one to conclude that the vector geographic data model uses sequences of points with known coordinates, and these points represent object geometry (Fig. 3.4). A single land cover patch will be a two-

[24] To put it in plain words, metadata are data about data. A good example of metadata is the bibliographical description of a book that can be found in library catalogues. This description contains various information about the book, allowing one to find it on the library shelf, and frequently helps to form some preliminary opinions about its content and to check whether it contains the information we are looking for or not.

[25] Accessible at https://www.ogc.org/standards/geotiff, accessed: May 2020. *Open Geospatial Consortium* proposes many standards in the domain of geographic information that later have become accepted globally.

dimensional object (2D object[26]), with points building its perimeter (a closed polygonal chain) with a default non-empty interior, resulting in a complex geometric shape referred to as a **vector polygon** (or simply polygon). A large number of such land cover patches will be represented by a set of polygons. For objects such as rivers or roads, with a width generally negligible (contrary to their length) one-dimensional (1D) objects will be used for their representation, with points delineating their course and shape, forming simple polygonal chains, referred to as **vector lines** (also **arcs** or polylines). Finally, for objects of a negligible size and for events, zero-dimensional (0D) objects will be used for their representation, referred to as **vector points** (or simply points). In this case, each point represents the position of a single object or event. In a vector model for 1D and 2D objects, points building polygonal chains (either open or closed) are referred to as vertices, and the order of their sequence is meaningful, because the shape of a vector line or polygon will differ depending on the order of its vertices (Fig. 3.4). Some of the vertices called nodes will be discussed later on.

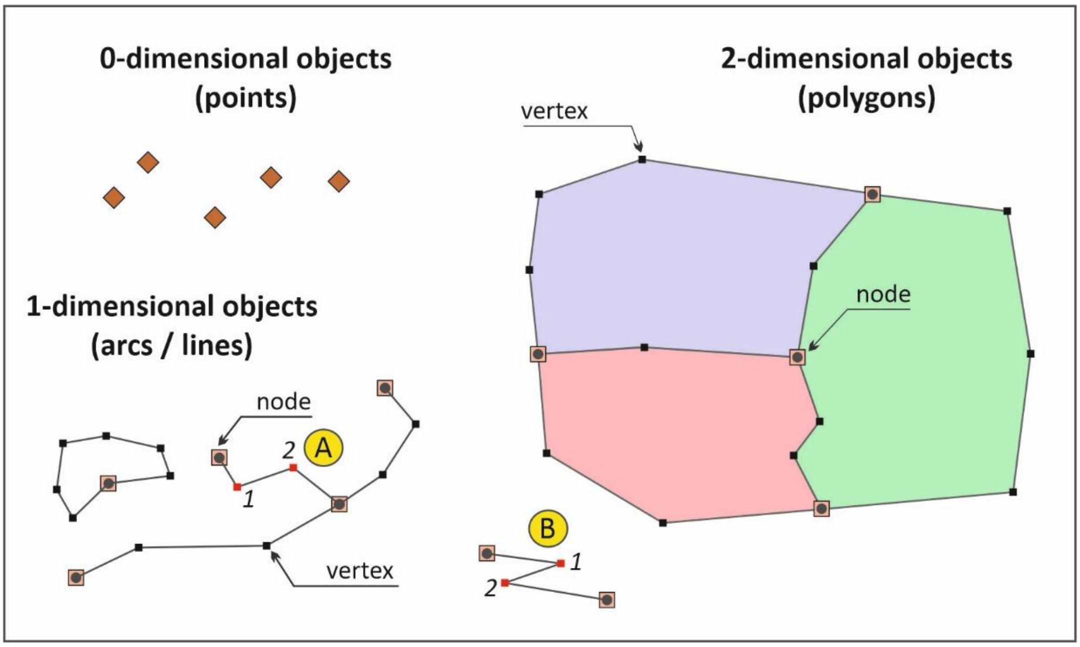


Fig. 3.4. Examples of vector objects.

For the vector lines marked A and B, the order of vertices 1 and 2 differs, resulting in a completely different shape for these two vector lines, though the relative position of the vertices is exactly the same in these two cases.

Adapted from: Jezioro, Kozak, 2004

Because the sequence of coordinates representing the shape of an object does not allow one to explicitly recognize the dimensionality of an object (e.g., a sequence of points of a closed polygonal chain may represent both a vector line or a vector polygon), the dimensionality of a vector object has to be declared independently, being a good example of vector model metadata.

If a vector geographic data model relies mostly on the geometry of objects (or events, however to simplify I will use exclusively the term *object* later on), so as to make it effective in real-life applications, it has

[26] Typically, an object (or event) is something that exists in the real world, while its vector representation (0-, 1- or 2D) is referred to as *feature* – however, to simplify this part of the vector geographic model description I have decided to skip the term *feature*.

somehow to store also the properties of the represented objects. In a vector model, properties are recorded in a simple way – through linking each object whose geometry is represented in a vector model with a record containing the values of its attributes, selected and ordered following some specific rules. As a vector data set typically contains many objects of a specific kind, so the records with attributes of objects are stored in tables, constructed according to the relational data model (Worboys, 1999). In this way, a vector representation has a hybrid character – the geometry of objects is stored separately in a vector model while relational tables store the attributes of the objects.[27] Each row of the table represents one specific object with a known and recorded geometry, while each column contains values of a specific attribute, for all the objects. The unequivocal connection of the object geometry and its properties in the relational table relies on the unique coding of each object in a vector model and in the attribute table ensured by a so-called primary key (Fig. 3.5). The collection of objects of a particular type, represented by a vector model, with tables storing the attributes of objects is a **thematic layer**, later referred to as a **vector layer**.

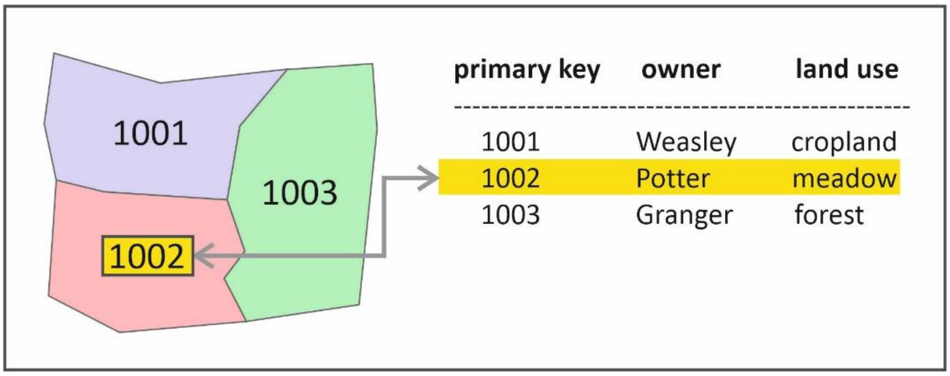


Fig. 3.5. Connecting vector polygons with an attribute table of the relational model.

An object coded 1002 in the vector model (e.g., a land parcel) has several properties stored in a relational table (e.g., owner, land use type), and can be found in a row with a primary key value equal to 1002.

An attentive reader might have noticed a minor understatement related to the description of how land cover patches are represented in a 2D vector model. If each patch has its own boundary (a closed polygonal chain built of a certain number of points) it means that in a vector layer storing 2D objects that are adjacent to each other, each common boundary segment is stored twice – for instance, a boundary between a forest patch and a lake will be recorded once as a part of the forest patch boundary, and once as a part of the lake boundary (see also Fig. 3.1).

How is this problem solved in a vector geographic data model? First, if the main rule of a vector model is equivalence between the objects in the real world and their representation in a vector layer (vector objects, or vector features) – for instance, one land cover patch is represented by one vector polygon (Fig. 3.6 I-II) – this duplication of common boundaries is exactly the case. Such vector models are called sometimes cartographic data structures (Theobald, 2001). But there is also another way to build vector polygons. A vector layer may store only the geometry of boundary segments, and each segment is then stored only once – in addition, however, the layer also records information in a separate table on how these boundary segments build the entire perimeter of each polygon (Fig. 3.6 III-IV). Such a model is called a **topological vector geographic data model** (or simply topological vector model), because it stores, using additional tables,

[27] In fact, the entire vector geographic data model can be written using the capabilities of the relational data model, including both the geometry of objects and their attributes.

various information on the neighbourhood, adjacency and intersections of vector objects.[28] In the case of vector polygons a topological model is not only highly effective, as the geometry of common boundary segments is stored in a database only once and redundant data are removed, but allows one also to avoid errors if the common boundary segment has to be modified – for this model, such a modification is made only once. For a vector model using cartographic data structures such a modification has to be done twice, for two identical but independently stored boundary segments of two adjacent polygons. This is more difficult and may be error prone.

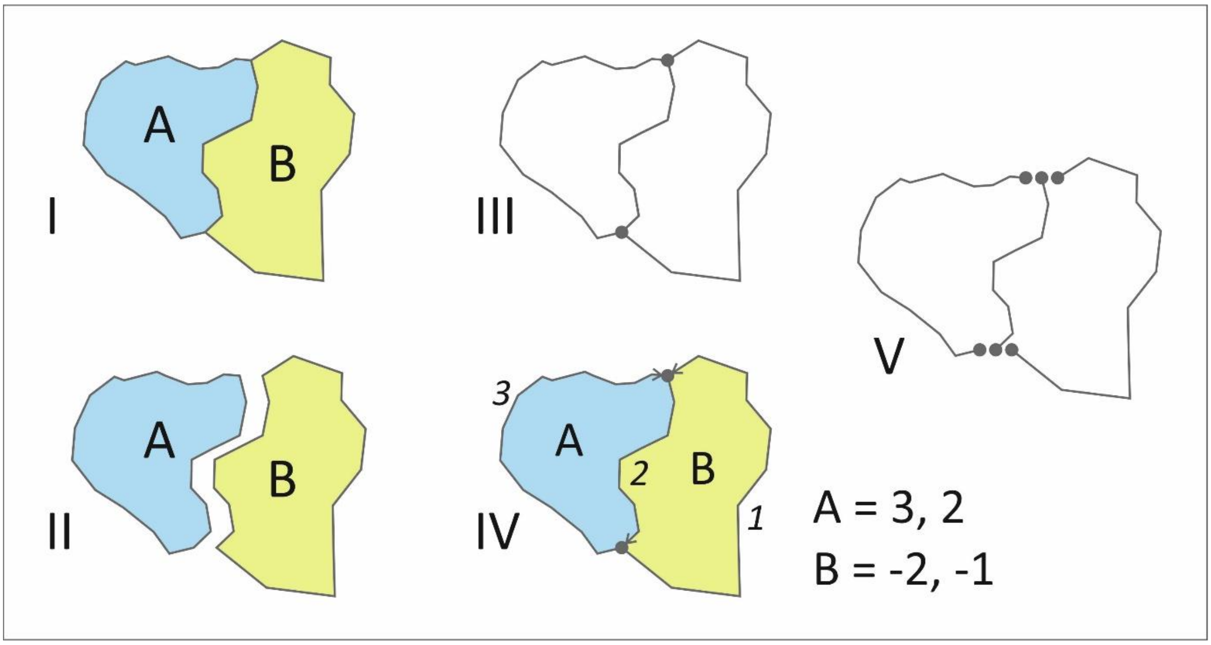


Fig. 3.6. Topological and non-topological vector models of polygons.

I-II: a non-topological model – a common boundary segment of two adjacent polygons A and B is both a part of the perimeter of polygon A and polygon B, this is why it is duplicated in the database. After moving apart the polygons, each polygon remains one complete entity. III-IV: a topological model – three vector lines connect with each other in nodes (see below for an explanation of the term *node*). Taking into account their coding (numbers) and direction one may declare how these vector lines build complete polygons. For instance, the complete perimeter of the polygon A consists of the vector lines 3 and 2, while the complete perimeter of the polygon B consists of the vector lines 1 and 2 ("-" sign denotes line direction). The vector line 2 is stored in the database only once, but it is declared twice as a part of the polygon boundary. In one case, it is a part of the polygon A perimeter, and in the other – of the polygon B. V: If the vector lines 3 and 1 are moved away from each other, the perimeters of the polygons are broken, and what remains are three vector lines that do not connect to each other in nodes.

Besides 2D objects, topological vector models are used also (and even more frequently) in the case of **vector networks**, that is vector layers consisting of vector lines that connect with each other. A topological vector model records in a table all the points where vector lines connect. These points have a special topological significance and are referred to as **nodes**. A node is any point that starts or ends a vector line, and any point where two or more vector lines connect with each other (at the same time, these lines either start or end in the junction).[29] As mentioned before, the points of a vector line that are not nodes are referred to as

[28] These relations, in a slightly different context, are dealt with by a branch of mathematics – topology.

[29] It means that lines in a topological vector model may only connect in nodes, but they cannot intersect in nodes, although it may seem possible to someone looking at a drawing of a complex vector network.

vertices.[30] An advantage of the topological vector model of networks is that as compared to a non-topological model it simplifies the analysis of network connectivity – to do this, it is sufficient to refer to the topological table only, without referring to the sometimes complex geometry of the lines that build the network (Fig. 3.7). Storing connections between nodes in the topological table allows us also to model movement in networks (more on this topic in Chapter 7).

As in the case of raster models, a vector geographic data model is typically stored on a computer hard disk drive as an ordered sequence of files, sometimes kept in a separate folder (Box 3.4).

Box 3.4. A coverage – vector geographic data model.

A coverage is an example of a complex vector geographic data model, invented by the Environmental Systems Research Institute, Inc. (now Esri), as a basic vector model of ARC/INFO software, whose first version was released in 1981.[31] To store attributes, coverage uses a relational data model, while the geometry of vector objects of various types (e.g., points, lines, polygons) are stored in separate files.[32] All files constituting a complete coverage are stored in one folder, with a strictly defined internal structure. A sub-folder *info* keeps attribute tables for vector objects, while sub-folders with the names of vector layers contain geometric data, in files with an *.adf* extension. An *arc.dir* file in the sub-folder *info* stores information on the relations of attribute tables and vector objects.[33] A coverage model allows one also to record topological relations.

Phenomena and their properties in raster and vector geographic data models

To a large extent, it is correct to say that raster geographic data model reflects the continuous aspects of the real world, while the vector geographic data model appeals more to the discrete ones. The raster model is mostly used to represent the properties of various fragments of the Earth's surface (that is, first level objects, sometimes even with no consideration on the nature of these objects), while the vector model focuses on the delimitation and representation of both first- and second level objects and events. The vector model therefore shows a very clear relation between phenomena in the real world and how they are represented in the model. Simply speaking, delimited objects or events are represented in a one-to-one relation using polygons (vector 2D objects), lines (vector 1D objects) and points (0D objects). In the vector model the geometry of real world objects or events has to be known, because representation is not possible without geometry. This requirement is not essential in the case of raster models because they focus on the properties of fragments of the Earth surface in pre-defined locations, while the boundaries between fragments of the Earth's surface showing different properties are not directly stored in the model. These boundaries, however, can be interpreted based on the values stored in the cells of the raster model.

[30] Vertices and nodes are also elements of boundaries of 2D objects (polygons).
[31] Accessible at https://www.esri.com/en-us/what-is-gis/history-of-gis, accessed: May 2020.
[32] Accessible at https://desktop.arcgis.com/en/arcmap/10.3/manage-data/coverages/what-is-a-coverage.htm, accessed: May 2020.
[33] Accessible at https://desktop.arcgis.com/en/arcmap/10.3/manage-data/coverages/contents-of-a-coverage-workspace.htm, accessed: May 2020.

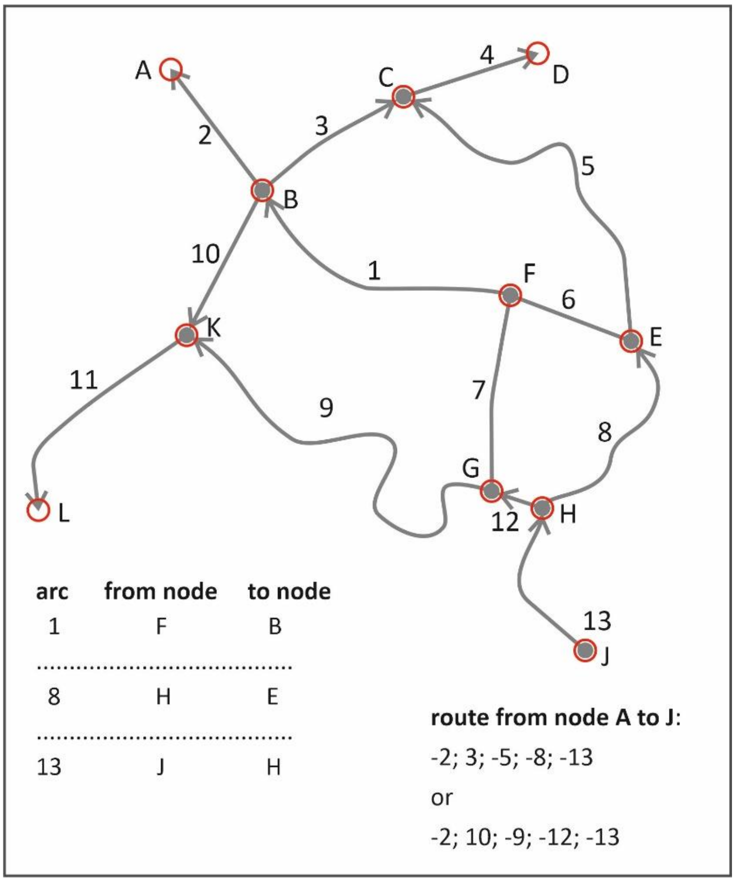


Fig. 3.7. A topological vector model for lines (network model).

A topological table stores information on how vector lines connect with each other in nodes. For instance, line 8 starts in the node H and ends in the node E, where it connects with lines 5 and 6. To find a connection in a network, that is a route from one node to another, it is enough to analyse the topological table, without referring to the geometry of particular vector lines.

Let us consider one more time land cover, and vector and raster geographic data models for this specific property (Fig. 3.8, see also Fig. 3.1). In the vector model, the boundary between *forest* and *lake* is explicit – one object, *forest* is adjacent to the other, *lake*, along the boundary whose geometry is recorded directly with a sequence of points representing the course of the boundary in the field (it does not matter whether the model is topological or not). In the raster model there are cells with values *forest* adjacent to cells with the value *lake*. The distribution of cells results from the regular spacing of the measurement points, with each point representing the cell centre. One may assume that somewhere in between these cells there is a boundary between the two land cover types, yet its exact course is not explicitly stored in the model. In the case of the raster model it is important to assume that raster cells are smaller than land cover patches – if otherwise, some patches might not be recorded in the raster model at all, hence it would not be possible to interpret the boundaries between these patches from the data.

It is worth paying attention to a slightly different meaning of the term *thematic layer* in the case of raster and vector geographic data models. In the raster model, a thematic layer refers to one specific property of a fragment of the Earth's surface. To represent other properties it is required simply to add new raster layers. In the vector model, a thematic layer refers to, and represents, one type of discrete objects, but it may store values of their various properties in the attribute tables. These tables may be easily appended with new fields (columns) storing other properties.

The vector model, similarly as the raster model, may omit small and insignificant objects, simplifying the complex patterns and spatial arrangement of objects in the real world. The ability to simplify is an important feature of both geographic data models, as well as of the entire process of representing reality. As it is in case of conceptual modelling (Chapter 1), various stages of the logical modelling – formalisation of models, measurements implying sampling and a limited capacity of storage – impose to reduce the complexity of the phenomena that are the subject of representation. I will return to this important topic in Chapter 8 of the book.

Fundamental differences between the raster and vector geographic data models should not lead to the conclusion that the raster model is limited only to continuous representation, being the only example of the field-based model, while the vector model is limited only to discrete representation, being the only example of the object-based model (see Galton, 2001). The raster model can be also used to represent discrete point and line objects, and events, because their spatial distribution is a property of any fragment of the Earth's surface. In such a case, the raster model will record if an object or event occurs in a certain location or not, using a simple 1-bit raster matrix, whose cells have only two values – 0 (no object / event) or 1 (object / event). The raster model may be also used to represent discrete areal objects – such objects are represented in a raster matrix by cells assigned one value representing a given object that form spatially contiguous groups. For 0D and 2D objects it is relatively easy to convert their vector representation to a raster, and the raster representation to a vector (Fig. 3.9). Regardless of this possibility, an optimal way to represent discrete objects is to use the vector model, especially when it comes to linear objects, whose raster representation – though possible – is in many ways flawed. On the other hand, the vector model, or strictly speaking its specific alternative, can be used to represent the continuous properties of the Earth's surface, being an example of a field-based model. This specific vector model will be discussed in detail in the next chapter.

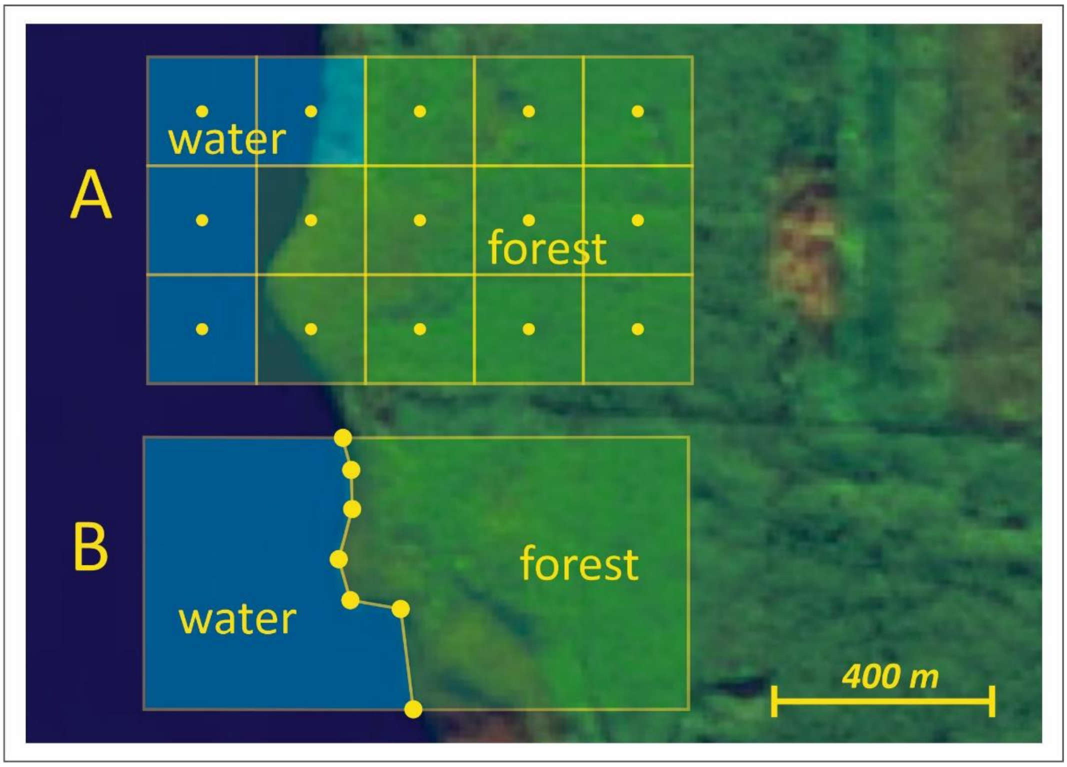


Fig. 3.8. Land cover represented with the raster and vector geographic data models.

A boundary between a lake and forest is explicitly stored in the vector model, and implicitly in the raster model.

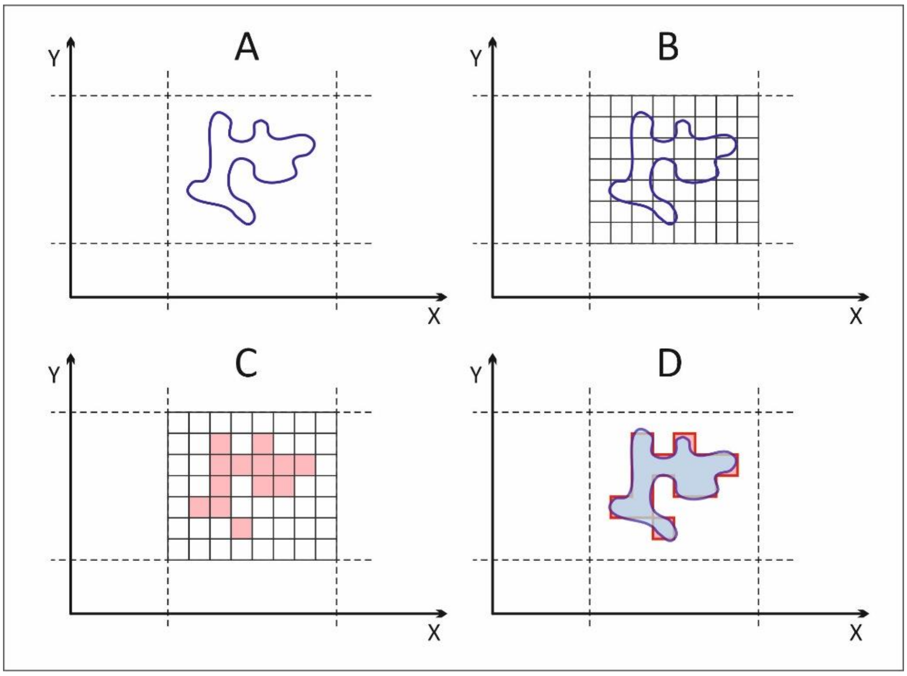


Fig. 3.9. Vector to raster conversion.

A – vector polygon; B – vector polygon with a grid of cells of a raster matrix; C – raster cells that are within the boundary of the vector polygon. Figure 3.9C shows a raster representation of the vector polygon shown in 3.9A; D – an overlay of vector and raster representations – some differences resulting from the vector to raster conversion are visible.

Geographic database

The last section of this chapter is short, but it might as well be written in bold. The process described in chapters 1-3 leads to a coherent digital representation of a certain fragment of reality which in this book is a selected fragment of the Earth's surface and phenomena that occur there. This coherent representation is referred to as a **geographic database** (geodatabase). A geographic database is an ordered collection of thematic layers, recorded using a vector or a raster geographic data model and storing certain properties of an area, the properties of objects that are in this area or events that have occurred there. The thematic layers have an additional description encoded in metadata. The database has also a range of software tools that allow one to manage and to carry out various actions on the data stored in the database.

4. A THIRD DIMENSION: SURFACE ELEVATION

Key words: TIN model, grid, isolines, Digital Elevation Model, 2.5D models, 3D models, voxels, levels of detail, map algebra, focal analysis

We already know what raster and vector geographic data models are, so it is worth now having a closer look at how these models allow one to represent the Earth's surface elevation and how these models store data about elevations or depths. Elevation has been mentioned several times in the previous chapters, as one of important quantitative properties of the Earth's surface. This property is represented by absolute heights (land) or depths (oceans), measured according to some accepted reference level (sea level). In this chapter the methods of height and depth measurements will be discussed within the context of the models for storing the measured values, and the analytical capabilities that these models allow for, here focusing mostly on land elevations.

From the point of view of geography, both physical and human, elevation is one of the most important properties of the Earth's surface. Elevation influences climate and hydrological processes, shapes vegetation and soils, and therefore creates various possibilities for environmental resources to be used by humans. The relations between elevation and other environmental features are most clear in mountain areas. Here, variability in elevations limits accessibility and forms various barriers for human traffic and the development of settlements, determines to a large degree agricultural suitability and defines areas where the risk of natural hazards is high. On the other hand, variable elevations in mountain areas contribute to high biodiversity, offer multiple opportunities to develop tourism as a result of scenic beauty and the possibility to practice a range of summer and winter sports.

Elevation: measurements and geographic data models

Absolute heights representing elevation are a continuous property of the Earth's surface, sometimes referred to as the third spatial coordinate – besides *x* and y coordinates, referring to the two-dimensional location, elevation and absolute heights are expressed as ***z* coordinate**. Absolute heights can be measured today using various methods (Box 4.1). Photogrammetry, radar interferometry and satellite laser altimetry allow a relatively regular sampling of the Earth surface while measuring heights. Field surveying results in measuring heights in irregularly spaced and carefully selected points that are characteristic for the topography of a studied area. Similarly, airborne laser scanning samples the Earth's surface irregularly, yet with an extremely high density. To understand this very important relation between the measurement method and geographic data model, let us first refer these methods to the raster model. However, as is explained later in this chapter, the measurements can be stored using also the vector model.

Measurements carried out using photogrammetry and radar interferometry are typically stored using raster data with a spatial resolution directly linked to the measurement sampling distance. In an ideal case, there is one single measurement for each cell (Fig. 4.1A). In field surveys, measurements are carried out only in some a priori selected cells (Fig. 4.1B), and the values for all other cells are estimated using spatial interpolation methods (more about spatial interpolation in Chapter 7). In the case of airborne laser scanning, there are several measurements for each cell, which means that the value finally stored in a raster cell has to be calculated in some way based on all the measurements recorded within the cell (Fig. 4.1C).

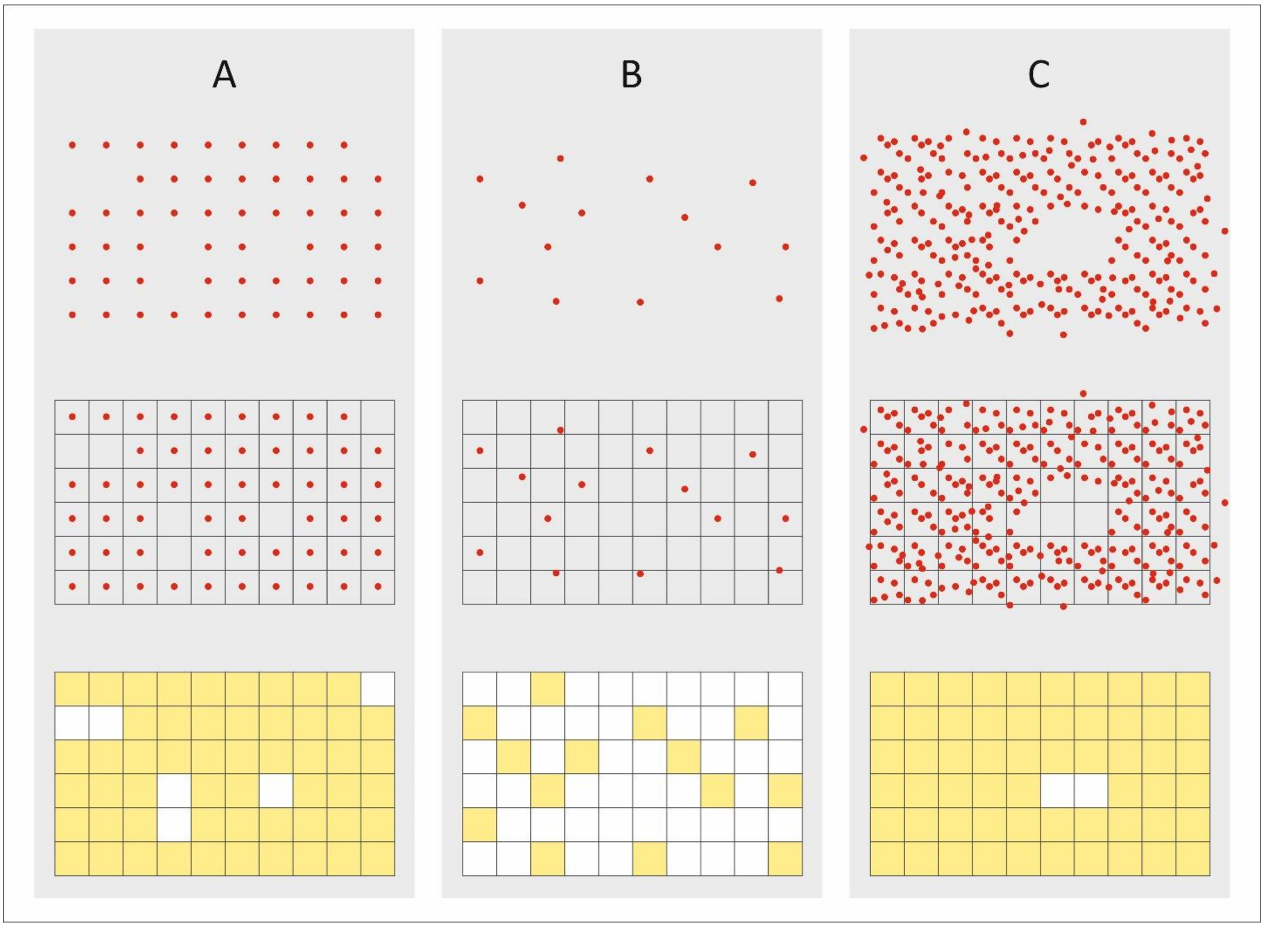


Fig. 4.1. Relation between a measurement and the raster geographic data model.

Top – measurement points, centre – raster cells overlaid on the measurement points, bottom – raster cells with at least one measurement (light yellow) and raster cells without any measurement, for which a value has to be estimated (white). A – measurement in a regular grid, with one measurement per cell; B – a measurement using a random or purposive point sample; C – high density measurement, with several measurement points per cell. In the A and C type methods, for various reasons measurements in some cells may not be carried out.

Box 4.1. Elevation data acquisition methods.

Currently, three methods are used to measure heights over large areas: photogrammetry, radar interferometry and laser scanning. Photogrammetry uses a method which is also a basis of our capability to perceive things in three dimensions: a possibility to measure the height differences of an irregular surface, or the distance to and from certain objects if the surface or objects are viewed simultaneously from two different vantage points. In the case of humans, eyes ensure two different view points, and height differences or distances contributing to three-dimensional, spatial vision result from how the left-eye and right-eye images are combined by our brains. In photogrammetry, two (or more) view points are typically related to taking aerial photos or satellite images from different positions of a plane on its flight path or a satellite on one or more orbits. Even two standard photographs taken from two different locations on the ground allow one to use photogrammetry to measure distances and the three-dimensional properties of objects. Photogrammetry was already in use in the first half of the 20th century, though then in an analogue form. Currently it employs digital imagery and dedicated software to process the imagery (Lillesand, Kiefer, 1994). A very good example of the contemporary capabilities of photogrammetry is a movie presenting how small drones with digital cameras were used to map the elevations of that well-known mountain in the Alps, the

Matterhorn.[34] Photogrammetry has been also used to map elevations on a global scale, with image data acquired by the ASTER sensor.[35] The other method, radar interferometry, employs two sensors separated by some distance that send microwaves and register how the microwaves are reflected back from the Earth's surface (note that in a similar way to photogrammetry, there have to be two vantage points). Radar interferometry has been used twice in global or quasi-global elevation measurements: in the *Shuttle Radar Topography Mission*[36] (SRTM), completed at the beginning of the 21st century, and the later Tandem-X project.[37] An interesting difference between SRTM and Tandem-X is that in the case of the first project two sensors were mounted on one platform (the space shuttle *Endeavour*), while in the latter sensors are mounted on two satellites flying in some distance to each other, with synchronised orbits. A great advantage of radar interferometry is that the method is not sensitive to clouds (microwaves penetrate clouds). The third method, laser scanning, uses LiDAR (*Light Detection and Ranging*) technology, that allows one to measure distance to target based on the accurate measurement of the time taken to send a laser impulse and its return to the sensor, after having been reflected back by some object (e.g., the Earth's surface). Laser scanners currently allow one to send hundreds of thousands of impulses per second, and to register their returns. The scanners are typically mounted on board planes or helicopters, so at altitudes much lower than in the case of satellites, therefore this method results in a very dense sampling of the Earth surface and the objects on it. An outcome of laser scanning measurements is the very large number of points from which the impulses are reflected back, with known *x*, *y*, *z* coordinates, referred to as a 'point cloud'. Still today, airborne laser scanning is not carried out at the global scale – rather, at local, regional or national scales. A very good example of this method's use in practice are the elevation measurements carried out recently in Poland within the framework of the so-called ISOK project.[38]

Digital Elevation Model

A thematic layer storing, for a certain area, height values that have been either measured directly or estimated using spatial interpolation method is referred to as **Digital Elevation Model** (DEM, also Digital Terrain Model, DTM). DEM is defined as *a digital representation of a section of the Earth's surface using a sample of points being a part of this surface and algorithms that allow one to approximate its shape based on the x, y, z coordinates of the point sample* (Gaździcki, 2004). The core of the definition means that what we have with a DEM is a set of points with known *x*, *y*, *z* coordinates and various methods that allow one to estimate heights in other locations – for instance, in locations where there were no direct measurements.

DEMs may use either raster or vector models. The raster model is used more frequently, and basically they do not require additional explanations beyond those provided already with Figure 4.1. The raster model is excellent to store elevation data when measurements methods provide regularly spaced data with at least one measurement per cell (Fig. 4.1A and 4.1C). It is important to note that raster DEMs are frequently

[34] senseFly presents: Mapping the Impossible. Accessible at https://www.youtube.com/watch?v=NuZUSe87miY, accessed: April 2020.
[35] NASA/METI/AIST/Japan Spacesystems, and U.S./Japan ASTER Science Team, 2019, ASTER Global Digital Elevation Model V003. NASA EOSDIS Land Processes DAAC. Accessible at https://doi.org/10.5067/ASTER/ASTGTM.003, accessed: April 2020.
[36] Shuttle Radar Topography Mission. Accessible at https://www2.jpl.nasa.gov/srtm/index.html, accessed: April 2020.
[37] TanDEM Science Service System. Accessible at https://tandemx-science.dlr.de/, accessed: April 2020.
[38] Projekt ISOK. Accessible at https://isok.gov.pl/o-projekcie.html, accessed: May 2020.

referred to as **grid models**, as a grid is simply a regular network, for instance made up of regularly spaced points that represent the centres of regularly spaced squares.[39]

In the case of measurements that are irregularly spaced an obvious choice to store the geographic data seems rather to be the zero-dimensional vector model than the raster one (Fig. 4.1B), because in such a case every single measurement is represented by a point with known *x*, *y* coordinates and a known attribute value (the coordinate *z*). Such a choice makes sense both for a low density of measurement points (e.g., obtained by traditional field surveying methods) and when the density of measurements is extremely high as is the case with airborne laser scanning. Such a set of vector points fulfils the first part of the DEM definition explained above, yet it does not fulfil the part of the definition requiring that the DEM allows for the estimating of elevation beyond the measurement points themselves. This problem, however, may be easily solved when a slight modification of the vector model is introduced. This modification allows one to store the continuous and quantitative properties of the first level objects, and in particular the Earth's surface elevation, leading to a model referred to as the Triangulated Irregular Network, or simply the **TIN model**.

TIN Model

In having a set of irregularly spaced points with known coordinates *x*, *y*, *z*, one may specify the various spatial interpolation rules that will allow one to assess heights in any location where the height has not been measured. A simple method used by the TIN model is arranging irregularly spaced points into a triangular network in which every point becomes a vertex of at least one triangle. Each triangle and its vertices determine an unequivocal plane in the 3D space, which means that for any location *x*, *y* falling within the triangle it is easy to find the corresponding *z* value given that the *x*, *y*, *z* location belongs to the plane of the triangle. In such a way triangles built from points may approximate the surface elevation (Fig. 4.2), allowing one to calculate heights in any location within the area covered by the points. Hence, both conditions specified in the DEM definition are fulfilled.

How triangles are built from irregularly spaced points is clearly defined through the so-called Delauney triangulation. Delauney triangulation fulfils the following condition: the circumcircle of every triangle is empty, that is, it does not contain any other point from the irregular set (Weisstein, 2020; Fig. 4.3). The way points create triangles is written in a table (e.g., in the case presented in Figure 4.3 four points A, B, C, D build two triangles ABD and BCD). Therefore the TIN model may be considered a topological vector model (see also Chapter 3).

As compared to raster models, the TIN model has two important advantages. First, its basic element are irregularly spaced and characteristic points that represent or are located within important features of the Earth's surface: peaks, ridges, valley axes, edges of any type, where typically measurements are carried out during field surveys. In this way the distribution of measurement points may be adjusted to the specific elevations of the area in question, contrary to regular sampling where such an adjustment is not possible. Next, as a result of this first advantage, the TIN model is much more efficient than raster models, because the distribution of measurement points may be in general regulated in such a way that areas with diverse elevations have more points than do the less diverse areas. On the contrary, in raster models the density of measurement points is constant and has to be fitted into the areas with the highest diversity, leading to redundancy in less diverse areas (Fig. 4.4). On the other hand, the TIN model is not as simple as the raster

[39] Regular sampling may result also in another regular network model, a lattice. A node in a lattice does not represent the centre of a cell with one elevation value for the entire cell; rather elevations are calculated in between the nodes taking into account the four nearest nodes. Accessible at https://blogs.ubc.ca/advancedgis/schedule/slides/spatial-analysis-2/lattices-vs-grids/, accessed: May 2020. The difference between a grid model and a lattice model becomes insignificant with the increasing spatial resolution of grids and lattices.

model – the latter is much more convenient if elevation data are analysed. Therefore the TIN model is used mostly to store elevation data, and when the data are to be analysed, the TIN model is converted to a raster with the required spatial resolution.

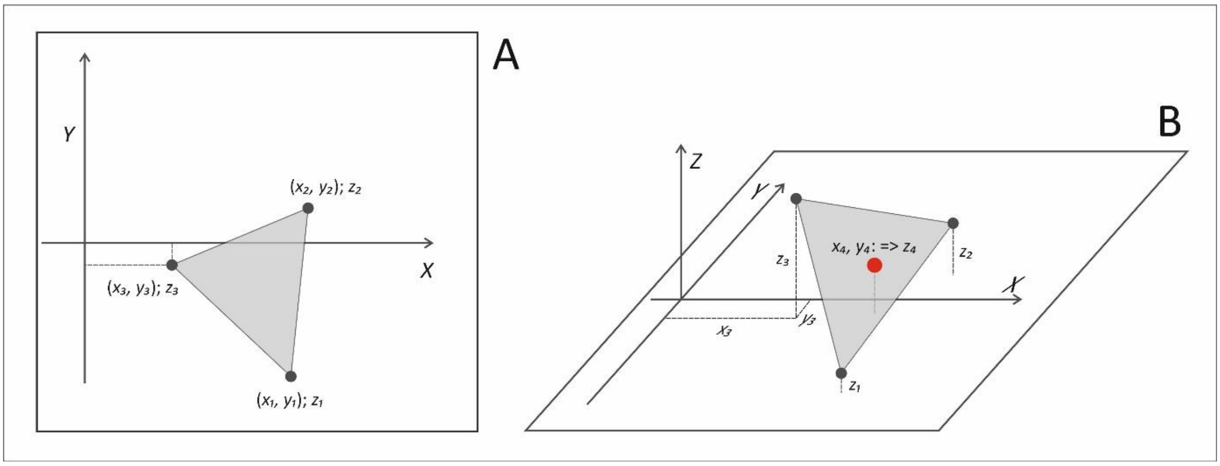


Fig. 4.2. TIN model.

A – a triangle with vertices x_1, y_1; x_2, y_2; x_3, y_3 in a Cartesian coordinate system. Values z_1, z_2, z_3 (heights determined for the vertices) can be considered point attributes; B – the same triangle in a three-dimensional coordinate system where the Z axis denotes elevation. In this coordinate system point locations are set by three coordinates *x*, *y*, *z*. Knowing the *x*, *y*, *z* coordinates of the vertices of a given triangle, the corresponding z_4 value can be calculated for any point (e.g., x_4, y_4) inside or outside the triangle.

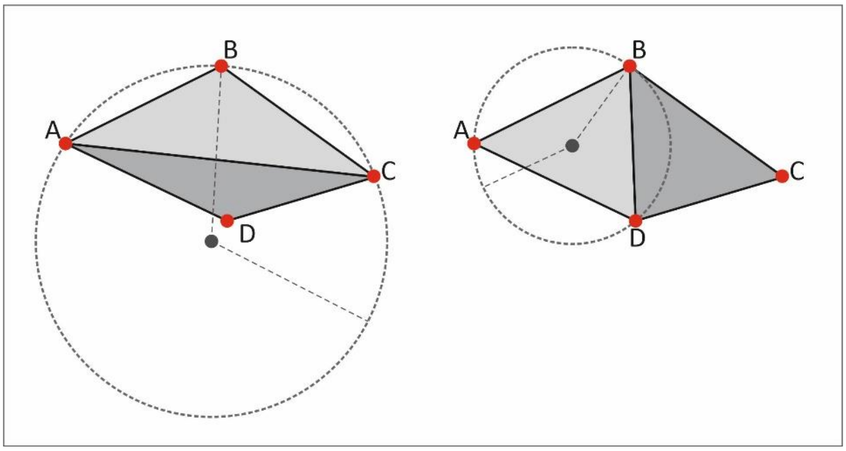


Fig. 4.3. Delauney triangulation.

Left: for the points A, B, C and D, the triangles ABC and ACD do not fulfil the conditions of Delauney triangulation – point D lies within the circumcircle of the ABC triangle. Right: the triangles ABD and BCD are constructed according to Delauney triangulation – point C lies outside the circumcircle of the ABD triangle.

With no doubt the TIN model is an example of the field-based data model (see Chapter 3). Worthy of mention is that although the basic elements of the TIN model are points with *x*, *y*, *z* coordinates – a typical feature of vector models – some researchers (e.g., Peuquet, 1990) regard the TIN model as a variant of raster models, because – similarly to raster models – it is based on a tessellation, that is a disjoint and complete division of a plane into geometric figures to which some values are linked that characterise the properties of the Earth's surface. The raster model is in fact an example of a tessellation made up of square cells, with every cell having assigned a certain value (for instance, representing height). Triangles building the TIN model could play a similar role, yet it is important to note that these triangles do not store a single value, but rather serve to compute height values in any location within the specific triangle – something that is not characteristic for a tessellation. This is why the TIN model should be considered a vector model, in the class of field-based data models.

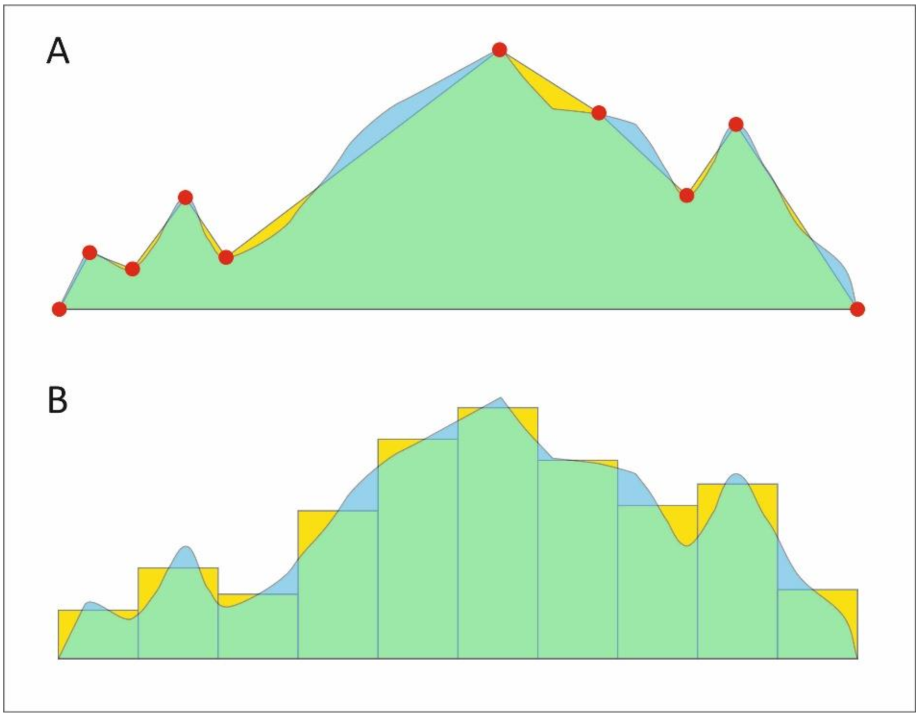


Fig. 4.4. An elevation profile and its representation with TIN and raster models.

The TIN (A) and raster model (B) have both the same number of measurement points (10). Blue denotes the sections of the profile where actual height is underestimated, while yellow – as overestimated by the model. Divergences between the elevation profile and the model are larger for the raster model.

Isolines

Cartography has invented and applied various methods to present the elevation of the Earth's surface. Finally, among all the methods applied, **isolines** have proved to be the best. Isolines are lines connecting points having the same value of some property – for instance, elevation. In the case of elevation isolines are referred to as contour lines. Contour lines allow one to obtain the heights of points located along a contour line and in areas between contour lines, the latter task can be solved with spatial interpolation methods (Fig. 4.5). Because contour lines can be easily stored using a vector model – contour lines are nothing more than mere 1D objects – so it is worth asking the question as to whether this is a proper model to store elevation data or any other continuous quantitative properties, like air temperature, which is frequently presented using isolines (isotherms). The answer to this question is basically "no", with there being three reasons for such an answer. First, the contour lines are not an effect of direct elevation measurements but rather of processing point-based measurements – which means that contour lines contain not only measurement errors but also errors of calculations that lead to their delimitation.[40] Therefore, contour lines are secondary data and of a lower accuracy than point-based measurements. Next, interpolation methods that allow one to assess heights outside contour lines are not easy to translate into efficient algorithms: typically these algorithms require one first to convert contour lines to points and then to construct a TIN model out of them. In such a case it seems much more reasonable to store point measurement data directly in the TIN model, skipping the additional and redundant conversion of point measurements into contour lines. Finally, if contour lines – for any computation – have to be converted into the sequences of points out of which the TIN model is constructed, it means that the contour lines store, unnecessarily, much more data than are needed for an efficient elevation representation.

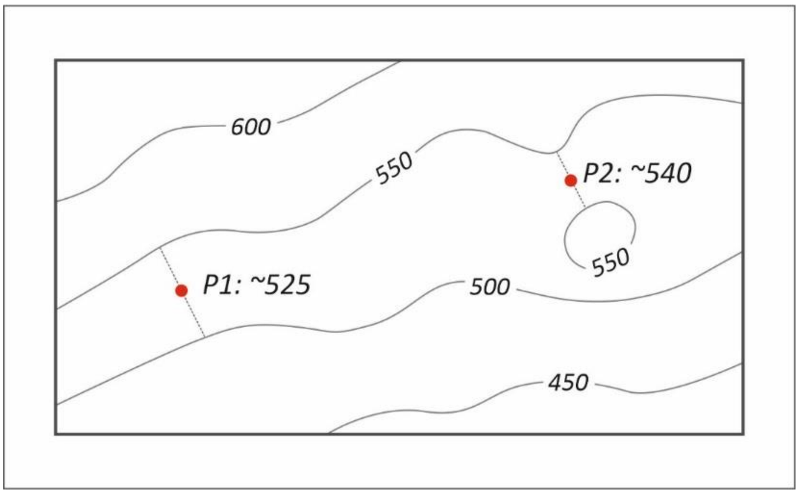


Fig. 4.5. Spatial interpolation based on contour lines.

For P1, estimating the height value is relatively easy because one may assume a relatively constant slope between contour lines of 500 and 550 m; in such a case the P1 value will be approximately a mean of the values of these two contour lines. Such a simple operation cannot be performed for P2 – here additional assumptions about the terrain have to be made.

Therefore, contour lines remain currently only a method for a visual presentation of elevation, a method which is nevertheless extremely useful. They are not, however, a geographic data model that is reasonable from the point of view of representing continuous quantitative properties such as elevation. In having a DEM of any area, we may easily check the layout of the contours in this area using a software to process and analyse geographic data that as a standard allows one to generate contour lines from raster or TIN models (Fig. 4.6).

Surface analysis

When analysing DEMs one interesting type of operations from the toolset of so-called map algebra is used. **Map algebra** (also cartographic modelling; Tomlin, 1991; Longley et al., 2011) are mathematical operations carried out on raster layers. The operations used in the surface analysis are **focal functions**, sometimes referred to as moving window analysis. Focal functions calculate, for a specific location (e.g., a raster cell), a value based on the values of the cell and the values in its neighbourhood. The most typical neighbourhood

[40] Regardless of whether the contour lines are computed mathematically using some interpolation method or are drawn intuitively based on a detailed knowledge of the terrain by a cartographer, the contour lines are always the result of an interpretation of measurements carried out in points.

of a cell in a raster consists of 8 cells around the cell and the cell itself (Fig. 4.7), however in practice neighbourhood shapes can be freely defined. Focal functions allow one to derive several basic layers from a DEM – for instance slopes, aspects and hillshading (Fig. 4.6). They are also used to conduct some more sophisticated operations resulting in topographic or hydrological indices. In the latter case, one of the basic layers for hydrological modelling computed from a DEM with focal functions defines water flow directions.

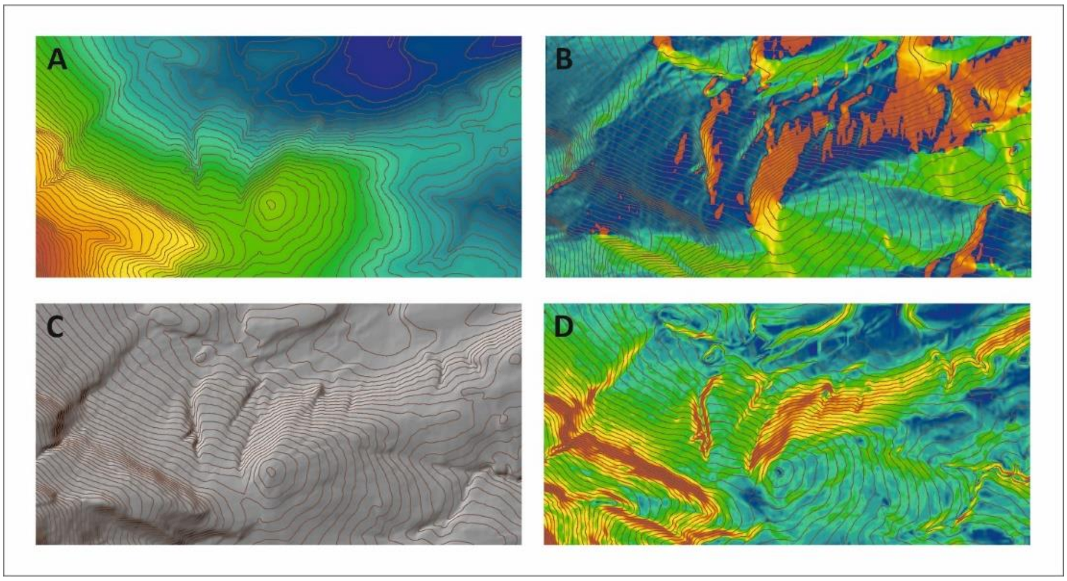


Fig. 4.6. DEM stored using a raster model and basic outputs of its analysis.

Contour lines are computed from the elevation data. A – DEM; B – aspects; C – hillshading; D – slopes.

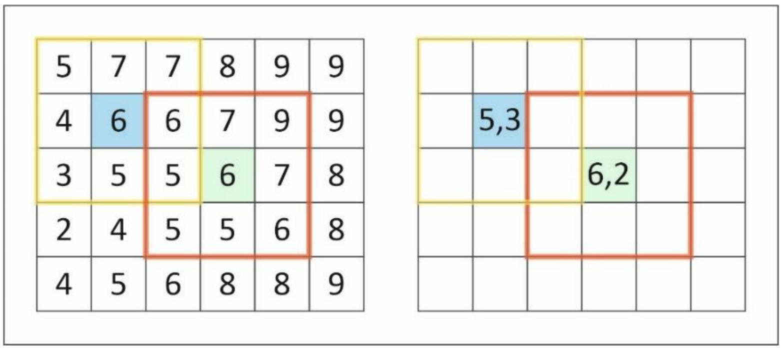


Fig. 4.7. Focal function and its operation.

For a cell marked blue (raster matrix on the left) its neighbourhood is shown with a yellow outline. The focal function is “MEAN” – its result, the mean value of all the cells of the neighbourhood, is entered in the output raster matrix (on the right), in a cell representing the same location. The same function is applied for another cell (marked in green). Typically, the function is computed for all cells and their neighbourhoods in the input raster matrix, while the results are stored in the output raster matrix.

True 3D models

Although raster and vector DEMs are frequently referred to as 3D representations, in fact they are not models of three-dimensional objects – solids – but only models of their surface. This is why DEMs are sometimes called **2.5D models**, clearly distinguishing them from true **3D models**.

True 3D models, representing solids – for instance the Earth or any of its part – have relatively little significance in geography, but are used widely in related research fields and disciplines such as geology, oceanology, climatology and geophysics, where they are used to represent the three-dimensional diversity of geological structures, the chemistry and thermal properties of ocean waters, the composition of the atmosphere or the dynamics of atmospheric processes. One type of 3D models is a three-dimensional extension of standard raster models. Their basic element is a **voxel** – a unitary cube that may be attributed a certain property occurring in a specific three-dimensional location in a solid. In this way, the entire solid can be made up of a huge number of such unitary cubes. An interesting example of how such a model can be used is offered by the *GeoCraft* project carried out in the Netherlands, where a 3D model of the entire country was constructed out of unitary cubes 1 x 1 x 1 m using the rules and capabilities of the popular game *MineCraft*.[41] In 3D models created using a vector model logic, a TIN model extension is used in which the substitute of the triangles are tetrahedrons, each constructed of four points being a subset of the irregular three-dimensional point set. Such 3D models are also used in architecture and urban planning where standards of so-called **levels of detail** (LOD) were proposed. Among 5 LODs the first one (LOD0) is simply terrain elevation – that is DEM. Higher LODs include various features of buildings (roofs, walls, their geometry and layout, interior infrastructure and design) and define which elements should be considered at which LOD (OGC City Geography Markup Language (CityGML) En-coding Standard, 2012).

[41] Accessible at https://geocraft.nl/english/, accessed: May 2020.

5. IMAGE DATA

Key words: remote sensing, sensor, image data, thematic data, greyscale levels (DN), radiometric resolution, spectral resolution, repeat cycle, spatial resolution, reflectance

Our sense of sight, in some simplification, works as follows: various objects illuminated with natural or artificial light reflect light (electromagnetic radiation), this reflected radiation reaches our eyes where photoreceptive cells are activated, these cells then send signals via the nervous system to the brain that interprets the meaning of the signal (Woleński, 2007). The sense of sight reduces the diversity of the real world because our eyes register electromagnetic radiation only within a very narrow range of wavelengths, so-called visible range (380-740 nm).[42] However, even this narrow range allows us to perceive and distinguish a full range of colours owing to the properties of various objects and their surfaces that reflect light in various ways depending on the wavelength. For instance, in the visible range green leaves reflect mostly light with wavelengths ranging from 500 to 565 nm, determining our perception of vegetation as green. With the sense of sight, we register not only colours but also contrasts, textures, relations among objects and their movements – all because of the immediate interpretation of signals transferred from our eyes to the brain via the nervous system. This means that we not only perceive things, but we know what we perceive. Our sense of sight is therefore an evolutionary shaped **sensor**, a device to register certain physical effects. Contrary to sensors constructed by humans, the sense of sight is almost perfect just because of how the registration of electromagnetic radiation is coupled with an interpretation that transforms physical signals into useful information allowing one to take relevant actions and decisions.

The registration of some physical effects by a sensor results in producing data that represent the specific properties of a studied object. If electromagnetic radiation is being considered, the result of the registration are **image data**. Image data record the amount of electromagnetic radiation, either reflected from or emitted by some surface, using a regular sampling of the surface, and the raster geographic data model for storage. In digital photography, regular sampling is ensured by a matrix of light-sensitive elements (detectors): radiation registered by a single matrix element in a very short time (fractions of a second) is converted to a number stored in the pixel of the raster matrix (see also Chapter 3 and Footnote 23), with the raster matrix having the same number of rows and columns as the matrix of light-sensitive elements. In the case of satellite sensors some other sampling techniques may be applied. For instance, a certain number of detectors can be coupled with mechanical devices that allow one to rotate the detectors, and for a given orbital parameters and speed, to register by a single detector the electromagnetic radiation incoming from regularly spaced areal units. Then, registrations combined from all detectors fill in a raster matrix with a pre-defined number of rows and columns.

Understanding how image data differ from other raster data, frequently referred to as **thematic data**, is very important. When it comes to thematic data, numerical values stored in raster cells have an unequivocal interpretation as for their meaning, regardless of how these data are displayed on a screen and perceived by an observer. An example of thematic data are elevation data, for which a cell value denotes simply a certain height occurring in a specific location, independently of how it is presented. For other thematic data, land use, a value in a raster cell denotes a specific land use category, according to its pre-defined classification scheme. In the case of image data, though numerical values depend on the physical properties of the objects under study and their surfaces, these values do not translate explicitly into information about what these objects are. Such information can be received only through image interpretation. One of the most simple

[42] Wavelengths for spectral ranges are given after Wikipedia (*Visible spectrum*; *Infrared*). Accessible at https://en.wikipedia.org/wiki/Visible_spectrum and https://en.wikipedia.org/wiki/Infrared, accessed: May 2020.

methods of interpretation is a visual interpretation carried out by humans. Looking at image data displayed on screen we are usually able to recognize quickly which objects we can see, and to name these objects, similarly as we do in everyday life. Image interpretation thus gives meaning to objects registered in image data, and converts image data into thematic data.

To sum up, image data recorded by various sensors are equivalent to what our sense of sight would register if we were able only to store the data, but not able to interpret them as we normally do. In a slightly different context Peuquet (1988) aptly stated that image data correspond to the world as seen, while thematic data[43] to the world as understood.

Recording image data: technical aspects

Image data are recorded by various sensors carried by a variety of platforms. In this section I will focus on satellite sensors that have been providing the largest amount of image data, monitoring continuously the state of the Earth's surface: so-called Earth observation (EO) satellites. Similarly as humans, the most typical sensors of these satellites are passive: for the observation they use sunlight reflected from the Earth's surface. There are also active sensors used in the EO domain: they emit electromagnetic signals themselves and register how these signals reflect from the Earth surface or various objects. The active sensors are used for instance in radar interferometry and LiDAR systems (see Box 4.1). Acquisition of various types data, for instance image data or the elevation data discussed in Chapter 4 is the focus of **remote sensing**, a domain of science and technology that elaborates and uses various methods to acquire data about remote objects using passive and active sensors. The active and passive sensors employ a range of physical phenomena, in particular electromagnetic radiation. Remote sensing deals also with methods of data processing.

A single detector of a sensor records the electromagnetic radiation of a certain wavelength, incoming in some moment from an areal unit with a certain size on the Earth's surface. Let us have a closer look at what it means, decomposing the statement into the following factors: registration, moment of registration, electromagnetic radiation wavelengths and areal unit of the registration.

The registration means that a certain integer value is recorded that corresponds to the amount of energy carried by the electromagnetic radiation incoming to the sensor. The value is referred to as a greyscale level or a Digital Number (DN). Integers are convenient as they can be easily stored as *n*-bit data depending on the number of levels one is going to record, wherein the number of levels will be equal to 2^n (see Chapter 3). Older sensors have typically used 8-bit registration, that is recording of the incoming energy using 2^8 (256) integers, which translates into DN values ranging from 0 to 255. DN = 0 means then the energy lower or equal to some pre-defined minimum level, DN = 255 denotes energy higher or equal to the pre-defined maximum threshold, while energy levels in between are linearly scaled (Fig. 5.1). Similar rules are applied also when registration employs a larger range of DN values, for instance currently used 12-bit registration level.

The number of DN values is sometimes referred to (rather imprecisely) as **radiometric resolution**. In fact, radiometric resolution is a technical property of a sensor and defines the minimal distinguishable difference of the incoming energy. However, with proper sensor design, 1 DN should correspond to this difference, hence the number of DN values may be used as a proxy of radiometric resolution. In this way, radiometric resolution determines if objects with similar properties (e.g., similar land cover types) will be distinguishable or not (Fig. 5.2).

[43] In the original phrase it is 'structure', but this term refers to vector data, and may be understood in the broader context of thematic, interpreted data. In fact, Peuquet (1988) considered the duality of maps being a structure or an image, but this duality holds also very well in the case of image and thematic data.

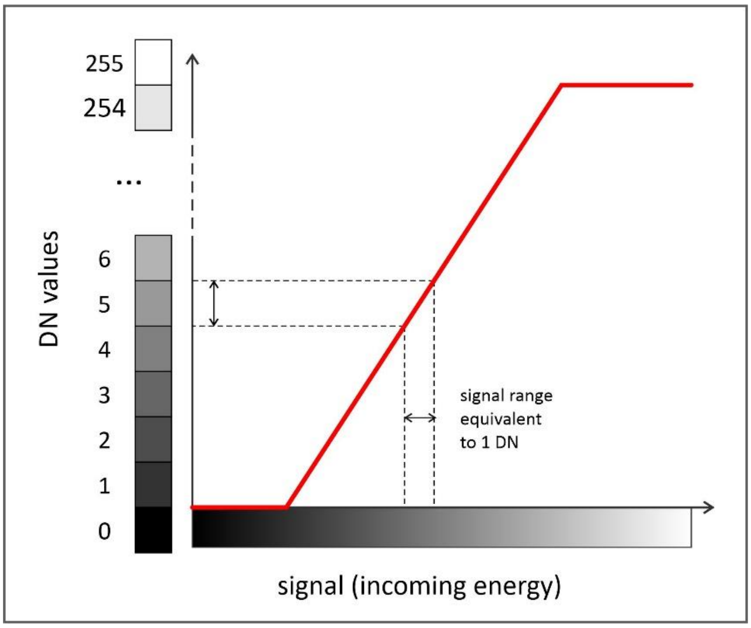


Fig. 5.1. Transformation of the signal recorded by a sensor (amount of incoming energy) into DN values.

Red line is an example of a transformation function.

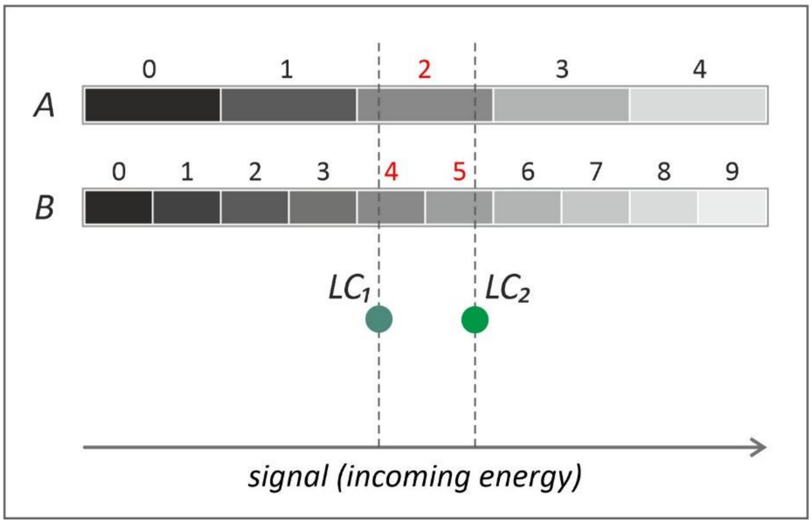


Fig. 5.2. Registration of energy incoming from two similar land cover types (LC_1 and LC_2) by two sensors having different radiometric resolutions.

Sensor A, with low radiometric resolution, records the same DN value for both land cover types, while sensor B, with high radiometric resolution, assigns to both land cover types a different DN value, allowing, contrary to A, to distinguish between these two land cover types.

The moment of the registration, and in particular how frequently it is possible to register data over the same fragment of the Earth surface, depends on the orbital parameters of a satellite and the so-called repeat cycle. Older missions had a pre-determined frequency of registrations and constant **revisit time**, which in many cases was problematic when image data were required to monitor rapidly changing phenomena. Now many satellites are able either to modify the orbital path or to tilt the sensor, in this way allowing some flexibility in acquiring data over a certain fragment of the Earth's surface. Even more frequently, identical sensors are mounted on independent satellites constituting constellations. In such a case, the frequency of registration increases, because image data with the same specification can be provided by any satellite of the constellation.

Wavelengths, for which incoming electromagnetic energy is recorded by a sensor, are selected based on three criteria. First, for obvious reasons, some parts of the electromagnetic spectrum (or spectral ranges, that is ranges of wavelengths of the electromagnetic radiation[44]) are not considered at all, because they are blocked by the atmosphere. Next, some spectral ranges in which the Earth's surface receives relatively a lot of energy, and hence relatively a lot of energy may be reflected back to the sensor are preferred.[45] Third, it is common to use spectral ranges which interact with major land cover types in different ways, allowing one to easily distinguish among them – these are spectral ranges that bear useful information about the diversity of the physical properties of the Earth's surface. The first and second criteria limit the potentially useful wavelengths to the visible and infrared ranges.[46] All three criteria taken into account result in selecting specific, relatively narrow spectral ranges, with this selection commonly referred to as **spectral resolution**. If image data are multispectral, that is data are recorded in at least several spectral ranges, then data recorded per each spectral range are stored in one raster layer. In this way, multispectral image data are recorded using a multi-layered raster matrix, with each layer storing the DN values for each specific spectral range of registration.

Four spectral ranges are basically a standard as for registration carried out by EO satellites and their sensors. These are visible green, 500-565 nanometres (nm), visible red, 625-740 nm, near-infrared (NIR), 740-1400 nm, and short wavelength infrared (SWIR), 1400-3000 nm.[47] In particular, combining information carried by electromagnetic waves in visible red and near infrared allows one to assess to what degree a certain area is covered by vegetation, and to discriminate vegetated from non-vegetated areas. However, there may be many other spectral ranges of registration (sometimes many more), and some spectral ranges may have very specific applications.

The size of the areal unit, from which a sensor collects reflected electromagnetic radiation is determined by instantaneous field of view (IFOV) that depends on the specific design of the optical part of the sensor, and the distance of the sensor to the Earth's surface (Fig. 5.3). The lower the IFOV and the distance to the Earth's surface, the smaller the areal unit from which the reflected energy is registered and stored using DN values. The size of the areal unit decides on the **spatial resolution** of image data (see also Chapter 3 and Footnote 21). Assuming that the energy incoming to the sensor is reflected from areal units of a certain size and that sampling distance, determined by sensor design and orbital parameters corresponds to the size of the areal unit, the simplest way to record and store the DN values is a raster matrix with pixel sizes identical to the areal unit of registration (Fig. 5.3).

---

[44] Typically, the term *band* is used in this context.

[45] Some sensors record also thermal infrared radiation. In strict physical terms, thermal infrared radiation is mostly emitted by the Earth's surface rather than being reflected. This aspect has been omitted in this book.

[46] An in-depth characterisation of electromagnetic radiation can be found, for instance, at Wikipedia: https://en.wikipedia.org/wiki/Electromagnetic_spectrum.

[47] What is given above, are minimal and maximal wavelengths for the entire specific spectral range. Minimal and maximal values that are used for specific sensors typically result in much more narrow ranges, and may differ among sensors themselves.

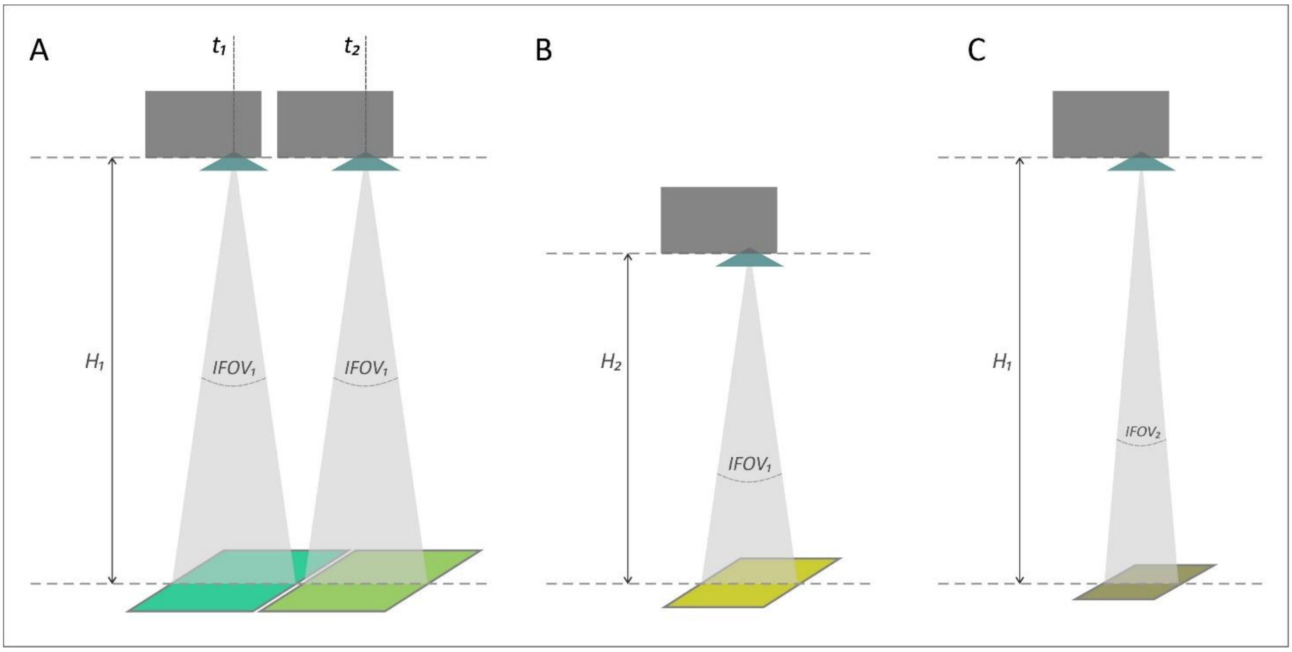


Fig. 5.3. Geometry of image data registration and the size of the areal unit.

A – two subsequent registrations by a sensor of one satellite with a specific IFOV and orbital altitude H refer to two areal units with a given size, which correspond to two adjacent pixels in a raster matrix; B – registration by a satellite sensor with the same IFOV as in A, but from a lower orbit, resulting in a smaller areal unit of registration; C – registration by a satellite sensor with a smaller IFOV than in A, and from the same orbital altitude as in A, resulting in a smaller areal unit of registration.

Earth observation satellite missions

Satellite observations and monitoring of the Earth have been carried out for several decades, with the number of non-military devices now being extremely large.[48] These are both very complex platforms created by large space agencies such as the National Aeronautics and Space Administration (NASA) or by the European Space Agency (ESA) that provide mostly open data, and a range of devices built by private companies all operating on a commercial basis. Undoubtedly, among EO satellites a prominent place is taken by satellites of the Landsat system – the longest EO programme, headed by NASA (Box 5.1). For some years, huge amounts of image data are provided within the framework of the European Sentinel programme. Within this programme, a pair of satellites operate: Sentinel-2A and Sentinel-2B, with sensor design following standards that were elaborated and established over many years of the Landsat programme. Similarly as in case of Landsat, Sentinel-2 data are openly and freely accessible, and can be downloaded through a portal of the European EO programme, Copernicus (*Copernicus Open Access Hub*).[49]

[48] EO satellites database is maintained at the website of the eoPortal Directory, accessible at https://directory.eoportal.org/web/eoportal/satellite-missions, accessed: May 2020.
[49] NASA, accessible at https://www.nasa.gov/, accessed: May 2020; ESA, accessible at http://www.esa.int/, accessed: May 2020; *Copernicus Open Access Hub*, accessible at https://scihub.copernicus.eu/, accessed: May 2020.

Box 5.1. Landsat satellite system.[50]

It is correct to say that Landsat satellites initiated the history of scientific EO programmes. The idea had arisen in the United States in the 1960s, when efforts to develop various space technologies were intensified. The first satellite of the Landsat series was launched on July 23rd, 1972 (named as *Earth Resources Technology Satellite*, ERTS-1). The first three satellites acquired image data with a *Multispectral Scanner* (MSS) sensor, registering electromagnetic radiation reflected from the Earth's surface in four spectral bands (visible green, visible red and two near infrared bands), with a rectangular areal unit of 68 x 83 m. Landsat 4, launched in 1982 had a significantly improved sensor, the *Thematic Mapper* (TM). The TM registered seven spectral bands: three in the visible range (blue, green and red), one in near infrared, two in short wavelength infrared and one in thermal infrared. The areal unit was significantly smaller than in the case of MSS, 30 x 30 m, except for the thermal infrared band (120 x 120 m), while the repeat cycle was 16 days. These parameters have become a standard for the subsequent satellites of the programme. The next generation sensor, the *Enhanced Thematic Mapper Plus* (ETM+), was mounted onboard Landsat 7, and launched in 1999. When compared to the TM scanner, the ETM+ had an additional panchromatic band – a relatively wide spectral range comprising visible and near infrared ranges, with an areal unit of 15 x 15 m. The two latest satellites of the Landsat mission, Landsat 8, launched in 2013 and Landsat 9 launched in 2021, have two independent sensors, *Operational Land Imager* (OLI) and *Thermal Infrared Sensor* (TIRS; for Landsat 9, the sensors are called OLI-2 and TIRS-2). Considering the areal unit and spectral bands, OLI and TIRS have quite similar characteristics to TM and ETM+, however, OLI has two additional spectral bands (one in the visible range and the other in the near infrared range), and TIRS acquires data with the spatial resolution of 100 m. When compared to previous satellites of the programme, Landsat 8 and Landsat 9 have a significantly improved radiometric resolution. Landsat 5, launched in 1984, is undoubtedly the most efficient satellite of the programme. It acquired data until 2011, exceeding many times its planned service duration, being even noted in the Guinness World Records in the category of *longest-operating Earth observation satellite*. Data of the Landsat programme collected over 50 years are publicly and freely available – they may be searched for in various archives, for instance such as *Earth Explorer*.[51]

Image data interpretation

We are able to distinguish hues and to perceive the whole richness of colours because various materials reflect electromagnetic waves of the visible range in different ways, depending on their exact wavelengths. A similar rule – differences in how electromagnetic waves are reflected by various surfaces, depending on the wavelength – is the foundation of image data interpretation. In ideal laboratory conditions it is easy to measure the ratio of the amount of reflected (E_o) and incident energy (E_p), for a given length (λ) of electromagnetic waves:

$$r(\lambda) = E_r / E_i \quad [5.1]$$

[50] Based on Landsat Science: History, accessible at: https://landsat.gsfc.nasa.gov/about/history/, accessed: May 2020; The Multispectral Scanner System, accessible at: https://landsat.gsfc.nasa.gov/the-multispectral-scanner-system/, accessed: May 2020; The Thematic Mapper, accessible at: https://landsat.gsfc.nasa.gov/the-thematic-mapper/, accessed: May 2020; Landsat 5, accessible at: https://landsat.gsfc.nasa.gov/landsat-5/, accessed: May 2020; Landsat 8, accessible at https://landsat.gsfc.nasa.gov/landsat-data-continuity-mission/, accessed: May 2020.
[51] Accessible at https://earthexplorer.usgs.gov/, accessed: May 2020.

The ratio, expressed as a unitless value *r* depending on the wavelength is referred to as a **reflectance**. The reflected energy is always a fraction of the incident energy, hence the reflectance has values in the range from 0 to 1 (or 0 to 100%), while how the reflectance of a specific material varies with wavelengths provides information about the material in question. For instance, reflectances of green vegetation increase suddenly when wavelengths change from visible to the near infrared range, while reflectances of clean water, being quite similar to reflectances of vegetation in the visible range, in the near infrared range are close to zero, which means that water absorbs most incident energy in the near infrared spectral range (Fig. 5.4). Of course, spectral reflectance curves presenting how reflectance varies with wavelength, can be specified for other materials.

Now let us imagine that we have image data that register electromagnetic radiation reflected from the Earth's surface in two spectral bands: visible red and near infrared, using an 8-bit radiometric resolution (DN values range from 0 to 255) – which means we have a standard satellite image, for instance acquired by Landsat 5. In a pixel of this image the following DN values were recorded for these two spectral bands: 15 and 90, and in another pixel – 14 and 5, and yet in another 74 and 57. Let us assume that the incident energy in this spectral range for all three pixels is similar. Because the larger the DN values are, the larger is the reflected energy recorded by the sensor, so it might be concluded that in the first pixel there is vegetation (or a similar material), as the amount of reflected energy is low in the visible band, and high in the near infrared band. A similar reasoning can be carried out for the second pixel, with the likely occurrence of water (as indicated by a very low DN value in the near infrared band). For the third pixel, DN values indicate that a non-vegetated soil surface, rocks or building materials can be expected. The Cartesian coordinate system, with one axis being the reflectance in the visible band (e.g., red) and the other axis in near infrared, is a two-dimensional feature space. Every pixel of an image made up of these two bands can be unambiguously located in such a feature space, and the location indicates quite clearly the type of material or land cover occurring in a pixel (Fig. 5.5). Because there might be many registration spectral bands, and there is also a virtually unlimited number of band combinations, a feature space can be created for a different combination of bands, and therefore can be extended mathematically to any number of dimensions.

The approach described above presents a general rule applied in **image data interpretation**, that is a transformation of DN values registered in various spectral bands into specific categories (for instance, land cover types). The interpretation can be visual – then variations of DNs in various spectral bands for various land cover types are interpreted by humans who know and recognize the reflectances of these particular types. However, automated image data interpretation is much more frequent. In automated image interpretation pixels or spatially coherent groups of pixels[52] of the raster matrix storing image data are allocated to selected categories following an accepted algorithm, which uses DN values from various bands stored in the raster layers. Such an algorithm is, in fact, a precisely defined method to divide an *n*-dimensional feature space into distinct spectral regions, each spectral region representing a specific category.

[52] These groups of pixels are referred to as segments, and the methods leading to their delimitation as segmentation (or image segmentation). If image interpretation is based on segments rather than pixels, it is known as object-based image interpretation.

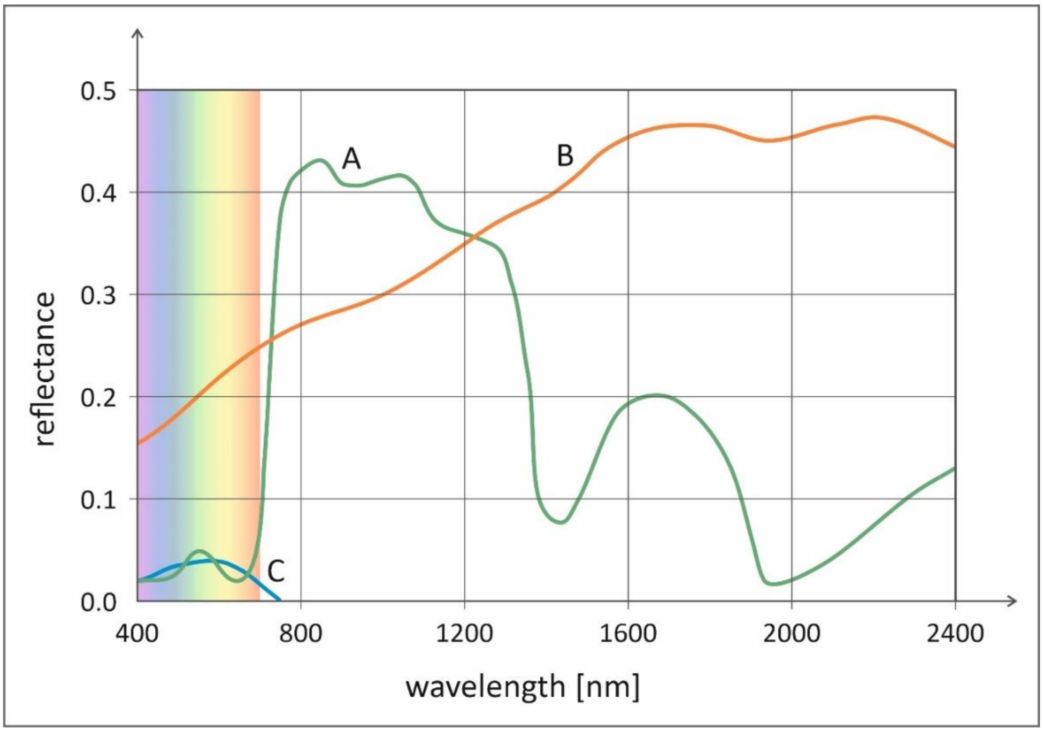


Fig. 5.4. Spectral reflectance curves for various materials.

A – green vegetation; B – bare soils; C – water. The left part of the graph illustrates a location of the visible range in the electromagnetic spectrum.

Adapted from: Lillesand, Kiefer, 1994

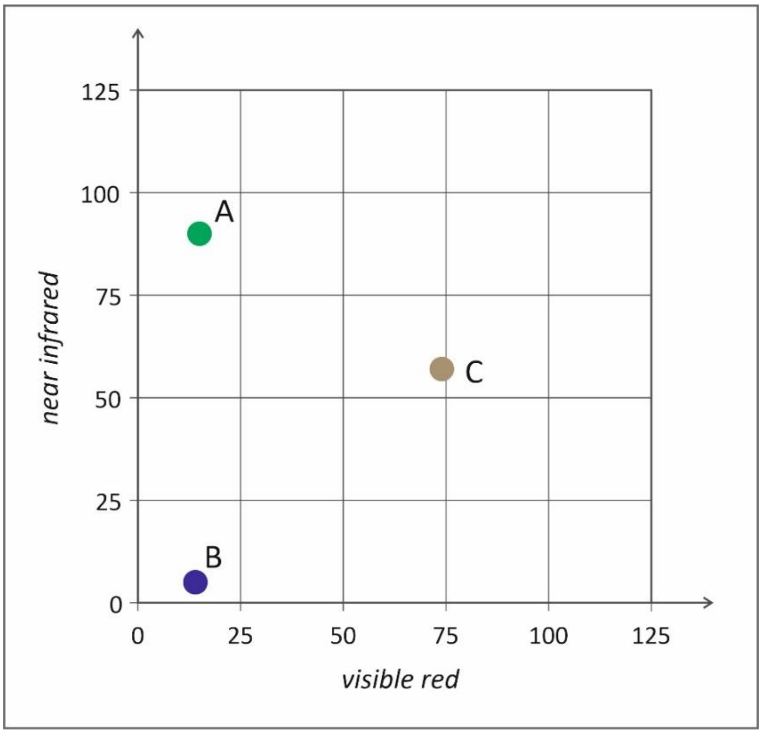


Fig. 5.5. Location of pixels with specific land cover in a two-dimensional feature space.

A – vegetation; B – water; C – bare soils, rocks or densely built-up areas (see also the text for details).

An advantage of image data, and other data acquired remotely, is not only that they provide information about various properties of the Earth's surface at a given moment in time, but, most of all, thanks to the continuous operation of EO programmes and missions (see Box 5.1), the time series of image data allow one to monitor and study changes occurring in the environment (changes and time will be discussed in more detail in the next chapter). Detecting changes occurring on the Earth's surface in any area for a given period using image data acquired at the beginning and at the end of the period can be done in a simple way, visually interpreting both images and finding the differences (Fig. 5.6). The simplicity of this method has undoubtedly many educational advantages, and allows the scientist to easily present and explain to the general public the environmental consequences of various human activities.

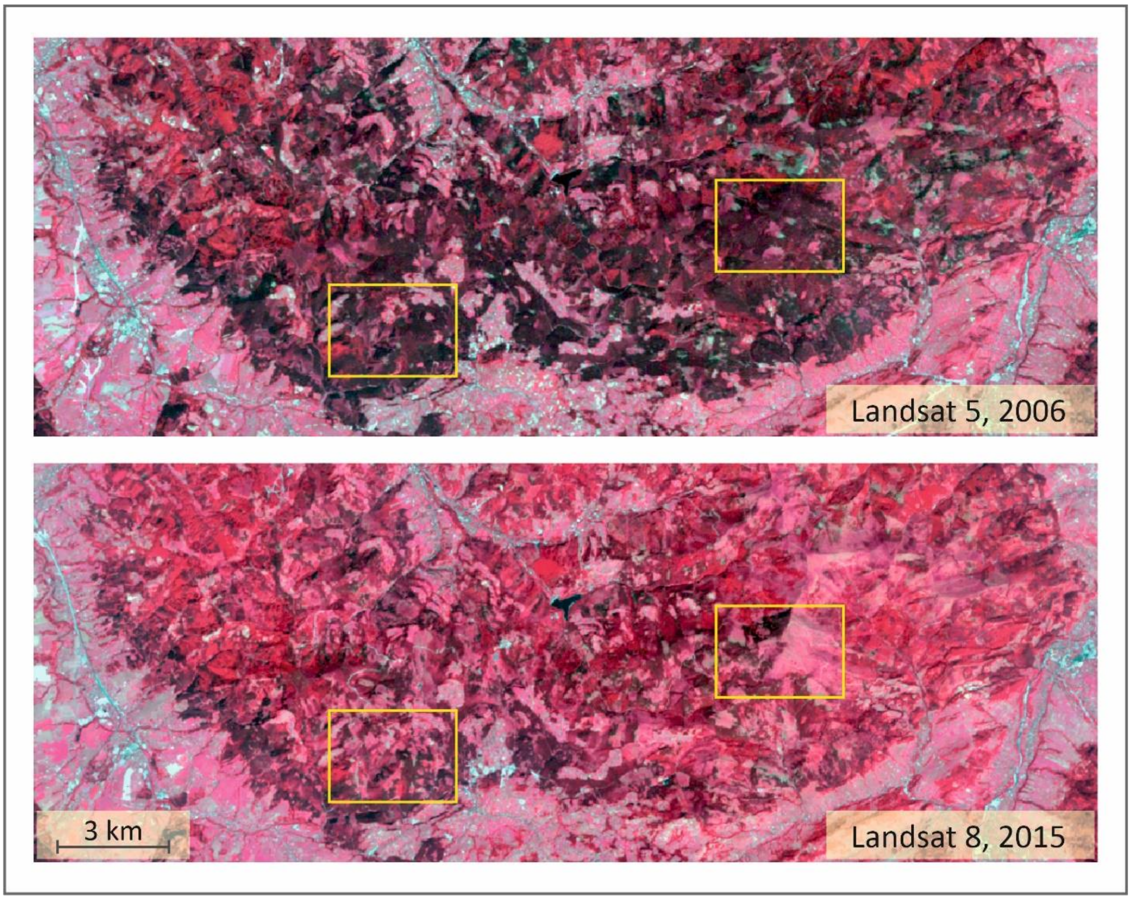


Fig. 5.6. Land cover changes in the southern part of the Silesian Beskid.

It is easy to find changes related to the ongoing spruce decline in this area – spruce forests are brown or dark brown, and yellow rectangles show exemplary areas where in 2006 extensive spruce forests are well visible, but they had disappeared to a large extent by 2015.

6. TIME AND CHANGE

Key words: time, event, change, state, process, system approach, time geography, space-time path, real time

The definition of geography given in Chapter 1 characterizes geography as a science dealing with various phenomena occurring on the Earth's surface. Geography, in particular, studies the various relations between humans and their products on the one hand, and nature and its phenomena on the other. In brief geography studies human-environmental relations. Geography is distinct through its spatial tackling of phenomena, its focus on their distributions in space and on spatial relations themselves. In the traditional approach to geography proposed by Immanuel Kant, geography is a synthetical science grasping things and reality spatially (Turner, 2002), with the other Kantian category, time, being used in a similar way by history. However, similarly as history is impossible without geography because everything happens not only in time, but also in space, so geography also does not hold out without time. Phenomena and their properties change in time due to the processes that shape them, and occur both in time and in space. Variability – how reality changes in time – is built into all phenomena, and humans plan their activities not only in space, but also (and most of all) in time.

In our every day experience, time is combined with space even if only through our movement in space with various velocities, which allow us to cover different distances in various time. We know how much time we will spend biking to work, or driving the car, or even walking on foot. Treating time as an additional fourth dimension may thus lead to the transformation of a three dimensional physical space into a four dimensional space-time. However, though we enjoy the freedom of movement in a three dimensional space, we do not have such a freedom to move across time. So if time has to be treated as an additional fourth dimension, one needs to remember that it differs from all three spatial dimensions. Similarly, in Sinton's (1978) concept of measurements described in Chapter 2, time is dealt with as an additional property or variable rather than an additional dimension. Therefore these various complications related to time have been the subject of discussion about how to implement time in creating a digital representation of reality (Langran, 1993). Some results of these considerations are presented later.

Events, states, changes, stability and time scales

Events have already been mentioned in Chapters 1 and 2. Galton (2001) treats events as temporal analogues of objects, mentioning that events, similarly to objects, are some characteristic combinations of processes and actions that can be named and delimited in time, having their beginning and end. An event is thus a result of a discrete perception of the passage of time. Many events are very short and occur within a very narrow space. An example of such an event could be a lightning flash, lasting mere fractions of a second and confined to a very small area. A vector data model is appropriate to represent such events, and to the x_i, y_i coordinates referring to the location of a zero-dimensional object (point) one may add a value for the moment when this event occurred (t_i), and any of its properties (w_i):

$$lightning_i\text{: } x_i, y_i, t_i, w_i \qquad [6.1]$$

There are many examples of such a notation similar to that above, where together with coordinates locating the event in space, a coordinate referring to time is added (Fig. 6.1).

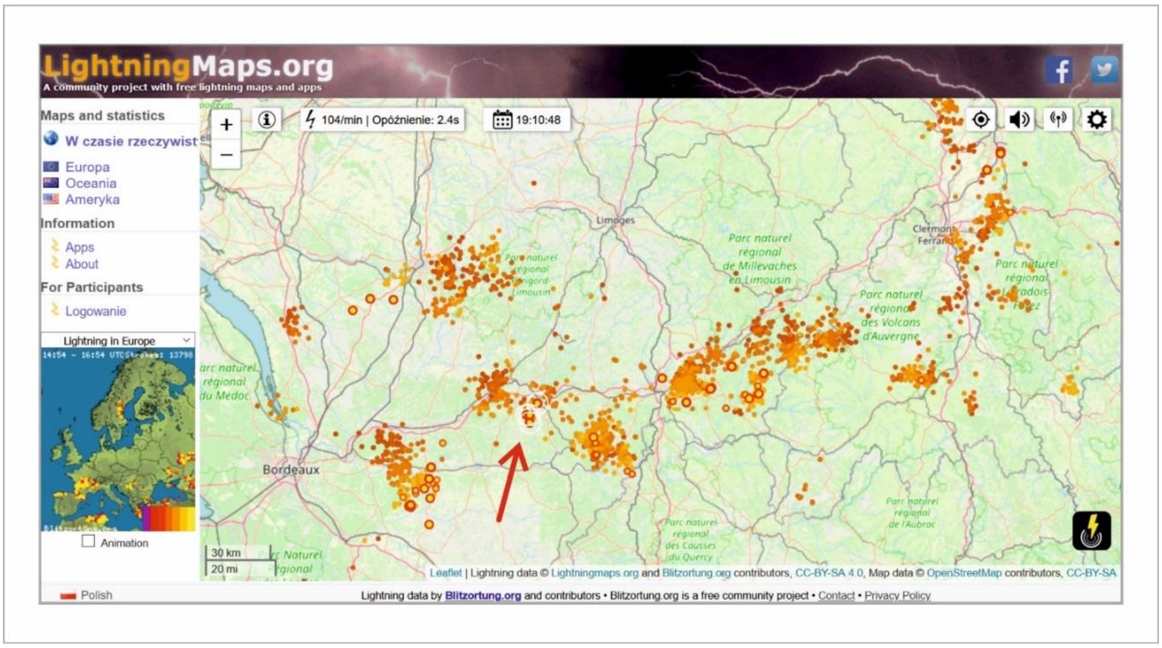


Fig. 6.1. A web portal presenting a map of lightning flashes.

The map shows lightning flashes on June the 1st, 2020, in southern France. The screen shot was taken at 7.10 pm, and the arrow shows the locations of lightning flashes that had occurred some seconds earlier.

Source: *LightningMaps.org* (CC BY-SA 4.0), https://www.lightningmaps.org; data: *Blitzortung.org and contributors*; map background: *OpenStreetMap and contributors*.

One of the more interesting events categories are crimes. Time is an extremely important coordinate in crime geography, because localising crimes in a spatial sense is usually simple and obvious (for instance in the case of a robbery or murder), but defining precisely the time when the crime happened is much more difficult and may be of paramount significance in finding the perpetrator (or perpetrators). Many famous crime novels use the trick of moving the hands of a clock or any other alteration of time done by a criminal to ensure him, or her, an irrefutable alibi for the time when the crime really occurred. A classic example are several novels written by the queen of this genre, Agatha Christie. In one of her best known novels, *Evil under the Sun*, the accomplice of the murderer moves the hands of the clock of an important witness, and then helps to feign the murder before it really occurs in the same location (so the location of the crime in space is true). These complications resulting from this double manipulation of the exact time coordinate of the murder are finally revealed by the eminent detective, Hercules Poirot.

How phenomena occur in time can be perceived and tackled not only in a discrete way, through the registering of events, but instead through a continuous approach, and the recording of **states**, the momentary configurations of values of some properties. In this case discrete events are opposed to **fluents** (Galton, 2001), representing how phenomena continuously unfold. With this approach one may record various values in time without grouping their sequences into events, accepting only that in some location in space something is going on, and that **changes** occur there over some time. The same or similar values in some period may witness a lack of changes and **stability**. A continuous registration over time relies, in fact, on sampling in time, that is the measurement of selected properties in pre-defined, regular time intervals. States can be recorded using either a raster or a vector model, but because of the potential of satellite or airborne sensors to record states of objects on the Earth's surface, the raster model is more frequently used. The analogy of the continuous and discrete approaches related to time with the continuous and discrete approaches related to space is very clear (Fig. 6.2, see also Fig. 1.2).

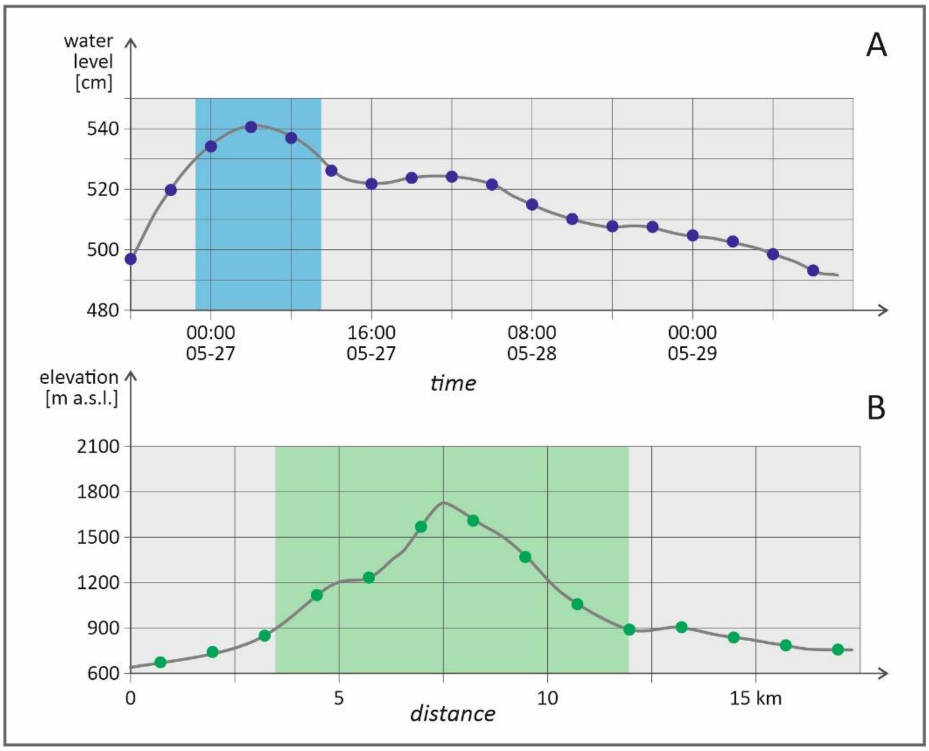


Fig. 6.2. Variability of phenomena in time and space.

Top: data of the water level measurement station in Jawiszowice, the Vistula river, in Poland. Bottom: hypsometric profile along the trail Zawoja Markowa – Markowe Szczawiny – Diablak – Lipnica Wielka (Beskidy Mts., Poland). In both cases the graphs present continuous variation of values in time and space, and the dots represent regular sampling (top: every 4 hours, bottom: every 1.25 km). Shading marks an event which might be referred to as *flooding* (blue, top) and the object is *Babia Góra* (green, bottom).

Source: hydrogram from http://monitor.pogodynka.pl, accessed: May 2020; profile from https://mapa-turystyczna.pl, accessed: May 2020.

Events, delimited in time, are also localized in space – one might say, that an event occurring in time refers also to a certain fragment of space. According to what has already been explained in Chapter 1, our perception of what an event is and what it is not, reflects our perception of permanency, or the lack of it. Here and now this perception is related to how we perceive time and to the patterns of our everyday activities. As for the past, the duration of events are not related to our direct perception, and an event may well be something that lasts for a very long time. For instance, in geomorphology textbooks one may easily find the following statements: *in this period, a vast plain was formed in the area*.[53] Such statements refer to events stretching over millions of years – if we were to observe such an event, we would not have any feeling of being in the middle of an event referred to as the *forming of a vast plain*. Quite likely, we would not be able to even record the critical values for such an event during our lifetime, having rather a feeling of permanency and stability for our surroundings. Quite a similar perception as to the rate of change refers to contemporary climate change. Looking at the climate from the perspective of our own life, the changes seem

[53] This is not a real quotation, yet for instance in the textbook *Rzeźba Tatr Polskich* [*Relief of the Tatra Mountains*] written by the famous Polish geomorphologist Mieczysław Klimaszewski (PWN, Warszawa, 1988), in a chapter discussing the formation of the relief forms (p. 489) one can find a similar sentence, describing an event lasting more than 10 million years: *Powierzchnia zrównania i doliny policykliczne zostały uformowane w długim okresie neogenu (od 17 do 2 mln lat temu)* [*A peneplain and valleys of multi-cycle character were formed in the long period of the Neogene (from 17 to 2 million years ago*)].

to be slight and slow, practically hardly perceptible at all, or only very weakly felt. Maybe this is why they do not seem as dangerous as scientists claim. However, scientists have a much broader temporal perspective thanks to the data located in the space-time, with time coordinate stretching back hundreds, thousands or millions of years, depending on the method employed. Putting the contemporary climate within a context of only a few hundred years immediately shows the abruptness of changes occurring in the last 50 years and the grave seriousness of the climate crisis. In this perspective, what seems stable to us, or changing only very slowly, transforms into a sudden and intense event, one comparable to a disaster (Fig. 6.3).

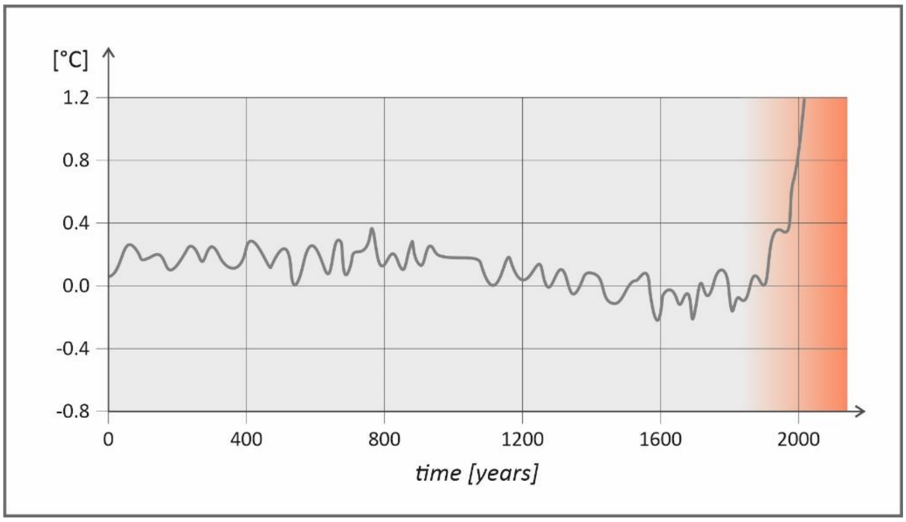


Fig. 6.3. Changes of the global temperature in the last two millennia, compared to the reference period 1850-1900.

Source: the graph in the *Climate Lab Book. Open climate science*, modified. Accessible at https://www.climate-lab-book.ac.uk/2020/2019-years/, accessed: May 2020.

To summarise: humans and their perception define in geography not only the spatial, but also the temporal scales. Many phenomena occur in timescales that are directly accessible to our perception and understanding, but there are other phenomena that occur in timescales of thousands or millions of years. Some properties of the Earth surface or of various objects may be considered constant in some timescales, yet variable in others (Schumm, Lichty, 1965).

Processes and the system approach

Behind the changes to the Earth's surface are diverse flows of matter, energy and information – that is, **processes**. The flow of water in the river bed, solar energy reaching the atmosphere, global atmospheric circulation, the photosynthetic activity of plants, and also an on-line purchase of a pair of shoes that initiates sending them from one place to another – these are all examples of such flows. A system, and a **system approach** are very useful concepts that allow one to address the magnitudes and intensities of flows between various objects and how these flows interact with the objects themselves. A system is a set of elements that are connected and interact with each other due to the processes occurring within the system itself (Chorley, Kennedy, 1971). The most simple system representing flows of matter, energy or information is a cascade system. The basic elements of this system can be considered as containers that receive incoming matter, energy or information, and send them on further. Important in a cascade system are the flows among the containers and the balance of matter, energy or information in each container being a part of the system. In Chapter 5 I have explained the concept of reflectance: here it is worthwhile adding that the reflectance characterises a cascade system related to the transformation of solar energy reaching the Earth's surface. Parts of this cascade are the incident solar energy and solar energy reflected from the Earth's surface back to the atmosphere. The containers are then the atmosphere, and the Earth's surface (composed of the biosphere, the upper parts of the lithosphere and hydrosphere), and the important property of this is the ratio of the reflected and incident energy, that is the reflectance. A simple illustration of a cascade system is also a water cycle. In the case of a simple system such as that of a terminant lake, the most important containers are the lake itself, its watershed and the atmosphere (Fig. 6.4). Rainfall feeds the lake directly when it falls on its surface or indirectly, through the inflowing rivers. Water evaporates from the lake surface back to the atmosphere. If evaporation is balanced by rainfall and river discharge, the lake may be stable for a long time (Box 6.1).

Box 6.1. Disappearance of the Aral Sea.

A well-known case of the disappearance of the Aral Sea, a terminant lake in Central Asia, at the boundary of Kazakhstan and Uzbekistan, is related to the significant decrease of the water inflow from two main tributaries of the Aral Sea: the Syr-Daria and the Amu-Daria. This decrease was initiated in the 1960s following the expansion of large-scale cotton cultivation in the region, and the use of water from these two rivers for irrigation. The lake was fairly stable until 1960, with a river inflow of 56 km^3 and rainfall ensuring 9 km^3, balanced by evaporation (66 km^3), with some minor share of subsurface inflow. With a relatively stable evaporation, decreasing inflow (since the 1980s the inflow of the Syr-Daria and the Amu-Daria has been below 20 km^3) disturbed the balance of the lake and initiated its drying. By 2009 the water level had dropped by 20 m, the area of the lake had shrunk from 68 000 km^2 in the 1960s to approximately 7 000 km^2, while the lake had split into several independent reservoirs. The environmental consequences of this change have been enormous (Micklin, 2010).

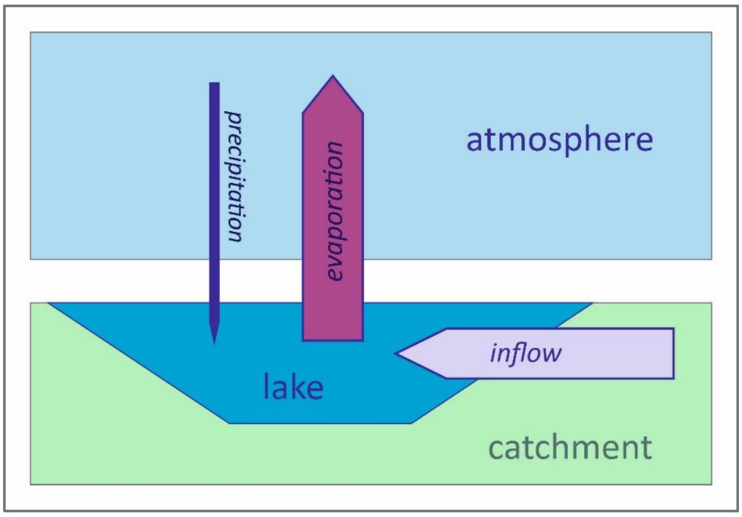


Fig. 6.4. The cascade system of a terminant lake.

The widths of arrows representing rainfall, river inflow and evaporation reflect the water balance of the Aral Sea in the early 1960s (see also Box 6.1).

The snapshot model

One of the easiest ways to record and study the temporal variability of phenomena is offered by the snap-shot model, proposed by Peuquet (2001). The snap-shot model relies on recording the state of the properties of a certain area (or certain objects) at specific moments. A comparison of the states at different moments of time allows one to assess the changes that have occurred between these moments, and inferring about the processes that have shaped the studied area or objects. A method that allows such a comparison in the snap-shot model is referred to as **map overlay**. Map overlay – in its most generic meaning – relies on comparing the value of a specific property in a certain location, recorded in a layer (stored using any geographic data model), with another value of the same or other property in the same location, recorded in another layer, for all the locations in the area where the two layers overlap (for instance, in the raster layer all the cells of the overlapping part of both layers will be analysed). In the snap-shot model the same properties are compared, yet the layers used in the map overlay the record values of these properties at different moments of time (Fig. 6.5).

One important theme in geographic research within which the snap-shot model is commonly used are land use and land cover changes.[54] The state of land use at a certain moment in time is relatively easy to record thanks to the variety of image data – for instance satellite imagery, discussed in more detail in Chapter 5, or simple photographs taken from a specific point at the beginning and at the end of a certain period. Two images, taken for example several years one after another are quite enough to observe and, if needed, to assess with relatively good accuracy, the area and rates of land cover change that occurred in a certain location (Box 6.2) and to infer changes in land use on this basis. Contemporary remote sensing technologies allow one to register properties of the Earth's surface with a high frequency, resulting in growing capabilities to compare satellite imagery acquired at various moments of a period that becomes longer every day (Box 6.3). The *Corine Land Cover* project mentioned in Chapter 3 (Box 3.1) uses the snap-shot model because the registration of the state of land use in Europe every 5-6 years allows one to compare the states of land use at various moments and delimit areas where changes occurred in a given period.

Box 6.2. *Mountain Legacy* project

The *Mountain Legacy Project*[55] uses photographs taken during various research expeditions carried out by the Geological Survey of Canada in western Canada in the second half of the 19th century and the first half of the 20th century. The locations from where these photos were taken are known and that allows one to gradually collect contemporary photos taken from the same locations, presenting the same objects and landscapes many years after. Archive photos were converted to a digital form and made available at the portal, together with contemporary photos taken until this day. The archive and contemporary photos are geometrically adjusted and therefore the portal allows for their interactive comparison. In this way it is relatively easy to study landscape changes that have occurred in a certain location. The *Mountain Legacy Project* is thus an excellent example of the use of the so-called repeat terrestrial photography method to assess landscape changes, frequently employed in geographic studies (Kaim, 2017).

[54] When talking about the registration of the state of the Earth surface using remote sensing and various sensors (see also Chapter 5) one should use rather the term *land cover* and not *land use* (see also Footnote 19), because land cover is a physical state of the Earth's surface that explicitly determines the amount of electromagnetic radiation reflected from the Earth's surface, while land use denotes how humans use the Earth's surface. Land use controls land cover, but relations between land use and land cover are quite complex. One form of land use may result in various land covers, and one land cover type may be characteristic for various land uses. Summarising this short explanation: what is recorded in image data is land cover, which typically allows one to infer about the land use.

[55] Accessible at http://mountainlegacy.ca/, accessed: May 2020.

Box 6.3. Global Forest Change database

Global Forest Change[56] presents the results of the automated interpretation of thousands of satellite images acquired in the Landsat programme, carried out to receive information about recent changes of forest cover on a global scale. The images taken in a given year were merged in such a way as to receive a coherent, seamless and cloudless image of the land surface of the Earth for that year, and later on, based on recorded DNs (or reflectances), the forest cover in the given year was delimited. Having annual data about forest cover for a certain period, year-to-year changes can be delimited. These changes are either forest cover losses that occur due to forest fires, catastrophic winds or clear-cutting, or forest cover gains that occur through afforestation or spontaneous forest succession in areas where forests had not grown before for quite some time. The first version of this global database comprised data for the period 2000-2012 (Hansen et al., 2013), and since then it has been continuously updated, covering currently the period 2000-2021.

The most important issue in the snap-shot model is the relation between the frequency of registrations and the rate of change of specific properties, that is the sampling of the phenomenon in time. In the case of typical changes in land use (for instance, deforestation, urban sprawl), the annual registration of states is fully sufficient because the rate of these changes is relatively low, and the comparison of snap shots received with such a frequency allows one to receive an apparently continuous registration of the phenomenon (Box 6.4). It is worthwhile adding that the resemblance of the snap-shot model to movie making is quite justified: in the end, a movie shows only the illusion of movement, being made of a huge number of snap shots taken one after another, with each snap shot being absolutely static (Fig. 6.5).

Box 6.4. Google Earth Timelapse.

Google Earth Timelapse[57] is an on-line animation presenting land cover changes on the Earth's surface thanks to thousands of satellite images of the Earth itself, acquired within the Landsat programme, here from the 1980s onwards (see also Box 5.1). For any location on the Earth, one can watch a sort of a movie in which subsequent snap shots show the land cover at a certain moment from the period 1984-2018. Many places on Earth have quite surprising land cover dynamics – for instance, the coast of Dubai with its rapid urban development or the Amazon basin with extensive deforestation.

Except for land use change studies, the snap-shot model may have various other applications. For instance, the measurements of elevations carried out every few years in areas with active erosional processes allow one to assess the changes to the Earth's surface topography that have occurred between the two consecutive measurements. Similarly, as in the case of land use change, in the case of elevation changes it is quite difficult to continuously record processes that cause these changes, because they are diverse, dispersed and their intensity varies significantly in time. It is, however, much easier to infer about the processes and their consequences based on changes of states, using a standard snap-shot model.

[56] Accessible at https://glad.earthengine.app/view/global-forest-change, accessed: May 2021.
[57] Accessible at https://earthengine.google.com/timelapse/, accessed: May 2020.

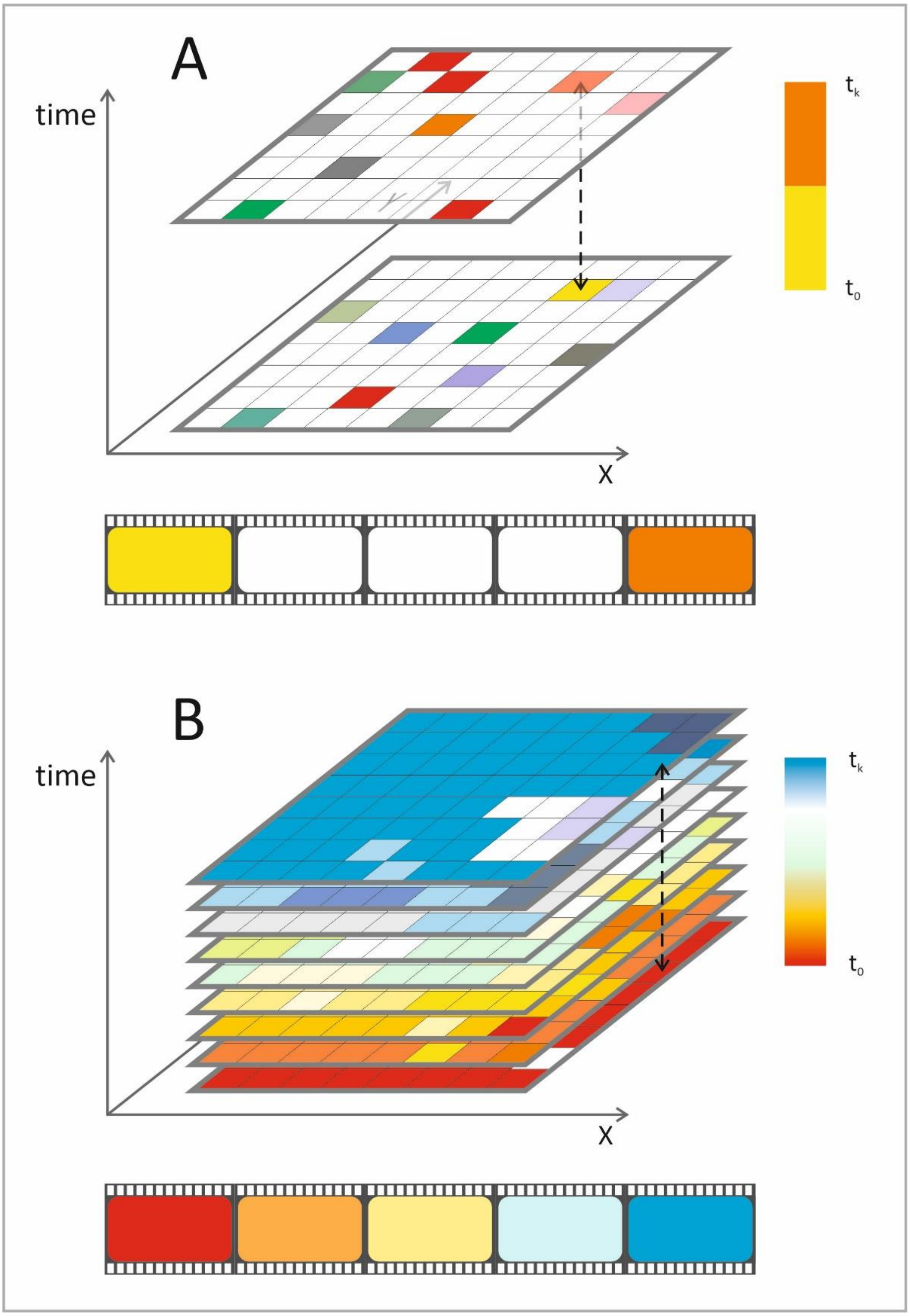


Fig. 6.5. Map overlay in the snap-shot model.

A – the snap-shot model for two registrations of states, at the beginning (*t_0*) and the end (*t_k*) of the studied period; B – the snap-shot model for a large number of registrations of states in the studied period.

A slightly different case is when the snap-shot model can be applied in parallel to continuous measurements of processes and their intensity. A good example are demographic studies. Using censuses carried out in regular intervals of time (e.g., every ten years) one may receive a complete state of the population of a certain area at a selected moment. Independently, in most countries various demographic processes that control population for a certain area are continuously registered (migrations, births, deaths). Censuses allow one to assess if the continuous registration is sufficiently reliable to receive an exact population number at any moment in between the census checks. In this situation, the snap-shot model is merely complimentary to any process-based registration and plays a role similar to inventories in shops that allow one to assess if process-based registration (supplies, purchases, returns) is carried out accurately.

In many cases the registration of state (and therefore applying a snap-shot model) is more difficult than continuous process-based measurements. This is the case with water balance research in catchments, where changes in water retention in a catchment are easier to estimate when all the processes of the water cycle (precipitation, discharge, evapotranspiration) are measured than when trying to estimate, with some accuracy, the amount of water retained in the catchment. To do the latter would require very complex measurements of surface and underground water reservoirs, technically quite demanding if a high level of accuracy is sought.

Time geography

All phenomena that are dealt with in geography occur in space and in time. Human activities are no exception: what we do has both a spatial and temporal dimension. If we move from one place to another – for instance from home to work – then we traverse some distance in space in a certain time. We may typically choose among several options of transportation means and routes, and our choices determine the time needed to reach the target. The choices may depend not only on the distance and time but also on several other factors. For instance, my travel to work in Kraków by public transport is rather complicated and takes approximately one hour with at least one change. Biking the same distance takes approximately 45 minutes. Using a car I may be able to travel to my destination in under 30 minutes, but possibly using a longer route to avoid the traffic jams in the city centre. For instance, the motorway may take me home faster, yet instead of driving 15 km via the city centre, the distance will be almost 30 km. For the takeable options other factors are also important: bad weather may decrease my chances of taking the bike, but the car may also be less likely because of the high probability of increased traffic and traffic jams. Good weather, a willingness to take some physical exercise, a pro-ecological attitude and running costs make the bike the most likely choice. If I wish to have some time to read during my journey, I will tend to use public transport. If there is anything heavy to be taken to work or brought back home, along with the necessity to call in at various places in the city on the way back, the most rational choice will be the car.

Let us pay attention to the fact that each activity, even the most simple and every day, requires us to consider spatial and temporal aspects and, based on them, make appropriate choices. These choices are motivated in many ways by our habits, our physical and economical possibilities. These varied factors that influence our activities in space and time are the basic subject of study of a certain branch of geography called **time geography**. Time geography originated thanks to the pioneering works of the Swedish geographer Torsten Hägerstrand in the 1960s and 1970s, dealing with the temporal aspects of man's spatial activities (Corbett, 2011). Taking into account how geography focuses on, and is entangled in various spatial approaches, this branch of geography is probably the one with the most surprising name. At present, however, thanks to the ubiquity of positioning devices and technologies, and huge amounts of geographic data that they collect about human activities with one of coordinates being time, time geography is with no doubt one of the most interesting and quickly developing research areas in geography itself.

Today, almost every smartphone has satellite navigation capabilities and, in this way, it allows the smartphone (and its owner) to be localised almost everywhere, with a very high frequency of measurements (Box 6.5). To find the position of the smartphone one may use also cellular or wireless networks, because the smartphone *knows* to which station it is connected at a specific moment. Regardless of the localisation technology, a smartphone or a satellite navigation receiver is localised in space and time, recording four coordinates – three related to space (*x*, *y*, *z*) and the fourth related to time (*t*). In this way, the user is positioned in a four-dimensional space-time. Sequences of coordinates x_i, y_i, z_i, t_i create a space-time path – a space-time trajectory of the mobile object.

Box 6.5. Global Navigation Satellite Systems.

Global Navigation Satellite Systems (GNSS) allow one to establish the location of a receiver on, or at the Earth's surface thanks to the simultaneous measurements of distance from the receiver to at least four satellites of the system. Measuring the distance is based on the time difference between sending the signal from a satellite and its registration by the receiver, and the known positions of the satellite at the moment the signal was sent. The first global navigation satellite system was the United States Global Positioning System (GPS) started in the 1970s, first as a military system, but already in the 1980s allowing for non-military applications. GPS became fully operational in 1995. Due to the primacy of the United States in the domain of global navigation satellite systems, these technologies and in particular the receivers are commonly referred to as *GPS*, though today three other systems are either operational or in development: the Russian *Glonass*, started quite soon after work on GPS had been initiated, the Chinese *Beidou* and the European *Galileo* system.[58]

Space-time paths can be stored using a vector data model as linear objects that can be visualised in a three-dimensional space-time cube (Mei-Po Kwan, 2004; Miller, 2005).[59] The modelling of space-time paths allows one to detect and study interactions among mobile objects (e.g., humans), because in this case checking the coincidence in space is not enough – time has to be considered as well (Fig. 6.6). In a space-time cube one may also delimit a section of space-time that is accessible for a person being in a specific place and able to travel with a certain maximum speed – a so-called space-time prism (Miller, 2005).[60] Quite soon the analysis of movements of individuals and the visualisation of its results in space-time cubes will become a standard in time geography. Actually, the barrier is less the technology of data capture and data analysis, and more privacy concerns and legal regulations pertaining to personal data protection. Where someone is, at a specific time, is undoubtedly a sensitive personal piece of information.

Retrospective and real-time studies

Still not long ago most geographic studies that were dealing in some way with time, for instance studies using a snap-shot model to assess changes on the Earth's surface, or studies of human activity in space-time carried out according to time geography concepts, had a retrospective character. In these studies researchers analysed what had happened in the past, based on measurements that had been concluded and transferred to research teams. Now various measurements (e.g., location technologies, remote sensing) and communication technologies (data transfer via the Internet) allow one to carry out research in real, or near-

---

[58] Based on https://geoforum.pl/gnss/historia and https://geoforum.pl/gnss/chronologia, accessed: June 2020; Śledziński (not dated); European Global Navigation Systems Agency, https://www.gsa.europa.eu/, accessed: June 2020.
[59] In this case, the *z* coordinate (elevation) is omitted as typically not significant for this class of problems; in this way space-time paths can be visualised in three dimensions (*x*, *y*, *t*).
[60] There is some analogy here to the light cone concept known in physics. In geography, however, the space-time prism depends on the individual capabilities of a mobile object (person), while in physics a light cone is simply defined by the light speed (https://pl.wikipedia.org/wiki/Sto%C5%BCek_%C5%9Bwietlny, accessed: June 2020).

real time. It means that researchers may analyse phenomena that have just taken place and immediately send their results to interested institutions, allowing them to act accordingly. The possibility to carry out research in real time is very important in situations where human safety is at stake. For instance, satellite imaging systems, thanks to the capability to control their orbital paths, allow one to register the state of the Earth's surface in areas affected by sudden environmental disasters and to mitigate their impacts as they unfold. Examples of real-time systems are, for instance, various early warning systems (Box 6.7).

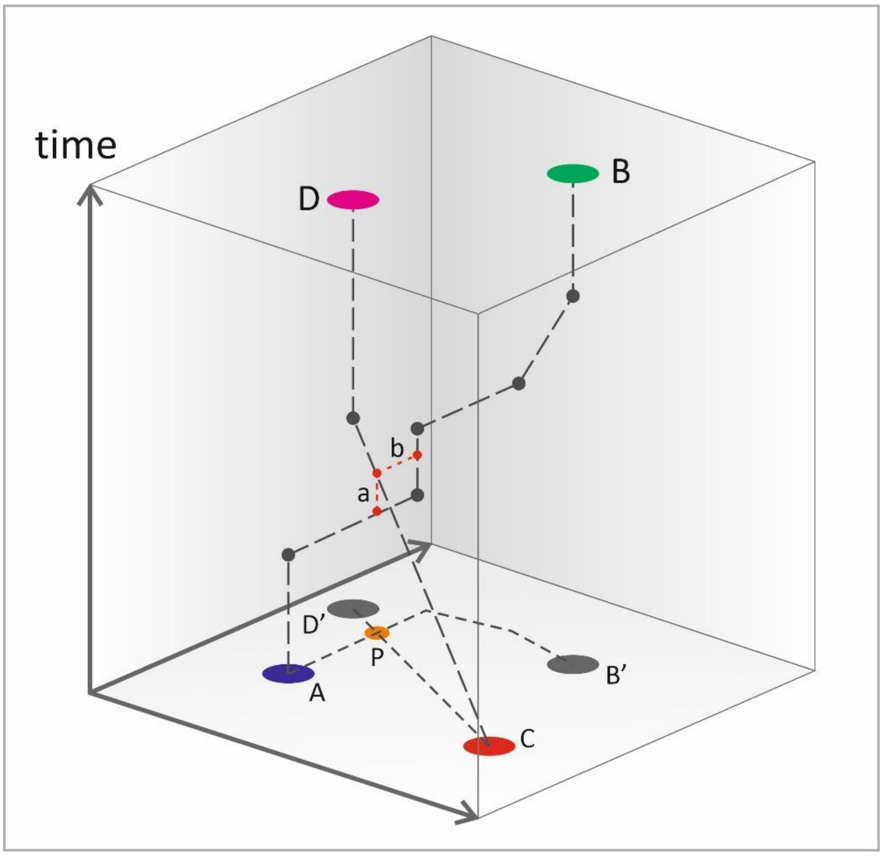


Fig. 6.6. Relations between paths reduced to two dimensions and space-time paths.

For a person moving from A to B and a person moving from C to D, paths AB' and CD' intersect on a plane in P (B' and D' are B and D projected onto the plane), thus delimiting an apparent location of the meeting of these two persons. Taking into account the relation of space-time paths, the distance between two space-time paths in P corresponds to segment a (which means that the person moving from A to B was in P some time before the person moving from C to D). At some moment in time, the distance in space between these two persons is presented by segment b. Anyway, these two persons travelling along their space-time paths have not met at all.

Box 6.7. U.S. Tsunami Warning System.

U.S. Tsunami Warning System[61] integrates systems registering earthquake occurrences, a set of ocean measurement stations and measurement stations located on the coast of the Pacific Ocean. The ocean measurement stations belong to the Deep-ocean Assessment and Reporting of Tsunamis (DART) system.[62] Each station is comprised of a bottom pressure recorder and a surface buoy for real-time data transfer to the analytical part of the system. Following an earthquake in the region, a tsunami warning is issued. The warning

[61] Accessible at https://www.tsunami.gov/, accessed: June 2020.
[62] Accessible at https://www.ndbc.noaa.gov/dart/dart.shtml, accessed: June 2020.

indicates which areas are endangered by a tsunami and when it may occur. The warning is updated based on the measurements of ocean motions coming from DART stations. Besides the location and time of a potential tsunami, updates include also information about the expected height of the tsunami wave and other collateral effects (e.g., strong currents that may threaten the sailors, fishermen and swimmers) and recommendations for local authorities.

In the Introduction I have presented a draft definition of geography, later on explained in more detail in Chapter 1. Now, after presenting the issue of time in geography, the characterisation of the subject of geography can be expanded. It was said earlier that **the subject of interest and research of geography are various phenomena: objects and their properties, processes and events, distributed in space and time, occurring on the Earth's surface and of importance to humans. Geography studies the spatial patterns of properties of first level objects using either a discrete or continuous approach, and the spatial arrangement of second level objects and events, taking into account the discrete approach.** Now I may add that **through studying processes: flows of matter, energy and information occurring in space, and the movements of mobile objects including humans, geography focuses on interactions among objects and the changes of their properties in time that result from these flows and movements.**

7. Points, lines and polygons

Key words: point, line, polygon, spatial autocorrelation, spatial interpolation, network, region, spatial pattern, zonal analysis, modifiable areal unit problem

It may appear rather strange that one of the most obvious representations related to discrete objects of various dimensionality is dealt with relatively late in this book. Points, lines and polygons, or 0-, 1- and 2-dimensional representations of discrete objects are as old as maps themselves, and one may assume that the very first maps were drawn using simple symbols for small objects, like trees, linear objects like rivers or areal objects like islands or lakes. These seemingly simple methods of representation conceal a vast range of interesting issues, and this is why they are discussed only now, when the reader is already aware of various contexts that are important for geography and may easily refer to them.

Points, lines and polygons are most frequently related to the vector geographic data model – yet the choice of this model is not necessary, especially when it comes to points and polygons, for which a raster representation is as good as a vector one (Fig. 7.1). Lines are an exception, because the raster geographic data model is very unwieldy and therefore rarely used to represent 1-dimensional objects (see also Chapters 1 and 3).

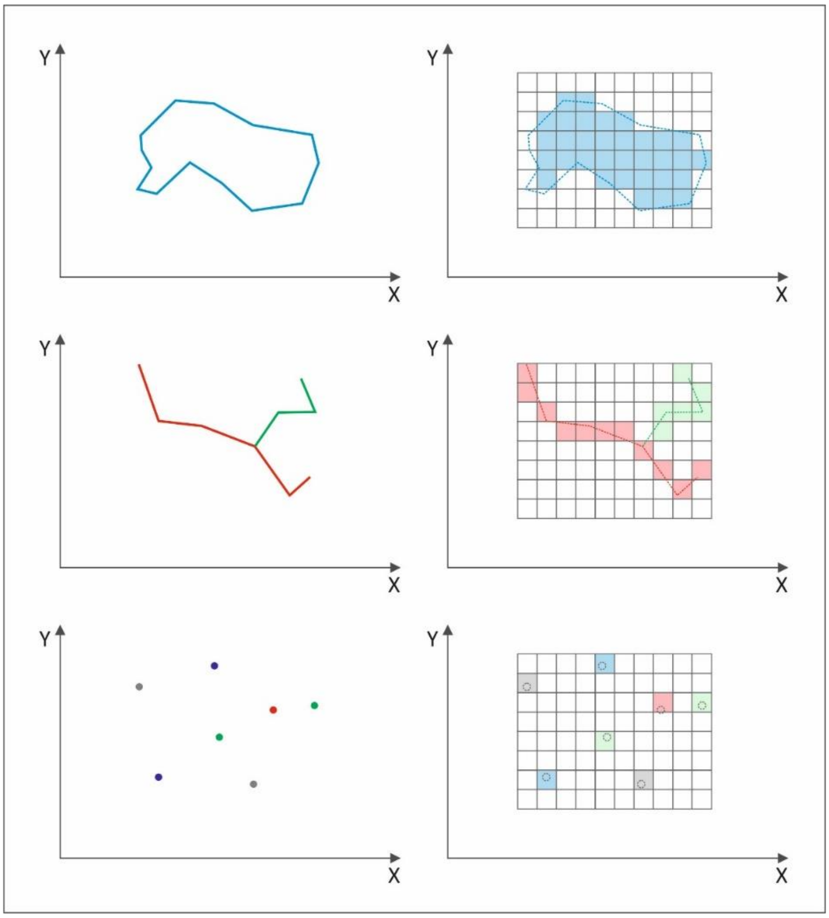


Fig. 7.1. 0-, 1- and 2-dimensional objects represented by vector and raster data models.

Points

Points are used to represent relatively small objects, small enough that for a given scale of representation and analysis their dimensions are negligible. They are mostly second level objects (see Chapter 1), for instance buildings, trees, monuments or springs. They may represent also various locations that are interesting for some reasons, e.g., that of a measuring station on a river, an elevation point or an address. Points may also represent events that occur in specific places. It is important to note that the occurrence of a second level object in a certain location at a given time may be an event – in this way a point may represent a mobile object and its ephemeral location. The distinction between the locations of mobile objects and events is relatively fuzzy: for instance, a car accident is typically an event related to the occurrence of two (or more) mobile objects in the same location and at the same time. Points may also represent characteristic places of continuous fields: in the case of elevation, these might be peaks, sinks or passes. Finally, a point may represent a set of other points or an object of a higher dimensionality, like a polygon.

The point representation of small second level objects and events does not involve anything of particular interest – a point is simply a pair of coordinates, denoting as accurately as possible the location of the object or event. The two latter cases listed above require, however, some more explanation. Characteristic locations of continuous fields can be received using methods of focal analysis (see Chapter 4), for instance, peaks are points (single cells of a raster matrix) with values higher than the values in their neighbourhoods. Points representing objects of a higher dimensionality (or sets of points) are e.g., centroids, geometrical centres of such objects or sets. In the case of polygons with an irregular shape, their geometric centre may fall outside the polygon (Fig. 7.2). If a point represents a set of points, its coordinates may be calculated as a mean of coordinates of all the points of the set, however other statistical measures can be used as well (e.g., a median).

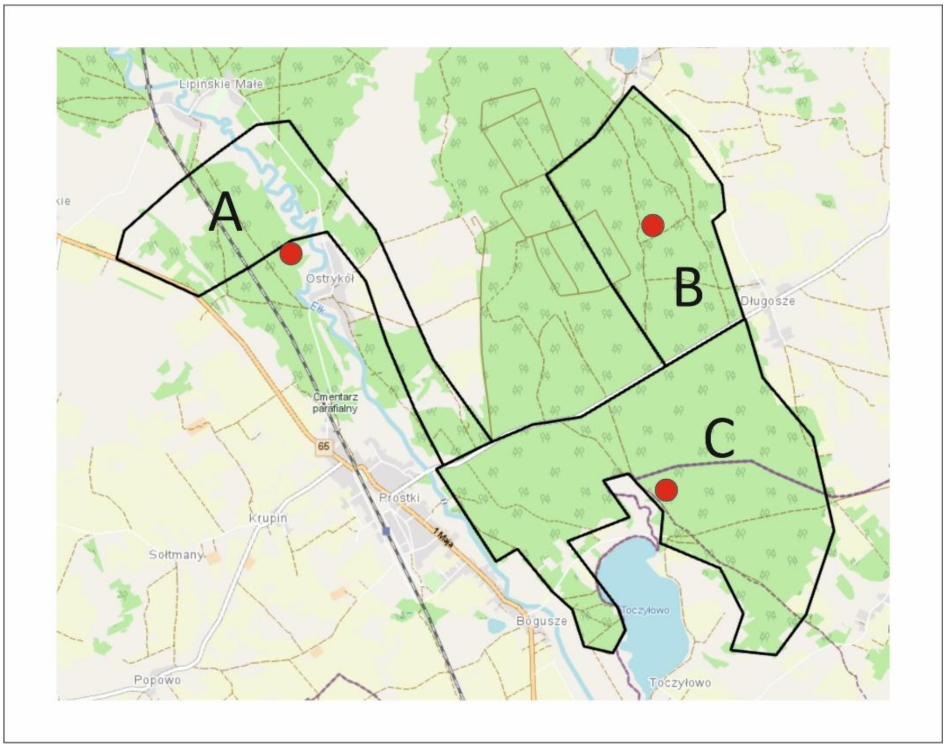


Fig. 7.2. Polygon centroids.

The centroid of polygon A is located outside its boundaries.

Background: OpenStreetMap, https://www.openstreetmap.org.

Spatial point distributions

Spatial point distributions, or point patterns may provide valuable information about the nature of phenomena represented by points. At the general level, point patterns can be random or non-random, the latter can be further divided into regular or clustered distributions (Fig. 7.3). The analysis of point patterns can be carried out based on point density distribution or on the distances among points.

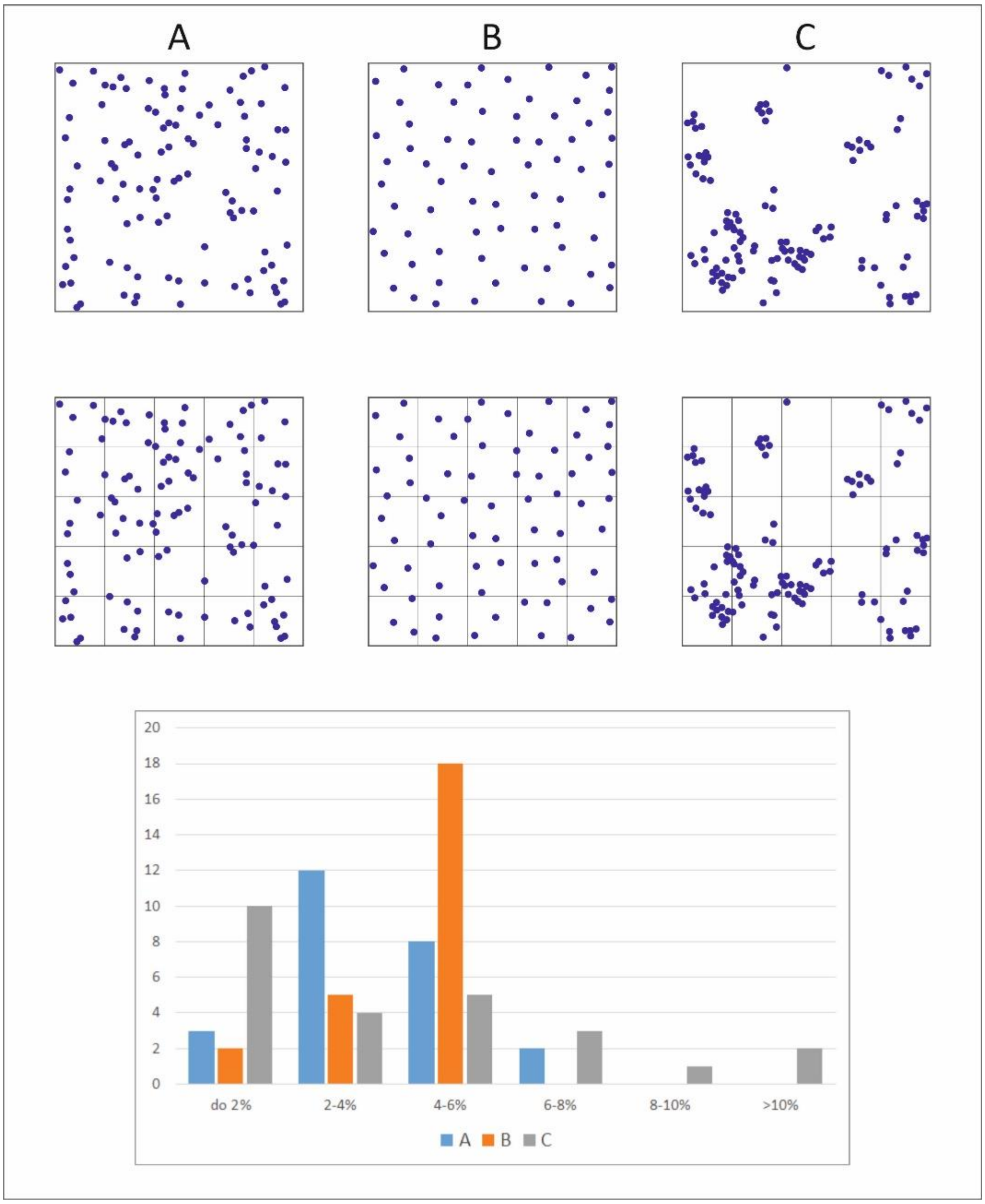


Fig. 7.3. Point patterns.

A – random pattern; B – regular pattern; C – clustered pattern. If the entire area where points are located is divided into 25 identical areal units then each pattern has its own typical frequency distribution (the diagram below). For a regular one, most areal units have around 4% of the total number of points (that is close to the mean value per areal unit) and for a clustered pattern areal units may have both very low and very high numbers of points. For a random pattern, the frequency distribution is approximated by the Poisson distribution.

Source: Baddeley (2010).

Point density analysis allows one to find places where points occur more or less frequently – which leads to the question as to what is the cause that the points, or rather phenomena represented by these points, assemble in specific places. Point density analysis relies on relating the number of points in a certain area to the size of this area. In such a case it is critically important to what area a certain set of points is related, or, simply, what is the areal unit used to calculate density (areal units were also discussed in Chapter 2). There are two basic options: stable or movable areal units. Stable units, for instance a regular square grid, divide the entire analysed area exclusively and exhaustively, allowing one to calculate the density for each areal unit. Movable areal units also allow one to calculate the point density for any location in the analysed area (as the division will be exhaustive), but areal units may overlap in various ways, as the division is not exclusive. In this case, point density can be calculated for identical squares, but the squares may be constructed in such a way that the locations for which we plan to calculate the point density become the centres of these squares. Instead of squares, other shapes can be used, e.g., hexagons or circles. Densities calculated for the overlapping areal units of any shape may be similar (Fig. 7.4).

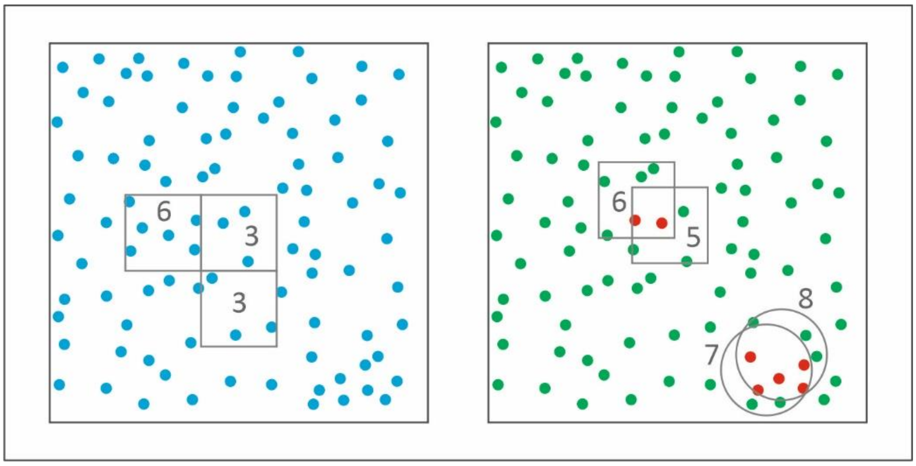


Fig. 7.4. Point numbers and density for stable (left) and moveable (right) areal units.

For moveable areal units, some points may fall in various areal units (red dots). For stable areal units, which exclusively and exhaustively divide the studied area, each point falls in exactly one areal unit. Based on the number of points in the areal unit and the size of the areal unit, point density may be computed. If point density is assigned to the centre of an areal unit, then such a calculation allows one to receive a continuous density representation, with density based on movable areal units visually more appealing than that based on stable units. It is worth noting that there is some similarity between moveable areal units and the moving window analysis described in Chapter 4.

The specific point pattern we are dealing with may be also assessed based on point distance analysis for a given set of points, because the distance distributions for a random set of points differ from distributions characteristic for regular or clustered patterns. Such a distance analysis may include distances between nearest neighbours or distances for any pair of points of a given set. For a regular pattern and the nearest neighbour method some specific distances dominate, e.g., for a regular square grid these are distances equal to the side of the square. For a clustered pattern relatively small distances will dominate, as almost every point of the set will have a neighbouring point located rather close. In a random pattern, there will be both small and large distances.

Seemingly trivial point pattern analyses form one of the fundamental methods of spatial analysis, with applications in various scientific disciplines, not only in geography. This is because points and their locations may refer to various objects: rare plants or animals, springs, rock formations, caves, monuments, wells, geographic names, services of a specific kind and various events – road accidents, earthquakes, meteorite falls or homicides. One of the scientific fields that extensively uses point patterns and their analysis is crime geography, introduced briefly in Chapter 6. Crimes are events that may be usually located quite accurately in space, and clusters of specific crimes in various places may have several causes. Careful analysis of crime patterns may therefore lead to efficient preventive actions. Another interdisciplinary field of science where point pattern analysis is of great significance is medical geography, in which morbidity or deaths related to a specific disease, and their locations are studied. One of the best known cases of point pattern analysis, frequently referred to in spatial analysis or GIS textbooks (though the case was described long before the first computer was constructed) is the analysis of cholera deaths during the epidemic in Soho, London, in the mid-19th century, carried out by Dr. John Snow (Box 7.1).

Box 7.1. John Snow and his famous map.

The cholera epidemic in Soho, London, broke out in 1853. John Snow studied its advances annotating on a map of the district subsequent deaths, easy to record due to the coffins put out every day in front of buildings. Each dash on the map was one death in a building. In some cases, with many deaths in a particular building, dashes drawn one over another form not fully legible bars. John Snow's map[63] shows clearly that cholera deaths cluster in the centre of a district delimited by Regent, Oxford, Dean, Coventry and Marylebone streets. Assuming the distribution of population in the district was relatively uniform at this time, John Snow searched for a cause that led to such a distribution. As the most likely cause he identified a water pump in Broad Street, more or less in the centre of the afflicted district. He decided to block the water pump, and this in the short time ended the epidemic, providing a solid confirmation for the entire analysis. Worth noting is that although the original map shows each death with a single mark, today it is easier to use points rather as a representation of the addresses of buildings where deaths were recorded. Then, each point, except for its location (e.g., an exact street address) will also have an attribute – the number of deaths recorded at the specific address. John Snow's data, converted to vector format (points with the *number of deaths* attribute) can be today found, downloaded and used for any analysis.[64]

Spatial interpolation

Representing point density based on a specific point pattern in fact converts a certain number of discrete objects (points) into a continuous field, made up of calculated density values. It is important to note that for this particular field its property – that is density – is only apparently continuous. While the density can be assessed in any location in space, under some assumptions (see also Fig. 7.4) it cannot be measured at a location, because an object either is or is not at any location, and cannot be at any location only partially, reflecting then the calculated density values that may be fractional or expressed by rational numbers. A good analogy here is probability theory – for instance, if the chance of tossing tails with a coin which is not fair is

[63] A lot of information about John Snow can be found on a web portal elaborated by R.R. Frerichs, available at www.ph.ucla.edu/epi/snow.html (accessed: July 2020). The map can be found there as well, in a part of the portal dedicated to the cholera outbreak in Soho: www.ph.ucla.edu/epi/snow/highressnowmap.html (accessed: July 2020).
[64] John Snow's map data in the shapefile format can be downloaded from the web page *Robin's blog: John Snow's Cholera data in more formats*, accessible at blog.rtwilson.com/john-snows-cholera-data-in-more-formats/, accessed: 2020. One can create these data alone, based on the map, especially as the pattern of streets in this part of London has not changed significantly since the mid-19th century, and John Snow's data can be therefore easily related to the contemporary geographic data presenting streets and buildings.

0.75 this in no way means that in one instance we receive 0.75 of tails and 0.25 of heads – rather, that in a series of four instances we may expect tails three times and heads once.

But there are some properties of space that can be – and they frequently are – measured in specific locations, and in addition we know that these properties may have other values in locations where no measurements have been done. In these locations the values of these properties can be estimated using **spatial interpolation** methods. These methods show some similarity and their effects may look comparable to the density analysis, but they have different assumptions and are applied in other contexts.

Spatial interpolation may be defined as estimating the value of a certain property in a location with known co-ordinates, in which this value is unknown, based on the known values of this property in a finite set of points with known co-ordinates (Fig. 7.5). Let us assume that such a property is air temperature, and for a relatively flat area 100 x 100 km we measure the air temperature at 25 stations, more or less evenly distributed over the study area (it is quite easy to find out that the distance between stations will be on average 20 km). We wish, however, to estimate the air temperature for a regular square lattice with a cell side of 5 km, that is for 400 points for the entire study area. Because air temperature does not vary significantly over such distances if the area is not mountainous, the assumption that for each node of the lattice the estimated temperature is very close to the temperatures measured at the four nearest stations seems rational. Such an assumption we may easily convert into a formula, e.g., in such a way:

$$t_i = (t_{i1} + t_{i2} + t_{i3} + t_{i4}) / 4 \qquad [7.1]$$

where t_i is the estimated air temperature for the node *i*, and t_{i1}, t_{i2}, t_{i3}, t_{i4} are the air temperatures in four stations P_1, P_2, P_3, P_4, located closest to the *i*-th node of the 5 km square lattice (Fig. 7.5). This simple formula may become more complicated – for instance, taking into account not only the air temperatures in the four nearest stations but also their distances to the *i*-th node of the lattice, assuming that the closer a station is located to the *i*-th node, the more the measurement at this station influences the estimated value. The formula to calculate the value might then look like this:

$$t_i = (t_{i1} / d_{i1} + t_{i2} / d_{i2} + t_{i3} / d_{i3} + t_{i4} / d_{i4}) / (1 / d_{i1} + 1 / d_{i2} + 1 / d_{i3} + 1 / d_{i4}) \qquad [7.2]$$

where d_{i1}, d_{i2}, d_{i3}, d_{i4} are the distances of the *i*-th node from the four nearest stations P_1, P_2, P_3, P_4.

Whether estimating values using spatial interpolation makes any sense depends on what **spatial autocorrelation** the studied phenomenon shows. Spatial autocorrelation defines how the similarity of values of a certain property changes in relation to the distance between the points where the values are measured. Spatial autocorrelation is expressed with a correlation measure of values ranging from -1 to 1. A high (positive) autocorrelation means that the values measured in points close to each other are similar, while those values measured in distant points will significantly differ. A low (negative) spatial autocorrelation means that extremely different values may occur at small distances between measurement locations (Gangodagamage et al., 2008; Fig. 7.6).

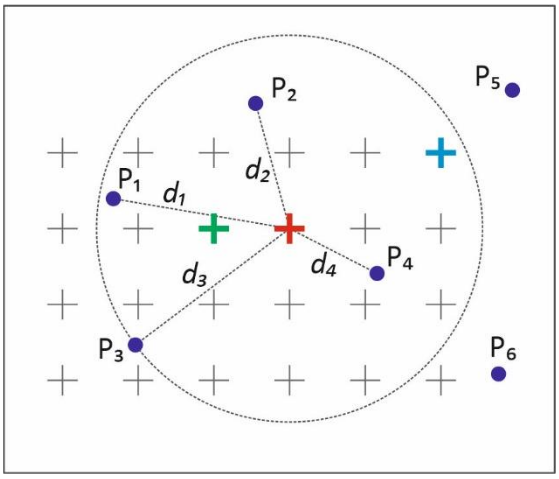


Fig. 7.5. Spatial interpolation.

The value for the lattice node marked red depends on the values measured at the four nearest measuring stations P_1, P_2, P_3, P_4. If the 7.1 equation is applied, then the value calculated for the neighbouring lattice node marked green will be exactly the same as for the red one, because the green node has the same four nearest measuring stations P_1, P_2, P_3, P_4. A different value will be calculated for the lattice node marked blue, as the four nearest measuring stations in this case are P_2, P_4, P_5, P_6. However, if the 7.2 equation is applied, then the calculated values for the red and green lattice nodes will differ, because the distances of the measuring stations P_1, P_2, P_3, P_4 to these lattice nodes differ.

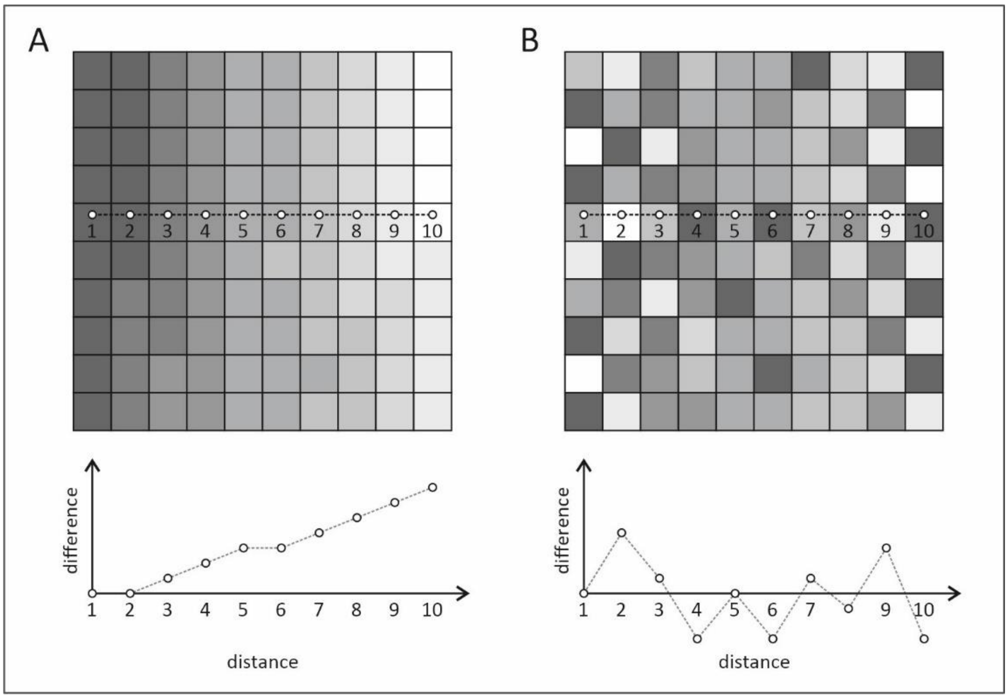


Fig. 7.6. Spatial autocorrelation for two raster layers.

Values of a certain property are shown in a grey scale (the lighter the shading – the higher the values). Left: values displaying a high spatial autocorrelation; right: low spatial autocorrelation. The graphs below show the differences between the value in the middle row, leftmost cell (marked 1), and values in cells 2-10. For the raster layer on the left, the differences increase with distance, which means that values close to each other are similar, and distant values differ significantly. For the raster layer on the right the differences do not depend clearly on the distances between the cells – both high and low differences, negative and positive, may occur for neighbouring cells.

It is no exaggeration in saying that geography is meaningful thanks to the positive spatial autocorrelation – that is, thanks to the similarity of values measured close to each other, and the differences between the measured values increasing with the distance between the locations, for most of the properties and phenomena in space (Box 7.2). Because spatial interpolation is, to some extent, a prediction of how values of a certain property are distributed in space, therefore it has to be preceded by assumptions as to the spatial autocorrelation of the property. Any lack of, or negative spatial autocorrelation excludes reliable estimates of values between measurement points with spatial interpolation.

Box 7.2. Tobler's First Law of Geography.

That most geographic phenomena have positive spatial autocorrelation is expressed by Tobler's First Law of Geography (TFL): *Everything is related, but near things are more related than the distant ones*. This, at the first sight apparently humorous definition, expressed for the first time by the American cartographer, Waldo Tobler, in a paper *A Computer Movie Simulating Urban Growth in the Detroit Region* published in 1970 in the journal *Economic Geography*, reveals a profound truth: the predicting and modelling of spatial distribution of phenomena is possible only when they are somehow structured in space. Such a structured distribution results from positive spatial autocorrelation that luckily characterises many different phenomena. The first formulation of TFL went initially almost unnoticed, only later in the 1990s with advances in GIS, did the popularity of TFL increase – both in geography and beyond (Sui, 2004).

Finally, it is worth adding some important remarks about spatial interpolation. Spatial interpolation may employ various methods, both mathematically simple and quite complex ones. **Geostatistics** is a branch of statistics that develops an entire group of relatively complex, yet efficient methods of spatial interpolation, referred jointly to as kriging. These methods estimate the spatial distributions of selected properties based on spatial autocorrelation models for these properties constructed from the set of measurement points.[65]

Because each interpolation method estimates the values of properties in any point of a selected area, and these properties can be measured in the same points, therefore regardless of the method used, a standard test for an interpolation method is comparing the interpolated (calculated) value at a point to the value measured in the same point, for a random sample of points. In the example shown in Figure 7.5, this may be done in such a way that for 10% of the grid points for which the values were calculated, the values are also independently measured. If the calculated values are similar to the measured values then the spatial interpolation model is well matched. If not – another interpolation model should be tested.

As noted above, spatial interpolation typically refers to values that may be measured in any point of a certain area, for instance like temperature, elevation, air pressure. In some cases, however, spatial interpolation may be applied also to properties that are not point-measurable, but may be estimated using density-related methods or approaches used to compute various characteristics of areal units (see also Chapter 2). For instance, let us assume that for each *powiat* (county) in Poland we calculate the annual per capita income, and the computed value refers then to the geometrical centre of each *powiat*. In this case, we receive points with known coordinates and with the values of a property allocated to points. Although computed, and not

[65] The term *kriging* originates from the name of Dani G. Krige, the South African mining engineer and statistician who employed geostatistics to estimate the amount of gold in ores, based on values measured in sparsely distributed drillings. Kriging was later developed in the 1960s by the French mathematician Georges Matheron, and now is available in spatial analysis software (based on Wikipedia, accessible at https://en.wikipedia.org/wiki/Kriging, accessed: July 2020).

measured, the values may be used to carry out spatial interpolation. As a result, instead of one value for an areal unit, we receive rather a continuous spatial distribution of the property in question, with values varying among the centres of areal units (Fig. 7.7). Whether such an approach is justified and what the consequences of choosing specific areal units to compute the values of the property being interpolated are will be discussed later, in the section relating to polygons.

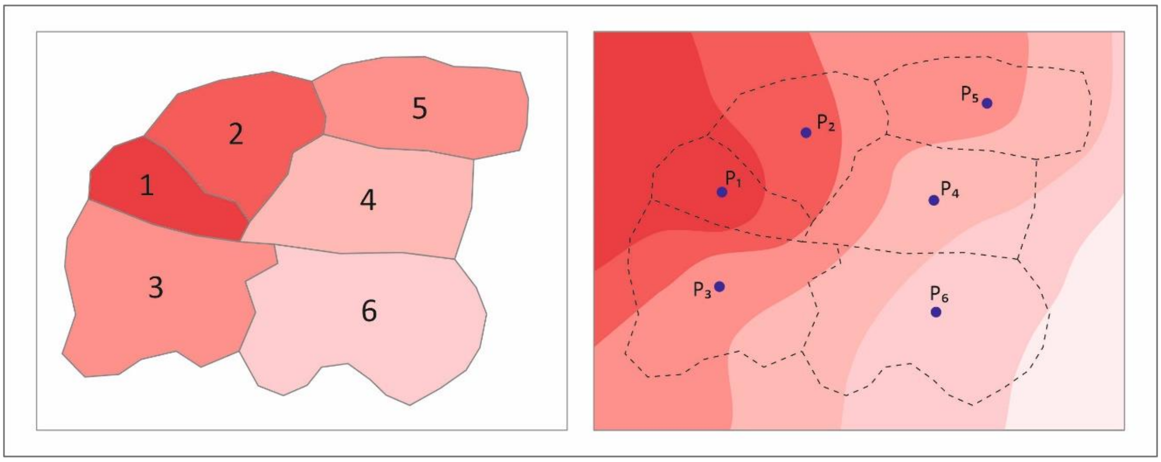


Fig. 7.7. Spatial interpolation of values of a certain property computed for areal units.

Left: areal units 1-6 (e.g., *powiats*), with the highest value in areal unit 1, and the lowest in areal unit 6. Right: each areal unit has its geometrical centre (points P_1, ..., P_6), assigning values for each areal unit to respective points allows one to carry out spatial interpolation that leads to a continuous distribution of the selected property. Worth noting is that after interpolation each areal unit has values of a certain range although from the way it is calculated one single value should characterise an entire areal unit.

Lines

Lines represent discrete objects with a width small enough, for a given scale of representation, to be negligible in most applications. This happens, for instance, when some discrete objects become transportation paths for matter, energy or information – regardless of whether the movement along the path is caused by natural forces or reflects certain human activities. As most frequently transportation paths for matter, energy and information form more or less complex networks, the best model for their representation is a topological vector geographic data model (see also Chapter 3) in which lines connect to each other in well-defined nodes, and have a specific direction which may determine the acceptable direction of movement. A topological network model allows one to easily identify connections based on information encoded in the topological table with no need to analyse the often complex geometry of the lines building the network (see also Fig. 3.7).

A simple example of a network are river networks. Typically, river networks have two features: as water flows only down, the direction of flow is strictly defined; in addition, connections within networks are such that streams originate in springs, and then connect to each other, merging into one outlet (Fig. 7.8). It means that in river networks matter is only relocated from springs to the outlet, and, for instance, a transfer of matter from one spring to another is forbidden. Each intersection of two lines forms a node: in most cases, this is the connection of two streams into one. Much more rare, and confined mostly to river deltas, are nodes where bifurcation occurs and one stream splits into two branches. Even more rare are natural channels connecting two independent river network systems, with the possibility of bidirectional water flow. An even more complicated situation can occur in karst areas because of underground channels that may – when projected onto a plane – visually intersect surface streams, but with no actual connection, because the water flow occurs underground, at a certain depth and in isolation from what is on the surface.

Man-made transportation networks are even more complex than natural river networks (Fig. 7.8), mostly because they do not impose limitations as to the direction of movement along the network segments, and also because mobile objects (e.g., people, cars) use various technologies that allow them to move with various speed, while in the case of river networks gravity is the only cause of water flow. Therefore the modelling of movement in the transportation networks requires one to not only consider the geometry of the network, but also those properties that condition the speed of movement. In practice, this is a task we do every time we plan a trip from any starting point (A) to a chosen destination (B). We look not only for the shortest path connecting A and B, but also check the travel time, in many cases selecting the path that is the quickest, and not the shortest. In many cases, we may also take into account the cost of the journey (Fig. 7.9).

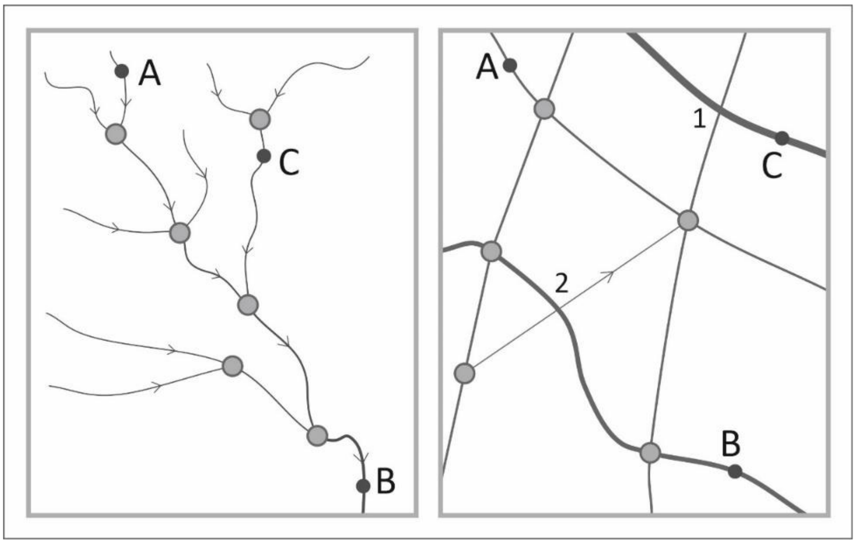


Fig. 7.8. Examples of various networks.

Left: river network, in which each vector line (arc) has a strictly defined direction, relating to the flow of water. Water may flow from A to B and from C to B, but any flows from A to C, from C to A and from B to either A or C are impossible. All line intersections are nodes (grey circles). Right: a road network, in which – except for the line with an arrow – movement may occur in both directions of any network segment. Not all intersections are nodes (grey circles) – two intersections marked with numbers are not topological intersections (e.g., one road may pass over another, with no possibility to take a turn). It is possible to drive from A to B or from B to A, but – taking into account the part of the network shown in the figure – it is not possible to travel from A and B to C, and from C to A or B.

Though a target data model for networks is a topological vector data model, in some cases line objects being part of a network may be created based on the analysis of raster data. River networks evolve through interactions of atmospheric, hydrological and morphological processes with terrain surface, with diverse elevations and geological structure. Flowing water uses (and shapes) terrain depressions, therefore knowing the terrain elevations of an area, one may try to model the geometry of line objects that build a river network. The accuracy of the model will depend on knowledge of various properties of the terrain, e.g., permeability. Close to the watershed we will not find streams, only walking further down are underground water reservoirs rich enough to feed springs and perennial streams. These streams flow down along the steepest slope line eroding the terrain surface, and connect with others, towards the outlet. The focal functions presented in Chapter 4 allow one to delineate the directions of the steepest slope based on a raster DEM, and therefore they allow one to build models of river networks, assuming a certain ground permeability. It is sufficient simply to accept that a stream starts in a cell which accumulates a water flow from a defined minimum number of cells located above. In this way, DEM-based modelling of a river network is just one example of how linear objects are delineated from continuous fields. As for terrain elevation, these objects are flow lines, which together with ridge lines may be used to characterise terrain morphology and patterns. Similar methods of analysis of raster data are used to delineate corridors for animal movements that may be used to assess a so-called habitat connectivity (Box. 7.3).

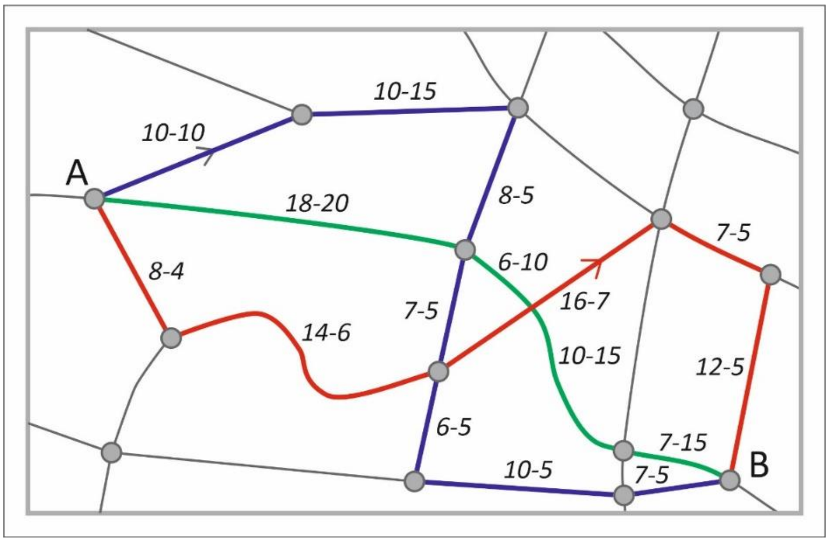


Fig. 7.9. Connections in a network.

Three paths from A to B differ with respect to distance (the first value in the pair describing each line segment of the network) and the travel time along the segment (the second value in the pair). For the green path, the distance is 41, and the time is 60. For the red path, the distance is 57, and the time – 27. The blue path has a distance of 59 and a time equal to 50. The green path is the shortest, but it takes the longest to travel this route. The blue and red paths are of similar length, but the red path is definitely the fastest one. Cost of travel can be an additional variable (e.g., the cost related to toll payments) – if the red path has sections for which it is necessary to pay, one may decide that the blue or green variants are more convenient. In addition, there are also other paths linking A and B that might be considered.

Box 7.3. Corridors and landscape ecology.

Corridors are relatively long and narrow strips of land that differ from their surroundings. Defined in this way, corridors are one of three functional elements of landscape, in a conceptual framework introduced to landscape ecology by Forman and Godron (1986). Two other elements are patches and a matrix (background). In landscape ecological studies, corridors are used to model the paths of animal movement between habitats that are patches with favourable conditions for specific animal species. In such a sense, corridors are analogous to the transportation lines constructed by humans. However, because corridors do not have any related infrastructure, their course can be delimited based on a spatial analysis of landscape structure (in particular, land cover and elevation) or on tracking animal movements using telemetry (Miller et al. 2019). As for the latter method, corridors are delineated as zones with a higher intensity of movement. As for the former, modelling methods are similar to the methods used to build river networks from DEMs where directions of the steepest slope are looked for, yet in the case of corridors, a continuous field of the cost of movement is analysed instead of the DEM. In such a layer each cell stores the value of cost (or time) needed to move through the cell by an individual representing a hypothetical species. The analysis relies on finding paths connecting habitats in the study area for which the total cost is the lowest.

Polygons

Polygons represent any fragments of the Earth's surface, delimited in any way – that is, first level objects (see Chapter 1). They may have more or less clear natural boundaries – e.g., islands, lakes, continents, catchments, climatic zones, habitats. They may also be fragments of the Earth's surface with boundaries defined by humans – countries, administrative units, national parks. Finally, polygons may represent any other fragments of the Earth's surface, for instance with geometrical boundaries, such as an area bounded by meridians with longitudes 10 and 20 degree east and parallels with latitudes 50 and 60 degree north.

Representing fragments of the Earth's surface with polygons is an extremely important topic in geography, rich in content and with a very long history, so describing it in a short subchapter is not possible. This is why the focus here is on the crucial issues related to what are **regions**, how their boundaries are defined and how they are used in geography and other scientific disciplines that deal with spatial patterns of phenomena occurring on the Earth's surface.

Regions

Region is one of the oldest geographical concepts, and, as with many concepts with a long history it has many definitions. In this book a region means a spatially coherent fragment of the Earth's surface, with boundaries delimited according to some accepted criteria. Typically when a region is delimited it is assumed that it shows a homogeneity for some of its properties, while beyond its boundaries properties have other values than those in the region. Sometimes the criteria are so obvious that region boundaries can be delimited without any doubt, for instance in case of an island (like Crete), whose boundaries are clear and stable, or a country (like Romania), with boundaries clearly defined and delimited with high accuracy. A property of *being Crete* (or *being Romania*) is a property of a particular region which does not occur outside the region. In many cases, however, criteria used to delineate region boundaries are complex, and their application in practice may lead to ambiguous results. This is quite common in the case of natural, historical or cultural regions of various sizes and hierarchy levels, with boundaries depending on which criteria are selected, how they are weighted and how the spatial differences of properties are assessed as for their importance.

The problem of region boundaries becomes even more complex when criteria do not refer to spatial patterns and properties that constitute regions, but rather to spatial interactions that may allow one to delimit fragments of space with specific functional relations. The boundaries of such functional regions are commonly quite difficult to be clearly defined and delimited. A classic example of such a region is the so-called isolated state (German *der isolierte Staat*) described in the 19th century by Johann Heinrich von Thünen. An isolated state is an idealised region, functionally cut-off from any external interaction and ideally homogenous with respect to environmental properties. In such a region a specific land use pattern emerges around the centre that is the major trading place. This pattern is governed by the distance from the centre, costs and transportation time (Crosier, 2001). In a real situation, when a functional region attached to a centre is never fully cut-off from external interactions, the boundaries of a region are difficult to delineate, because the influences of the regional centre and other centres located beyond the region's boundaries interfere. Therefore, if several criteria of regional delimitation are applied, various boundaries may arise in consequence. One of the frequently applied criteria is how people commute – in such a case a boundary between two functional regions of two major centres is a zone from which a similar number of people commute to each centre (Fig. 7.10).

If regions are considered to be separate and real fragments of space, building clear hierarchical systems expresses then geography's desire to unambiguously categorise the space and describe it as accurately as possible. Such an approach has several cognitive values and, following the assumptions of scientific realism, it accepts that delimited regions represent real entities. Its disadvantage, on the other hand, is that it is time-consuming, focusing on the most accurate delineation of regional boundaries as possible and on searching for delimitation criteria that might be widely accepted. For many researchers, however, a region is not a representation of a real-world entity with well-defined boundaries that may be studied, but rather a useful tool to study a spatially variable Earth's surface. This approach follows the position of instrumentalism,[66] within which researchers will not identify regions as real-world entities but construct them in some arbitrary ways to facilitate the study and description of selected properties and spatial interactions. Here a region can be any delimited fragment of the Earth's surface, for instance with simple geometrical boundaries.

To some extent a choice between scientific realism and instrumentalism in region delimitation depends on a discrete or continuous perception of space and of its properties (see Chapter 1). A discrete approach favours perceiving space as a mosaic of regions of various significance and size, because the discrete approach focuses on boundaries, discontinuities and, somehow by default, results in partitioning space into regions that are classical first level objects. On the other hand, in a continuous approach regions (or their boundaries) are redundant, because the space is filled with properties spatially varying, and boundaries are not necessary to perceive or study them.

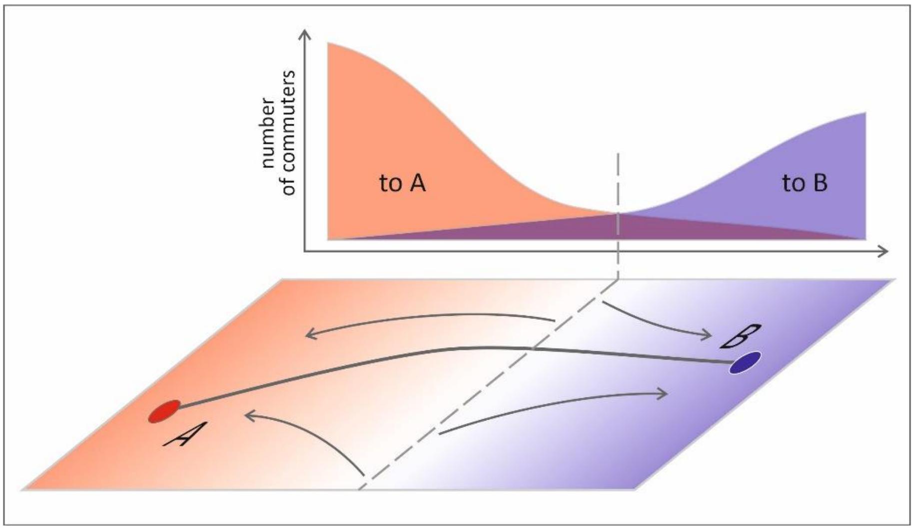


Fig. 7.10. Delimitation of functional regions.

The diagram presents the number of commuters travelling to cities A and B, along a major communication line connecting these two cities. There is a zone between A and B where the number of commuters to each city is similar – this zone is a boundary of the functional regions of cities A and B (dashed line). The areas on one side of the boundary tend to be attracted by A, and on the other side – by B.

[66] As stated in Chapter 1, realism and scientific realism accept that the real world exists, can be a subject of scientific inquiry, and that our best theories adequately describe the world. Antirealism is a generic term for various views that question at least some entities that are of interest for science to be real, or consider this issue as a purely academic problem. Instrumentalism, one of the positions of antirealism, regards theories as instruments to predict observable phenomena or tools to systematise observations (Chakravartty, 2017).

Spatial structures

Delimitation of fragments of the Earth's surface using a set of criteria results in a mosaic of regions that catch and characterise spatial distributions of some properties, representing a spatial structure (sometimes referred to as a spatial pattern). An example of such an approach are studies of land use and land cover spatial structure. Here, regions with relatively homogenous land use and land cover[67] are represented by polygons (2D objects). The geometry of these polygons is such that their entire set covers exclusively and exhaustively a certain part of the Earth's surface (Fig. 7.11). Regardless of the geographic data model, raster or vector, the spatial structure in such a case may be described with various metrics related either to the area of polygons, their shapes, spatial relations (neighbourhoods) or the spatial distributions of the various properties in the polygons themselves. One of the metrics depicting single polygons is the **perimeter-to-area ratio** for the polygon, normalised by the perimeter-to-area ratio for a circle with an area equal to the area of the polygon. On the other hand, the **density of boundaries**, that is the sum of the lengths of boundaries per unit area, is one of the metrics that refer to the spatial structure of a larger area with many polygons representing regions.

Such a model of spatial structure is a discrete one, and it requires space to be divided exclusively and exhaustively into regions, represented by polygons. If the properties used to delineate regions vary in space in such a way that there are areas where these properties vary only slightly, and zones where these properties change very quickly from place to place, then such a discrete model is adequate (Fig. 7.12). In such a case, the boundaries between regions will run along zones of quick change of properties, while regions will comprise areas in which the values of properties remain relatively stable. If, however, properties vary in space in such a way that gradients (that is, how quickly values change when moving from place to place) do not vary significantly, then, although the division of space into regions is possible, and regions may be represented by polygons, such a discrete division can hardly be considered a useful and sensible representation of the spatial structure.

Modifiable areal unit problem

Regardless of how we approach the problem of the delimitation of regional boundaries, any fragment of the Earth's surface, that is a first level object, can be considered to serve in the analysis as an areal unit. Areal units are represented by polygons recorded either using a vector data model or a raster data model. Here it is worth saying that the concept of an areal unit has been quite frequently referred to in this book – for instance, areal units were discussed in Chapter 2, and some methods of analysis using areal units have been already mentioned at the beginning of this chapter, in the section dealing with points (see Fig. 7.4). With a set of points representing second level objects we may – for any set of polygons representing areal units – calculate the densities of second level objects.

An interesting set of operations using polygons representing areal units are zonal functions, belonging to the wider group of operations discussed in Chapter 4 – map algebra. Zonal functions allow one to compute some characteristics of areal units, using a selected property or any set of properties that vary in space and are stored using a raster geographic data model. For instance, with a set of polygons representing areal units and a DEM for a certain area, one may compute the average elevation for each areal unit (Fig. 7.13). The values computed for areal units are then stored in a table, which may form a part of the relational database for these areal units. Administrative units are quite commonly used as areal units in zonal analysis, and as they are typically characterised by various other data collected, for instance, during the standard operations of

[67] In landscape ecology (see Box 7.3) such units are referred to as patches.

respective institutions, zonal functions add value in creating even more extensive databases and enrich the statistical analysis of various properties of administrative units.

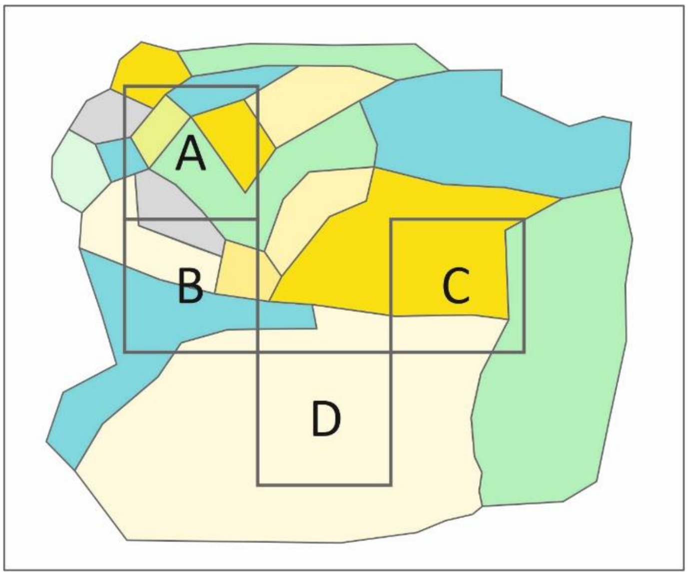


Fig. 7.11. Land use and land cover spatial structure.

The sizes of patches (polygons shown with different colours) vary in space – they are much smaller in the north-western part of the study area, and the largest in the southern and eastern parts. This pattern can be expressed with the number of patches per unit area. In four identical areal units (squares A, B, C, D) located in various parts of the study area the number of patches are 10, 6, 3 and 1, respectively.

In the case of us using regions instrumentally, as areal units and tools that serve to better characterise the Earth's surface, the delimitation of these regions – areal units – is simple, as any rule or criteria leading to region delimitation may be used. With the instrumental approach, the regions do not need to reflect any structures in the real world, and therefore any fragment of the Earth's surface may function as an areal unit. Such an approach brings about, however, a danger that is known in geography as the **modifiable areal unit problem** (Box 7.4, Fig. 7.14).

Box 7.4. Modifiable Areal Unit Problem.

Modifiable Areal Unit Problem (MAUP) denotes a dependency of the results of an analysis using areal units on their size and shape. Though known in geography for some time, it was thoroughly studied only in the 1970s and 1980s by Stan Openshaw. Openshaw proved that when applying any arbitrary divisions of space into polygons representing areal units, one may receive almost any result referring to the various aspects of the spatial distribution of selected properties. For instance, manipulating the shapes and sizes of areal units allows one to obtain various values for the correlation coefficients of two variables computed for these areal units – Openshaw showed that for the correlation any value from a range of -1 to 1 can be computed (Openshaw, 1983). Knowledge that the result describing a certain fragment of the Earth's surface depends on the shape and size of areal units has been used, among others, in politics, with an attempt to influence election results through the manipulation of the shape of voting districts, in elections carried out at various levels. Such practices are referred to as gerrymandering, the term originating from the name of Elbridge Gerry, an American politician, who used them in the 19th century elections in Massachusetts.[68]

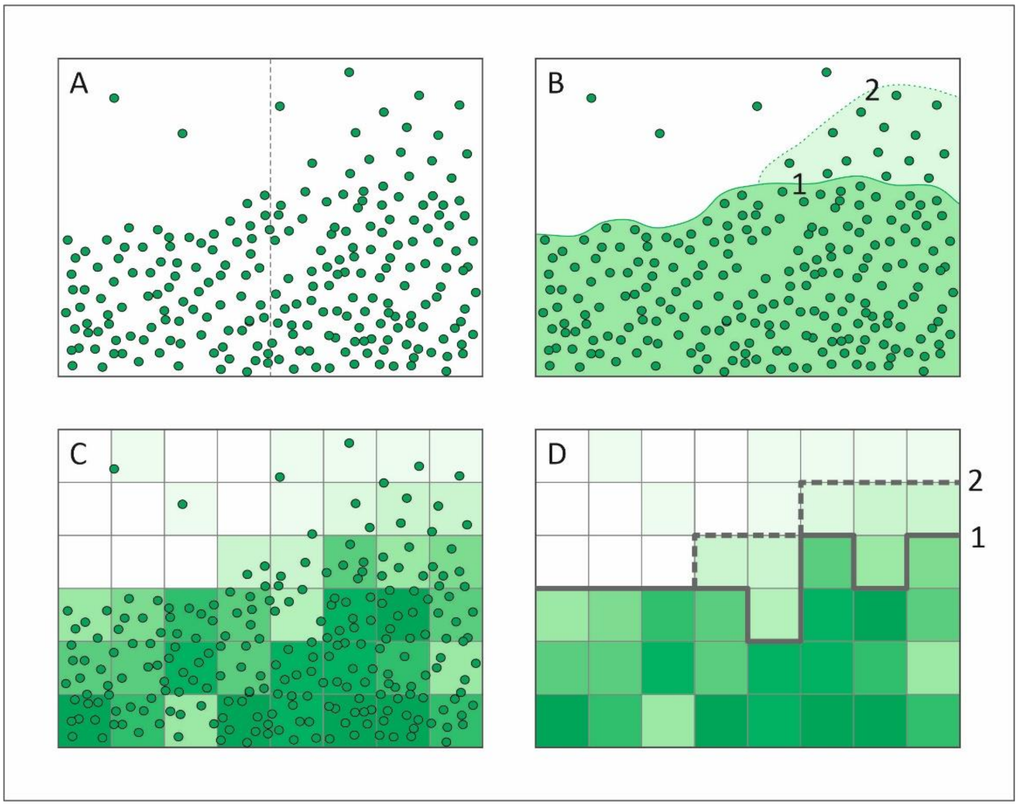


Fig. 7.12. Discrete and continuous variations in space and regional boundaries.

A – distribution of second level objects (for instance, trees on a mountain slope close to their climatic range, similarly as in Figure 1.3, presenting tree distribution on the south-facing slope of Cyl in the Babia Góra Mt. massif). Distribution of objects in the west differs from that in the east, and locating a boundary between the tree-covered and treeless area is quite easy in the west, while in the east the boundary may be delimited in various ways (for instance 1 and 2 in B or D). In D, the boundary was delineated based on regular areal units for which the tree density was calculated (C, higher density is shown in dark green), but, depending on the accepted density threshold, the boundaries in the eastern part are likely to be delineated in different ways. It is worth paying attention to the clear links between the method of boundary delimitation shown in B and the vector data model, and the method shown in C and D and the raster data model. The figure presents also a practical application of the point density analysis methods discussed at the beginning of Chapter 7.

Modifiable areal unit problem is a serious methodological challenge if regions are used instrumentally, as areal units to characterise some selected aspects of reality. To constrain its side effects, one should follow some rules on how areal units are delineated, trying to adjust their shapes and sizes to the specificity of the properties that will be analysed. Although regular, identical areal units seem to be a safe choice, in many cases they cannot be applied, either because they do not match the studied phenomenon or because it is impossible to find other relevant data for them. An example of the first case are hydrological studies, for which an ideal areal unit is a catchment, while an example of the latter case are socio-economic studies, for which a number of variables is known for administrative (or census) units, and the calculation of respective values for other areal units is difficult and may result in errors.

[68] Gerrymandering, after Wikipedia, accessible at https://en.wikipedia.org/wiki/Gerrymandering, accessed: July 2020.

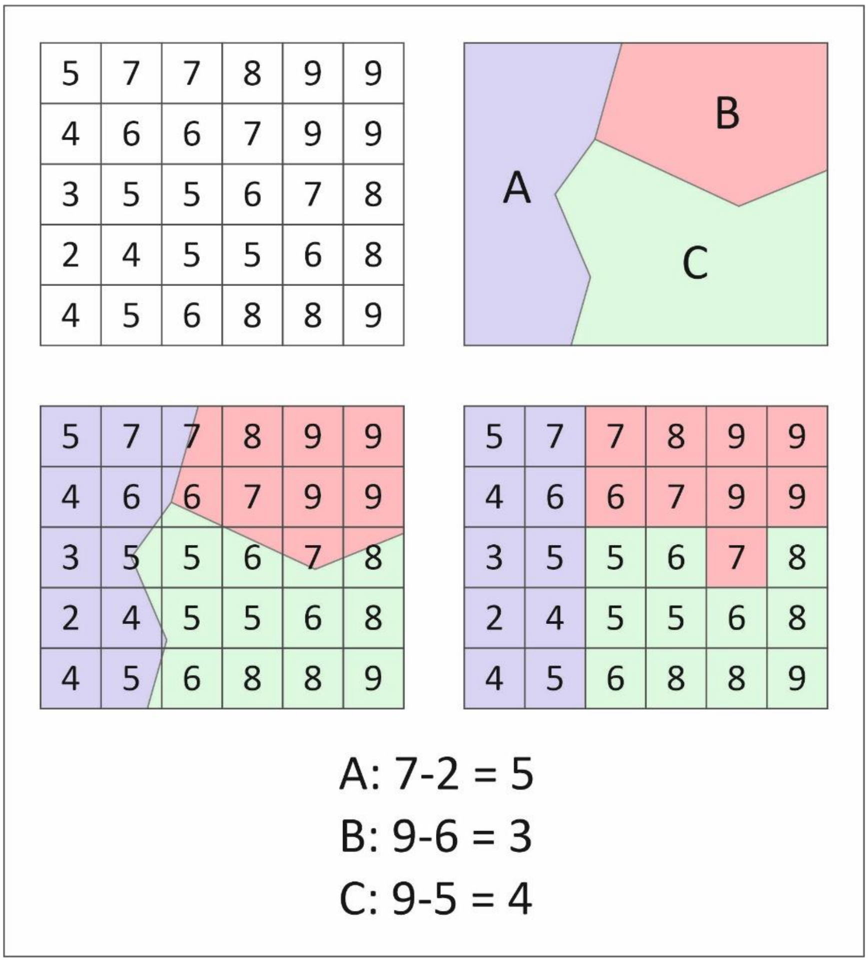


Fig. 7.13. An example of a zonal function.

Top: values of a property recorded with the raster model (left) and areal units for the same area in the vector model (right). Centre: areal units overlaid on a raster model of properties (left) and the transformation of areal units from a vector to a raster model (right). Bottom: the calculation of a value range for the areal units A, B and C.

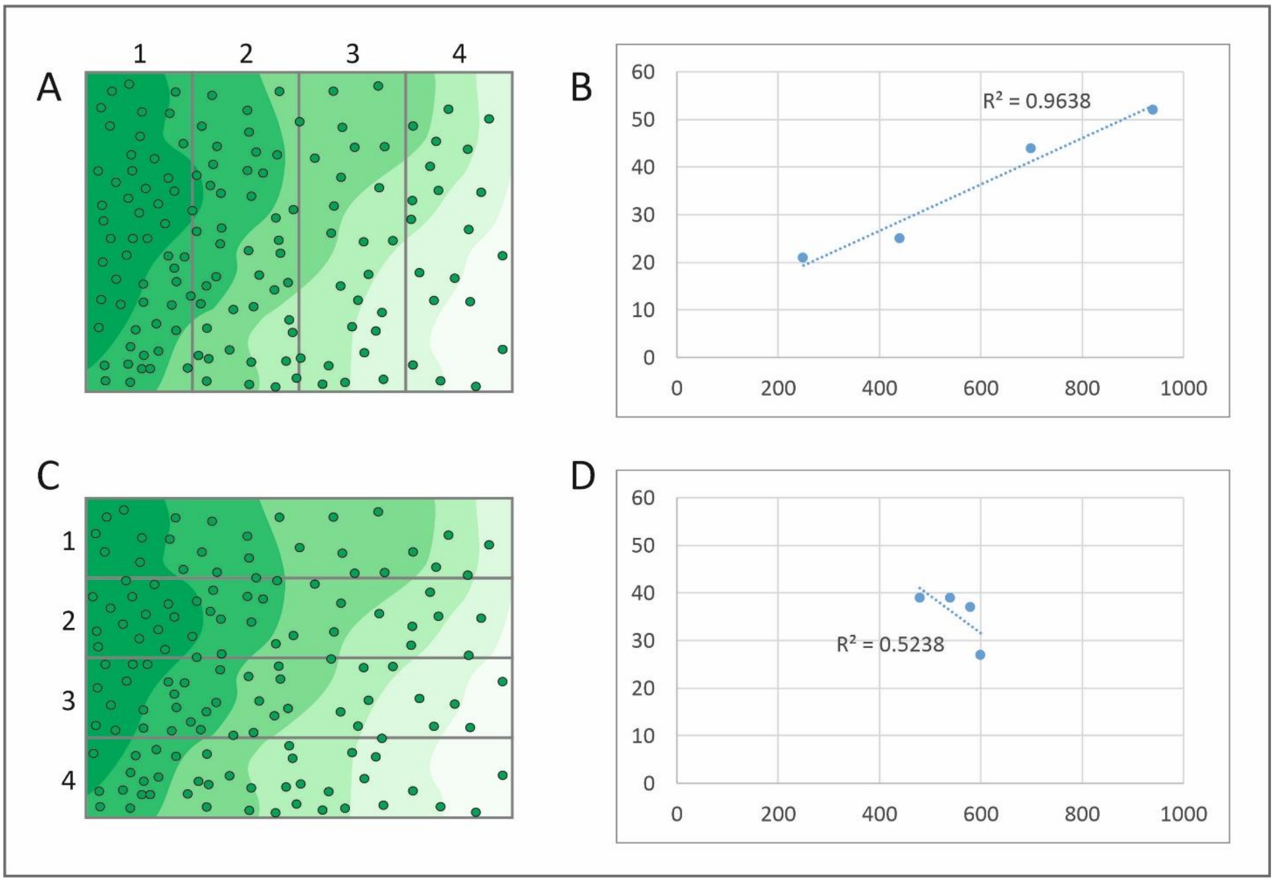


Fig. 7.14. Modifiable areal unit problem.

The figure presents a property that is continuous (shown with green shading and isolines, dark green being the highest value) and the distribution of discrete point objects. Top: for four areal units there is a clear relationship between the number of point objects in an areal unit (OY axis) and the mean value of the continuous property for the same areal unit (OX axis) – the higher the value, the more objects. Bottom: for four areal units delineated in a different way, the same relationship is weak, and moreover the number of objects is decreasing with the increasing mean value of the continuous property (OY, OX axes as above).

8. Geographic data quality

Key words: accuracy, error, coherence, completeness, uncertainty, generalisation, error propagation

There are two important features of any set of geographic data that fall into the wider concept of geographic data quality. The first one relates to how truly data represent selected aspects of reality. The other concerns the methods of data acquisition and how coherent they were when a specific data set was acquired. The measure of the first feature is **accuracy**. Measures of the other one are **coherence** (consistency) and **completeness**.

It is quite obvious that we wish to have data of the best quality. The most frequently met barrier to obtain such data is their cost. In theoretical considerations one may always follow the quality aspects mentioned above, in real-world applications, however, accuracy, coherence and the completeness of geographic data have to be related to the costs of their acquisition or production.

Scale and geographic data quality

Data quality strongly depends on how finely grained, or how detailed, is the representation of the real world by the data. Detail is a function of measurement density, with measurements understood quite widely, also as the recording locations of various objects. Obviously, for raster geographic data models, detail is closely related to spatial resolution: the higher the resolution, the higher the detail (at least, at first sight). If geographic data are stored on paper, a barrier to increase measurement density is the legibility of how the measurement results are recorded on paper. For digital geographic data such a limitation does not exist, and, in effect, **geographic databases can be considered scale-free**. It means that – in theory – we can record data with any finite density. Such an approach to geographic data storage would not be practical. Therefore, because of the processing time, the cost of production and the maintenance of geographic databases as well as the limitations of geographic data visualisation it is assumed that raster and vector models record data resulting from measurements of a defined density and falling into a certain spatial scale range. Therefore detail mostly corresponds to the spatial scale which is planned for a specific geographic database. These scale – detail relationships can be easily showed for raster data because of the close link of detail and spatial resolution in the raster data model. For instance, at global scales, in the range of 1:10 000 000 or smaller, data with spatial resolution of approximately 1 km are appropriate. On the other hand, Landsat satellite image data with a spatial resolution of 30 m are suitable for a spatial scale range of approximately 1:100 000. It is worth adding, however, that the relation between the spatial scale and detail (spatial resolution) of geographic data is not constant and changes with the increasing technical capabilities of geographic data processing. In recent years, global-scale analysis is being carried out using data with spatial resolutions and detail adequate for much larger scales (or smaller areas), resulting in processing geographic data sets with very large volumes – so-called big data. A good example of such a global analysis which uses remotely sensed big data from the Landsat programme is Global Forest Change, described in Chapter 6 (Box 6.3).

Geographic data with high detail can be converted into low-detail data – such a transformation is referred to as the **generalisation** of geographic data. Generalisation originates from traditional cartography and its simple limitation that arises when a large-scale map has to be contracted and reproduced in a smaller scale. With simple mechanical contraction, too many details would make reading the map difficult, so at least some phenomena shown on the large-scale map have to be removed or simplified, optimally using a logically clear set of rules. Such a generalisation is a one-way process, leading from maps in larger scales to maps in smaller scales – and there is no possibility to carry out an inverse operation, that is increasing geographic data detail, without new measurements and acquisition of additional data. Here it is worth adding that important issues related to generalisation have been touched on already in Chapter 1, where the relationships between spatial scale and object dimensionality are discussed (although the concept of generalisation has not been explicitly mentioned).

For obvious reasons one may accept that large-scale geographic data, with high measurement density, better reflect and represent selected aspects of reality than data created for small scales, with a low density of measurements (Fig. 8.1). This does not mean, however, that various components of data quality depend fully and exclusively on the spatial scale ranges that these data match. Measurement density does not anticipate the accuracy of individual measurements, moreover problems related to the coherence and completeness of geographic data may refer to various data sets regardless of the spatial scale and detail of the data.

Accuracy

Accuracy refers to how well data reflect and represent selected properties of the real world. This short definition requires, however, some clarifications. First, we do not have any ideally true representation of the real world and any of its properties. Our perception and cognition of reality, even the most tangible and reliable, is burdened with **uncertainty** that is embedded in a complexity and diversity of the real world and the imperfections of our cognition. These issues have been studied for hundreds of years, primarily by philosophy, and their extensive overview can be found in a number of textbooks.[69] Next, geographic data represent the real world in a specific scale range, referring both to time and space – and that means that the accuracy of geographic data should not be considered without reference to the scale range, for which the data were produced.

The accuracy of geographic data, for any phenomena, refers to three different aspects of these phenomena: their location in space, their location in time and the values of their attributes. Therefore we typically refer to spatial accuracy, temporal accuracy and thematic accuracy, respectively. To illustrate these three various accuracies, let us return to crime geography and the example discussed in Chapter 6, dealing with geographic data representing specific types of crime. As for burglary, we are able to provide a detailed location in space, with a high accuracy (known address, the unambiguous location of a building where a burglary occurred), but time is typically not well known (most frequently, this is rather a period, sometimes a long one, for instance in the case of the longer absence of hosts). As for the attribute – losses related to the incident – these are sometimes also difficult to assess, and may require employing experts. For pickpocketing, both the location in time and space may be difficult to assess, because – as for the time – the victim commonly realises the fact only after some time, and being mobile means then a lack of certainty also as to the location in space. However, with this crime it is much easier to accurately define what was lost and what its value was. To conclude, burglary is an event represented by data with a high spatial accuracy, low temporal accuracy and commonly a low thematic accuracy, while pickpocketing represents an event characterised by low spatial and temporal accuracies and mostly high thematic accuracy.

[69] For instance, Woleński J., 2007, Epistemologia. Wydawnictwo Naukowe PWN.

In practice, instead of measuring accuracy, it is much easier to measure its inverse, that is **error**. Error is the difference between the result of a measurement of location in space, location in time or an attribute value recorded in the data, and the true value of the locations or attribute. Taking into account that the true value is unknown (see the concept of uncertainty discussed above), an error *E* is then the difference between the measured value *m* recorded in the data and a certain **reference value** *r*:

$$E = m - r \qquad [8.1]$$

The reference value is a measurement about which one may judge that it is significantly more accurate than the data being tested.

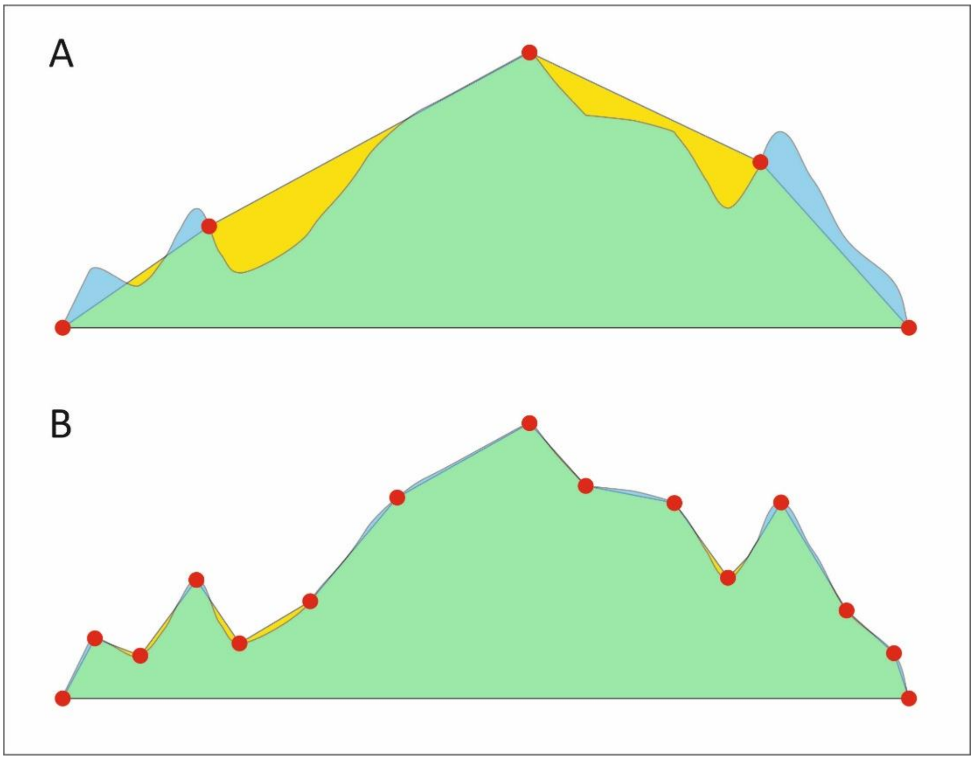


Fig. 8.1. Terrain elevation profile, measurement density and data detail.

A – 5 measurement points, low density and detail; B – 15 points, high density and detail. Blue shows the parts of the profile where elevation is underestimated, yellow – overestimated. For low detail measurements, (A) over- and underestimation are clearly visible while for high detail measurements (B) over- and underestimations are insignificant.

Let us assume that at a certain point elevation was measured to be equal to 401 m a.s.l., using remotely sensed methods (these methods were briefly discussed in Chapter 4, Box 4.1). At the same point, the elevation was measured to be 396 m a.s.l., using methods relying on highly accurate satellite navigation. If the latter measurement is considered to be more accurate than the former (for instance, based on detailed knowledge about how these two measurement technologies work), then the error of the remotely sensed measurement is equal to 5 m.

To assess the accuracy of any dataset we commonly do not limit our assessment to a single error estimate – rather, the assessment is based on a sample of measurements out of its entire population that builds a specific dataset (Fig. 8.2, Table 8.1). In such a case, accuracy can be determined with various measures that describe some aspects of error distribution. The most frequently applied ones are mean error (ME), mean absolute error (MAE), and root mean square error (RMSE). For *n* independent measurements for which reference values are known, ME, MAE and RMSE are calculated in the following way:

$$ME = \frac{\sum(m_i - r_i)}{n} \quad [8.2]$$

$$MAE = \frac{\sum|m_i - r_i|}{n} \quad [8.3]$$

$$RMSE = \sqrt{\frac{\sum(m_i - r_i)^2}{n}} \quad [8.4]$$

where m_i and r_i denote again measured and respective reference values.

ME may have either positive or negative values, or even 0, if negative and positive errors have the same absolute values. MAE and RMSE are always non-negative, and their value are affected mostly by how much measured values differ from reference values, regardless of if the differences are positive or negative.

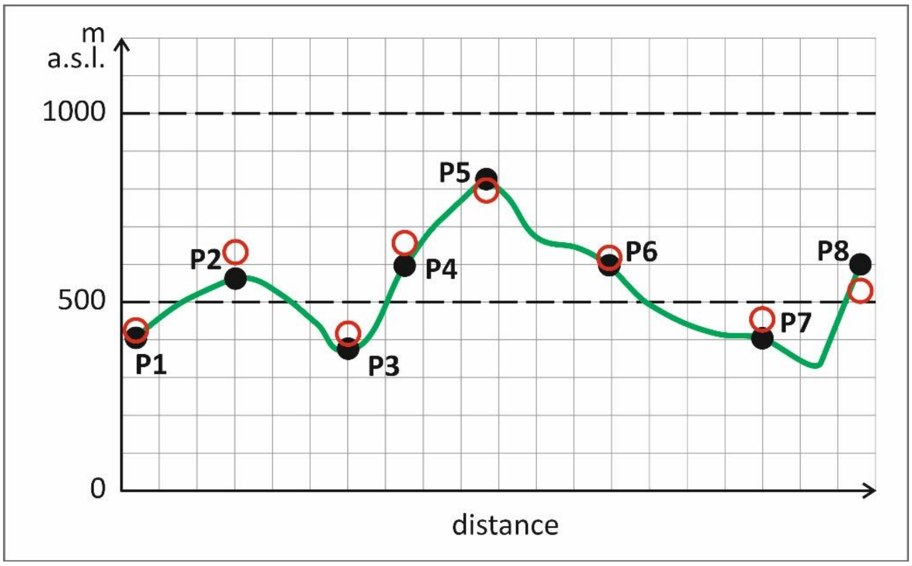


Fig. 8.2. Elevation profile, measurements and their errors.

Elevation measurements were carried out in 8 points along the profile (P1-P8; *n* = 8), using a remote sensing technology. For these points reference values are known. For a specific point, the difference between the measured value (red circles) and the reference value (black dots) is the error (see also Tables 8.1 and 8.2). Note that the error shown on the graph is exaggerated tenfold with respect to the scale of the OY axis.

Table 8.1. Measurement and reference values for elevation data.

Point ID	Measured value	Reference value	Error	Error absolute value	Error squared
P1	409	407	2	2	4
P2	570	563	7	7	49
P3	386	382	4	4	16
P4	605	599	6	6	36
P5	811	814	-3	3	9
P6	610	608	2	2	4
P7	407	402	5	5	25
P8	594	601	-7	7	49
		Sum	**16**	**36**	**192**
		ME	**2.00**	x	x
		MAE	x	**4.50**	x
		RMSE	x	x	**4.90**

Errors are resultants of two factors. The first one is purely random – we are not able to provide any explanation why such errors occur. The other one is related to the measurement method that may result in specific errors, for instance a systematic over- or underestimation of measured values. Errors of the first type are referred to as random errors, while the latter are systematic errors. When dealing only with random errors during measurements it is expected that ME will be close to 0. It does not mean that there are no errors in individual measurements and measured values are in perfect agreement with reference values (though such a case is theoretically possible), but rather that errors in different locations average out to 0 – sometimes measured values are higher than reference values, sometimes the other way round. With systematic errors ME will differ from 0, indicating that the measurement is systematically over- or underestimated. Importantly, ME can be computed also as a difference between the average of measured values and the average of reference values – in the example presented in Fig. 8.2 and Table 8.1 the first value equals 549 m, and the latter 547 m, with ME equal to 2 m. Such an ME indicates that on average, measured values are overestimated by 2 m as compared to reference values. The easiest way then to remove the systematic error is simply to subtract its value from all the measured values. After such an operation the ME will be equal to 0, and MAE and RMSE will be minimised, depending only on random errors (Table 8.2).

An example of the terrain elevation measurement method that leads to systematic errors in some contexts is radar interferometry (see also Chapter 4, Box 4.1). In forested areas this method overestimates elevations due to the reflection of microwaves from tree crown tops instead of from the ground, which is the reason that recorded elevations are systematically greater than reference values, with a difference approximately equal to the height of the trees (Fig. 8.3).

Table 8.2. Correction of systematic errors.

Error measures calculated for data shown in Fig. 8.2, after subtracting ME from measured values (the 'corrected value' column) to remove systematic error.

Point ID	Measured values	Corrected value	Reference value	Error	Error absolute value	Error squared
P1	409	407	407	0	0	0
P2	570	568	563	5	5	25
P3	386	384	382	2	2	4
P4	605	603	599	4	4	16
P5	811	809	814	-5	5	25
P6	610	608	608	0	0	0
P7	407	405	402	3	3	9
P8	594	592	601	-9	9	81
			Sum	**0**	**28**	**160**
			ME	**0.00**	x	x
			MAE	x	**3.50**	x
			RMSE	x	x	**4.47**

In many cases it is quite easy to determine which data may serve as reference data for other measurements. The DEM mentioned above, created from data acquired in the SRTM project (see Chapter 4, Box 4.1) may be easily compared to DEMs created from aerial measurements, using either photogrammetry or airborne laser scanning that provide very accurate elevation measurements, sometimes with centimetre accuracy, and with very high detail. Such data are certainly sufficient to estimate errors of SRTM DEM, as its global accuracy amounts to several metres. Sometimes, however, having various datasets, it is not easy to say which data are more accurate and which data might serve as reference data to assess the accuracy of other datasets. A good example of such a difficulty is the accuracy assessment of cloud cover – a standard product received from the processing of satellite data acquired by various sensors (Box 8.1).

Box 8.1. Cloud cover data assessment.

In his PhD thesis, defended in 2011, Andrzej Kotarba compared data about cloud cover recorded at meteorological stations in Poland with the cloud mask product resulting from the processing of data acquired by Moderate Image Spectrometer (MODIS) on board the satellites Terra and Aqua. Cloud mask data are one of standard MODIS products, created using a special algorithm operating on selected spectral bands registered by MODIS.[70] In the PhD thesis, raster layers containing four categories of cloud cover (cloudy, uncertain clear, probably clear, confident clear) were used to estimate total cloud cover for a selected area, for instance around a meteorological station. On the other hand, cloud cover data recorded at meteorological stations are human observations encoded in a nine-stage scale, with 0 denoting a lack of clouds and 8 denoting full cloud cover. In remote sensing, ground data are frequently considered to be accurate and serve as reference data – hence they are sometimes referred to as *ground truth*. However, A. Kotarba proved in his work that in the case of cloud cover relations between ground observations and remotely sensed data are much more complex, depending also on the type of clouds. If the area around a meteorological station is dominated by clouds with a well-developed vertical structure, an observer located on the ground typically

[70] MODIS Data Products, accessible at: https://modis.gsfc.nasa.gov/data/dataprod/, accessed August 2020.

overestimates the cloud cover, because such clouds may obscure the sky. Such a problem does not exist for a hypothetical observer on board a satellite (or simply for a sensor), who takes into account only the ground projection of the largest horizontal cross-section of a cloud, regardless of its vertical structure, defining thus which areas are under the cloud, and which areas are not. On the other hand, ground observers – skilled staff of meteorological stations – are much better in capturing cloud cover related to the occurrence of delicate *Cirrus* clouds that may not be fully detected by satellite sensors. Concluding, in some situations ground data could be assessed in the context of the more accurate satellite observations serving as a reference, while in some situations one may test the same data the other way round.

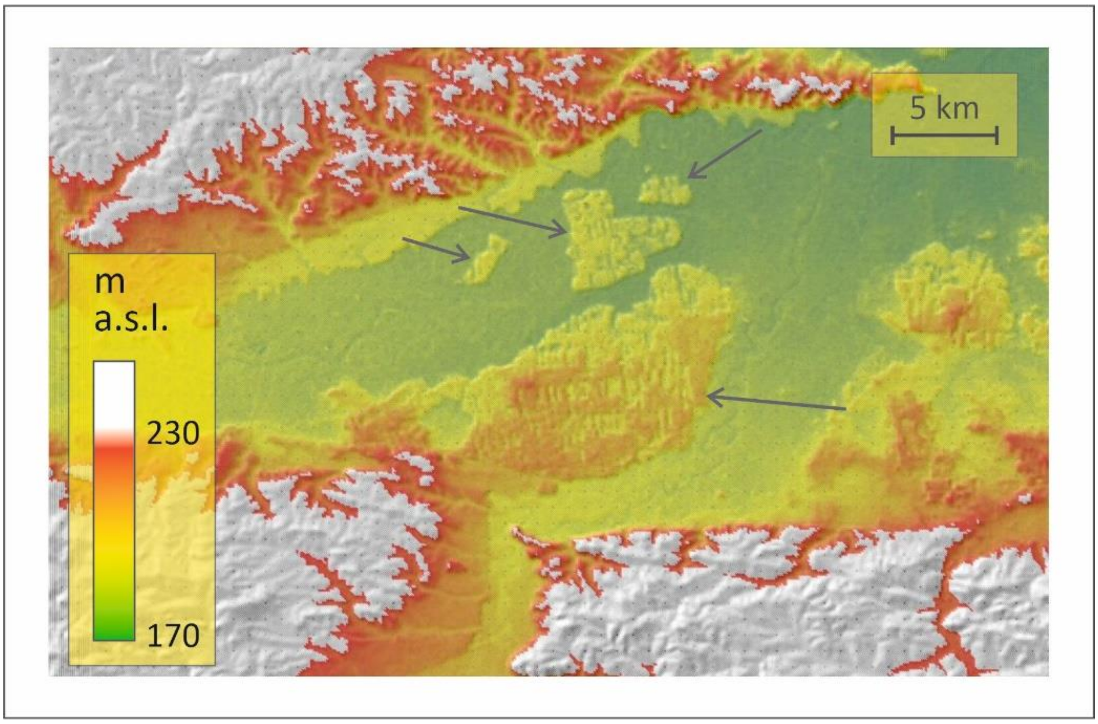


Fig. 8.3. Systematic errors of SRTM DEM.

Elevation data acquired using radar interferometry in the SRTM project, for the Niepołomice Forest region in Poland. Forested areas (marked with arrows) are clearly visible as areas with a slightly higher elevation than the surrounding grasslands.

Coherence and completeness

Coherence and completeness are two aspects of data quality that are to some extent complementary, and therefore it is easy to discuss them jointly. Completeness defines whether data refer to the entire set of phenomena of a certain type or only to its part. In the case of geographic data, completeness is typically related to areas and defines whether data have been collected for the entire study area, or whether some of its parts have been omitted. For instance, one of the standard criteria to select for analysis satellite optical image data is cloud cover: the occurrence of clouds means that image data are not complete (of course in the case we are interested in that of studying the Earth's surface – the most typical situation, and not in studying clouds per se). Completeness may, however, refer also to the temporal aspects of the data: for instance, for a series of meteorological data for a specific station one may find that the station was closed for some time and the series is not complete with reference to the specific part of the period that needs to be analysed. Finally, in data that are complete as for space and time some specific attribute values might be missing, for instance due to the fact that some data were not acquired in a given location or for some time – in such a case we may deal with a lack of completeness related to thematic aspects.

In the history of cartography and geographic discoveries incomplete data led to incomplete maps with empty spaces. These empty spaces reflected the incomplete knowledge of cartographers, and filling them either with lands that existed only in the imagination of cartographers or images of fantastic beasts was a way to hide the scale of ignorance about the area presented on a map. Currently the situation is completely different: if a certain dataset is not complete, we are able in most cases to find another one or to use mathematical methods that help to fill the gaps. Such solutions reduce problems related to the completeness of geographic data, but create new problems, related to the lack of coherence of the data. So today we are not dealing with incomplete data, but rather with data that are not fully coherent, for instance data with gaps related to some data acquisition method that later on were completed in some way.

The coherence of geographic data, for any selected area, refers then to measurements that were used to collect data and to the stability of the spatial, temporal and thematic accuracy of these measurements. Earlier in this section the example of incomplete satellite optical image data with clouds was brought up. Let us assume that we intend to create a satellite land cover map for a specific area, based on one image. In places under clouds, the map will not be complete. A simple solution then is to use another image, for instance acquired a few days later, for which the areas under cloud are well visible, and to run again a land cover classification algorithm just for the areas that are missing in the first image (Fig. 8.4). The output land cover map, even if accurate, will not be coherent in the temporal context, because in fact it is made up of two different land cover maps, representing two distinct moments of time. The longer the period between the acquisition of the first and the second image, the higher the significance of the incoherence for the analysis of the land cover and its changes.

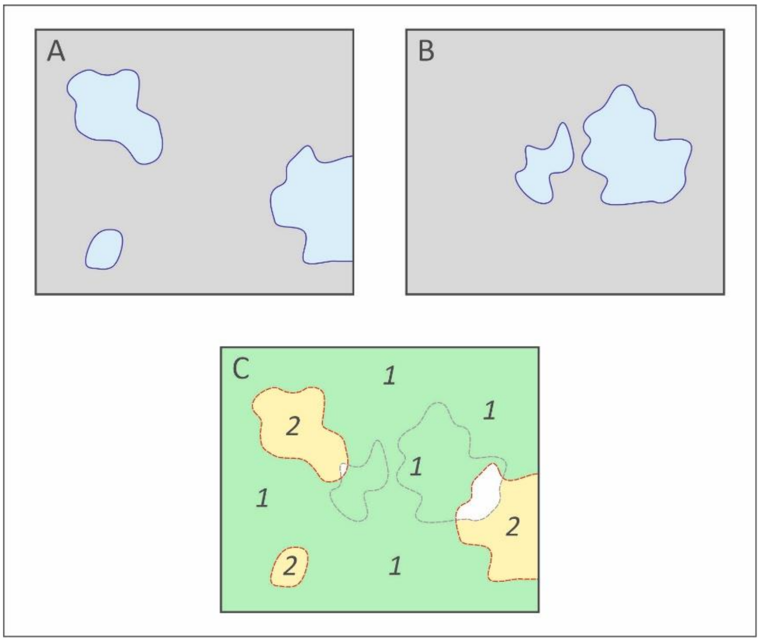


Fig. 8.4. Lack of coherence and completeness of a land cover map created from two partially cloudy satellite images.

A, B – images used for land cover classification, grey – clear areas, light blue – clouds; C – output map made based on image 1 (light green) and image 2 (light yellow), white shows areas under clouds on both images and therefore not mapped (however, such a gap can be completed using another image). The red outlines show clouds from image 1, grey – clouds from image 2.

Such a method of filling gaps in the satellite data, and improving spatial completeness at the expense of the temporal coherence of the output product was used in compiling the Global Forest Change data, discussed in Chapter 6 (see Box 6.3). Global Forest Change contains a complete set of global data for each year of the period starting in 2000, yet each annual layer representing forest cover and its changes is received through merging satellite images acquired across the entire vegetation season for a given year, in order to eliminate areas under cloud cover using an approach similar to that presented in Figure 8.4 (the number of images was typically much higher). For forest losses, assessed in Global Forest Change on an annual basis, a potential influence of this temporal inconsistency on output data quality (that is the accuracy of determining the moment of deforestation) will be rather minor.

A very good, though historical example of complete, but incoherent data is the global DEM GTOPO30, complied and made accessible in mid-1990s. GTOPO30, with a spatial resolution of 30 arc seconds (approximately 1 km in the equatorial zone), covered all lands thanks to the merging of many different DEMs of various detail and accuracy.[71] The most important sources for GTOPO30 were elevation data prepared based on the interpolation of contour lines from vectorised topographic maps in a scale of 1:100 000 (Digital Terrain Elevation Data, DTED) and elevation data that also used contour line interpolation, yet the contour lines were derived from the Digital Chart of the World, in a scale of 1:1 000 000. The latter source had a much lower accuracy than did the former, due to the much higher generalisation of contour lines from maps in the scale of 1:1 000 000 as compared to maps in the scale of 1:100 000, with obvious consequences for the detail of the final output (Fig. 8.5).

Contemporary DEMs, even those obtained with efficient methods of data acquisition based on remote sensing technologies, may not be complete. For instance, due to the specific properties of terrain and the geometry of microwaves used in radar interferometry, the data for the SRTM project were not acquired with full coverage for several desert areas, while data gaps were also quite common in high mountain areas. Data gaps were therefore filled in on the subsequent versions of the SRTM DEM using other available datasets or via spatial interpolation[72] – such an approach, however, may contribute to some coherency-related issues when analysing these elevation data.

A need to fill gaps in source data is one of the most obvious reasons for the lack of coherence in geographic data, with some impact on data quality. Incoherence in the geographic data may be, however, much more subtle. For instance, if one geographic database is created by various teams, problems related to data coherence may appear even if the data processing procedures leading to the final output are very well detailed. An interesting example of such a subtle incoherence are land use and land cover data for Europe, Corine Land Cover (CLC), discussed in Chapter 3 (see Box 3.1). CLC data are compiled by national teams of experts for each country involved, according to a single well-established methodology, and later on merged into a pan-European product. Such a methodology may result in slight differences in the interpretation of various land cover types among the national teams that result ultimately in incoherences within the database. The author of the book could document these incoherences while analysing raw CLC2000 data for some Central European countries. These data were created with an approximate 1 km wide buffer beyond the national boundary, and in consequence a 2 km wide strip along a national boundary was interpreted twice and independently by the national teams of two neighbouring countries. Considering CLC nomenclature of the 3rd level (44 land use and land cover categories), interpretation differences in the boundary zones reached around 30%. Such a difference proves that incoherence may occur over the entire CLC dataset. Of course, being aware of the incoherence one may attempt to minimise its impact on analysis

[71] USGS EROS Archive – Digital Elevation – Global 30 Arc-Second Elevation (GTOPO30), accessible at https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science_center_objects=0#qt-science_center_objects, accessed: August 2020.
[72] SRTM 90 m DEM Digital Elevation Database, accessible at http://srtm.csi.cgiar.org/, accessed: August 2020.

outputs; further, perfecting CLC data processing methods may help to remove the most important causes of interpretation differences.

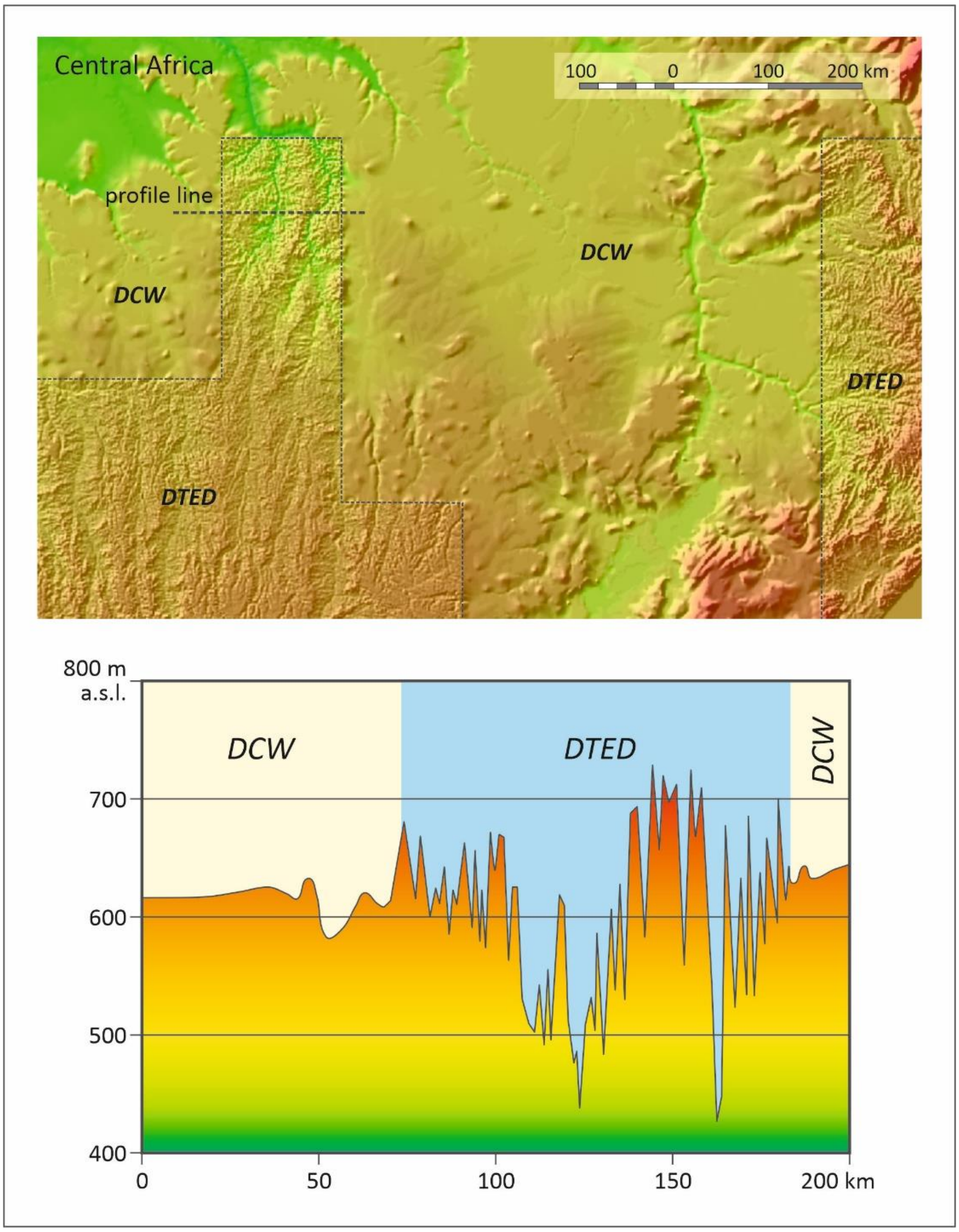


Fig. 8.5. Coherence issues of GTOPO30 DEM.

Above: a fragment of GTOPO30 DEM for Central Africa, with clearly visible tiling and differences between the data derived from Digital Terrain Elevation Data (DTED) and the Digital Chart of the World (DCW); below: the elevation profile showing how the level of detail differs for DTED and DCW data.

Source: Kozak (2004).

Another example of the incoherence of geographic data are global population data in the Gridded Population of the World (GPW) database, briefly introduced in Chapter 2 (see Box 2.1). Let us recall that these data, with a spatial resolution of 30 arc seconds (approximately 1 km in the equatorial zone), are compiled from census data collected in the administrative (or census) units of all world countries and then resampled into a regular grid of cells. Incoherence in the data result from the various sizes of areal units used in different countries (e.g., due to the availability of population data for various levels of administrative divisions). For instance, in Portugal GPW data come from more than 265 thousand very small administrative units, with an average area of 0.35 km^2. In neighbouring Spain, GPW uses population data from around 8 thousand areal units, with an average area of 63 km^2. Regardless of the size of the areal units, the population allocated to these units is always converted to raster cells of the same size (30 arc seconds). If areal units are of a similar size to raster cells, then the spatial resolution optimally reflects the measurement density. But in various countries areal units are much larger than raster cells which means that many cells are redundant, and most store the same information as the neighbouring cells, from the same areal unit. On the contrary, in some cases areal units are smaller than raster cells, with the consequence of information loss when transforming the population data from areal units into the raster model. Anyway, population data collected for areal units with such area differences may result in a relatively incoherent pattern of population distribution (Fig. 8.6).

In this way, the spatial detail of GPW varies significantly in space, due to differences in national census systems, population reporting and size of the areal units for which population data are made accessible. Accordingly, GPW data do not allow one to carry out a global analysis of population distribution with no precautions, as data incoherence may result in detecting fake patterns of population distribution. In the example illustrated in Figure 8.6 someone might conclude that vast areas in eastern Portugal are only poorly populated, because population density is less than 1 person / km^2, contrary to neighbouring regions in Spain that are more densely and evenly populated. Such an effect is, however, solely the outcome of differences in the sizes of the areal units used in population mapping on both sides of the state boundary.

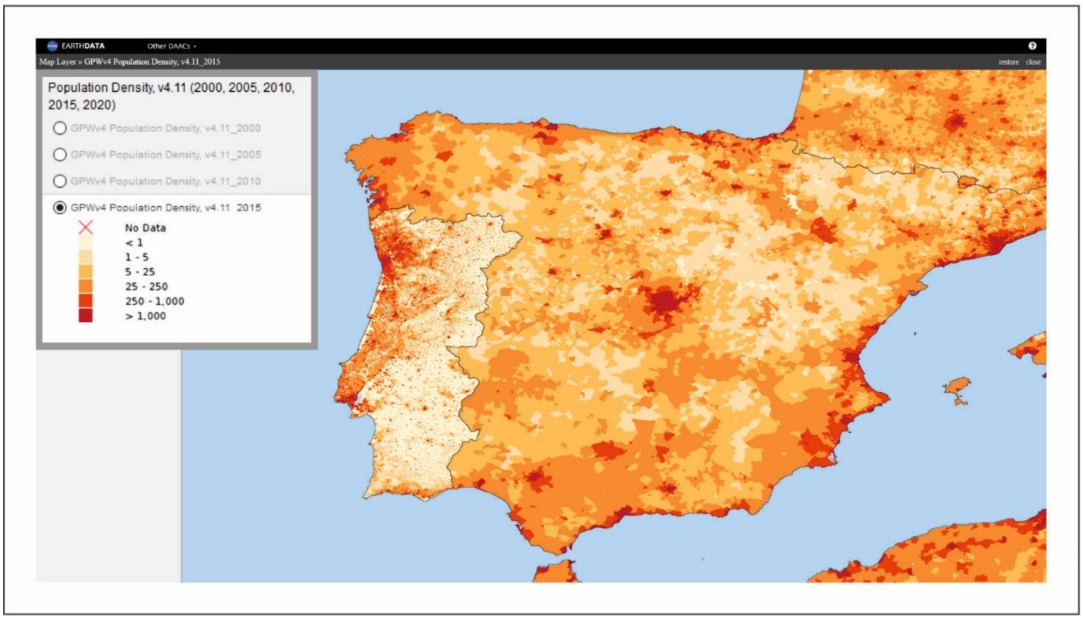


Fig. 8.6. *Gridded Population of the World*, version 4.

Population density for Portugal and Spain, with a clear incoherence in population distribution on both sides of the state boundary.

Source: https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11, accessed: August 2020.

Error propagation

Geographic data are used in analyses that rely on the processing of both accurate and inaccurate data – that is, data containing errors. When processing data with errors, errors are typically multiplied, and become higher in the outputs than in the input data. Such an effect is referred to as **error propagation**. A good illustration of error propagation is an analysis of an inaccurate DEM (Fig. 8.7). Small errors in assessing elevations may lead to significant errors in determining slope gradients and water flow directions, resulting in receiving river networks generated from DEM with a channel pattern completely different than the real one.

Error propagation is quite common also in remote sensing change detection. In the simple case of land cover change analysis for two land cover maps, one received at the beginning and the other at the end of the studied period through automated image data classification, the accuracy of land cover change (the output of map overlay of the two land cover maps) is a product of accuracies in the input layers, assuming errors in both maps are randomly located and mutually independent. For instance, with input accuracies equal to 0.9 and 0.8, the land cover change map has an accuracy of 0.72 (or 72%), with an error of 0.28 (28%).

Error propagation has to be carefully considered before analysing geographic data. It is necessary to take into account that merging various data, with various errors, may result in some unexpected outcomes that will emphasize the errors and incoherences, even if they were – at the beginning of data processing – not clearly visible in the input datasets.

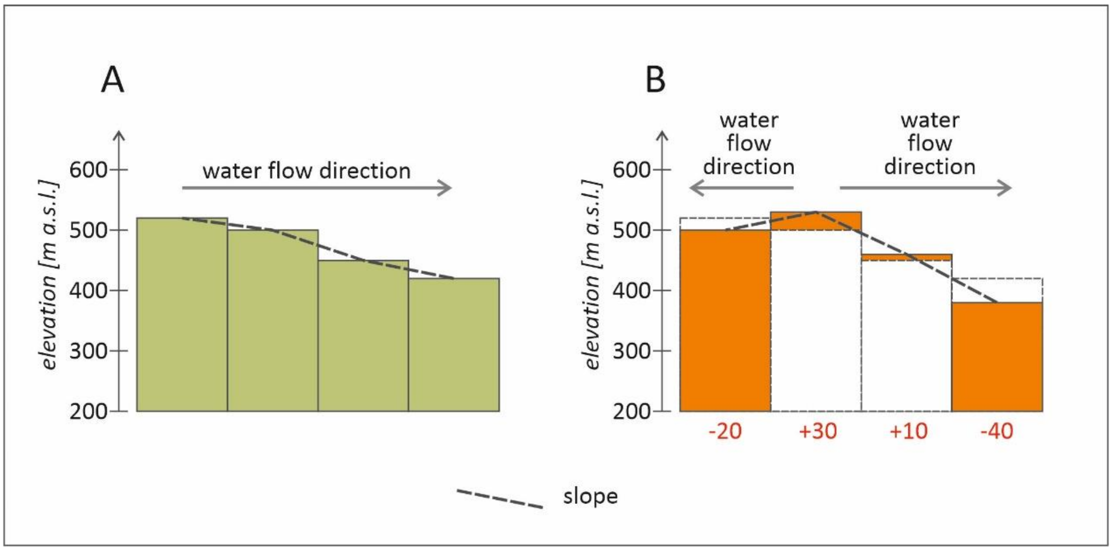


Fig. 8.7. Error propagation in a raster DEM.

A – accurate elevation values allow to compute slope gradients and water flow directions; B – relatively minor elevation errors in various cells (errors given in metres, at the OX axis, in red) lead to significantly different slope gradient values and a completely erroneous (left side of the profile) determination of water flow direction.

EPILOGUE

With this book, I have intended to present, as simply as possible, how various phenomena occurring in the real world around us are represented with geographic data models, and to discuss various problems related to these representations encountered by geographers or specialists working in other fields who frequently use geographic data. First, I have shown that geographic data are a relatively simple way of recording the properties of phenomena, resulting from how we perceive the reality and what methods of measurements we use. Next, I have attempted to present the variety of phenomena these data represent and the wealth of analyses and interpretations they offer. I do hope this purpose was at least in some degree successfully realised.

At the end of this book I wish to refer to one particular issue that – quite unexpectedly also for the author – has not found its place in any chapter, and its absence may raise serious doubts among readers. This issue, obviously, is **a map**.

Initially I wanted to dedicate an entire chapter to maps, tackling mainly how a paper map[73] differs from geographic data recorded in a digital form and stored in a database. Finally, with resolve, I have put aside this idea, limiting the scope of issues related to maps – paper maps – to this brief commentary placed outside the main body of the book. I think it is necessary to provide reasons for this decision that for many experts dealing with geographic information, to whom cartography remains important, may be at least surprising, if not shocking. This is because maps are so closely linked to geographic information that the development of many fields currently employing geographic information seems unimaginable without the centuries of previous cartographic achievements. Maguire et al. (1992) proposed that in GIS, in a broad sense, one may find three overlapping approaches, none more important than each other: the database approach, spatial-analytical approach and cartographic (or map-based) approach. So if the cartographic approach represents one of three equally essential roots of GIS, with a map being its key metaphor, it is worth explaining why this approach, and maps, have been left out of this book.

The paper map, obtaining in more or less its final shape somewhere in the middle of the 20th century (though modified also later on), was the first geographic data model invented and perfected by humans. In this model, geographic phenomena and their properties are represented with graphical symbols. Their proper choice required a thorough consideration of the nature of geographic phenomena – so this is not strange that so many issues covered in Chapters 1-8 (for instance, continuous and discrete approaches to geographic phenomena, the dimensionality of objects, spatial interpolation) were studied in detail by cartographers well before a digital representation of geographic data entered the scene, as they lay also at the foundations of paper maps and their making.[74] Thanks to the development of cartography, the paper map has become a respectable achievement in the field of representing geographic phenomena, successfully performing its duties, and, in part, doing this even now. However, as compared to digital representation with raster or vector geographic data models, a paper map displays a number of limitations. These limitations are serious

[73] One may use also in this context the term *traditional map* as it is not the material, but the graphic way of recording and organising information that really matters (traditional maps can be painted on canvas, or carved in stone, for instance).

[74] A digital representation inherits a lot from the cartographic legacy. For instance, coding 0-, 1- and 2-dimensional objects was almost directly transferred from paper maps to the vector geographic data model. Some methods, however, have found their use only with the onset of GIS&T: here, the concept of raster geographic data models may be indicated, which was practically useless in traditional cartography but proved its value when computers were applied in spatial analysis. On the other hand, some methods elaborated in cartography and of value in a paper map world, are not really useful to store geographic data today – isolines, for instance, are among them (this issue was briefly discussed in Chapter 4).

enough to provoke questions about how sensible – while discussing geographic data and information – it is still to employ the old ways and concepts that are appropriate for paper maps.

Paper maps fulfil two functions at the same time: they store geographic data and they present (communicate) geographic information to a potential user. As with many multifunctional devices, paper maps subsequently require a compromise between different functions, in this case between data storage implying a paper map is some kind of a database, and communication with the user. A simple example of such a compromise is representing a 0-dimensional object on a map in the scale 1 : 100 000: if the object is presented using a dot with a diameter of approximately 0.5 mm (clearly visible on paper), its real size, taking into account the map scale, is 50 m, which does not allow for an accurate location of the object. Meanwhile, in the vector data model the location of the same 0-dimensional object is as accurate as its coordinates in the geographic database, regardless of its graphic presentation (e.g., on a screen). This is because the main virtue of digital geographic data models is that **data storage is fully independent from how data are visualised**[75] – a point object is simply a pair of coordinates in the database which if necessary may be presented on screen using a symbol of any size or shape that does not influence the accuracy of the point location – it is determined fully by how the coordinates are stored. This virtue is not limited only to point objects, but to any phenomena that are represented with raster or vector geographic data models.

Separation of storage and the visualisation of geographic data in a digital form allows one to simplify the storage itself and to increase its precision and accuracy. It provides also – as compared to paper maps – a much higher capability for data analysis, as from the very beginning one may act on structured sets of numbers to what computers are very well suited, omitting graphic symbols that need to be first interpreted. The value of digital storage does not harm data visualisation – quite the contrary, it gains a lot because when visualising data stored in a geographic database one may focus entirely on the quality of the visual communication (and sometimes on the speed with which the data appear on the screen[76]). This can be done even at the expense of how truly phenomena are represented, because **a true representation of these phenomena is safely and unambiguously stored in the geographic database**.

For digital data, visualisation is a means to extract some specific information from the geographic database, most frequently through displaying something on screen (much less common way is to print something on paper). Contemporary GIS software offers a range of fantastic capabilities in the domain of geographic data visualisation, that may lead to simple or complex, yet impressive outputs directed on screen or to the printer. I disagree therefore with the opinions of many experts in the field of cartography that a multitude of geographic data visualisation options in GIS software harms cartography, because thanks to them everyone can make maps, frequently with few proper cartographic skills or knowledge. I think that GIS software rather allows many people, with different levels of expertise, to experiment and learn how to make maps – which on the one hand may lead to obvious errors, but on the other – to interesting and valuable ideas. This virtue of *democratising* cartography is, in my opinion, much more important than any resulting problem, especially when the visualisation of scientific data is such a quickly developing field, regardless of whether the data are geographic or coming from other research domains.

---

[75] How data are recorded is sometimes referred to as the database model, while the visualisation of data is referred to as the cartographic model.

[76] An example of decreasing visualisation quality in order to speed it up is building additional raster data layers with a lower spatial resolution as compared to the original ones, so-called pyramid layers. Pyramid layers are used to visualise raster data in various spatial scales. If data are visualised in a small scale, a computer uses low resolution data (so smaller files) which can be displayed quickly. With zooming in and increasing the scale of visualisation, data with better spatial resolutions are used, finally, at some scale threshold, the original data are used (but then, typically, only part of the raster layer is shown on screen and the file size used is comparable to what is used at the small scale).

The methods of presenting geographic phenomena on paper maps – that is methods of cartographic presentation – had been elaborated long before any idea of digital representation of geographic data emerged. This is why geography together with cartography accepted first a graphical representation of geographic data instead of a digital one. The graphical representation, however, has many disadvantages and remains ineffective if spatial analysis is concerned. It is worth changing then how geographers conceive representation of geographic phenomena: the rules and models of digital representation should be considered primary rather than the graphical representation commonly used in paper maps. Essentially, a paper map – or any other analogue representation of geographic data – does not need to constitute any part of the transformation process leading from the real world to its representation. It may be, however (and quite frequently is) one of the outcomes of a successful completion of a geographic database, analysis of its content and synthesis of research outcomes based on collected data. Longley et al. (2011, p. 91), referring to the relation of paper maps and digital representations, make a very good point stating: *So while the paper map is a useful metaphor for the contents of a geographic database, we must be careful not to let it limit our thinking about what is possible in the way of representation.*

For reasons explained above I have deemed that in 2022 a paper map issue can be mostly omitted when discussing geographic phenomena and their representation. Obviously, a paper map is necessary whenever geographic data visualisation is considered, as well as their analysis, because the analysis of geographic data requires a preliminary, visual interpretation of data by researchers at the stage of geodatabase exploration. This remark is made just in case someone might interpret my words as a partisan manifesto of teaching geographic information without maps: no, I do not support such an extreme view, but I think that a paper map should occupy its rightful place in such teaching.[77]

So what is the equivalent of a paper map in 2022? With no doubt two ideas that appeared around the turn of the 21st century give some impression as to the future evolution of geographic databases and current access to geographic data for users: spatial information infrastructures and digital globes. Spatial information infrastructures organise acquisition and access to geographic data produced by officially recognised institutions that, to some extent, guarantee high data quality. In the European Union, a spatial information infrastructure has been developed since 2000s based on the INSPIRE Directive adopted in 2007 (Gaździcki 2009). Digital globes, with the best known example of their kind being Google Earth launched in 2005 (Dalton, 2013), are mostly commercial platforms providing access not only to institutional datasets but also to a range of geographic data produced by various users, for instance Volunteered Geographic Information (VGI; Goodchild, 2007), and a variety of remotely sensed image data (Harvey, Kozak, 2011). The services and visualisation tools in geoportals, digital globes and in GIS software allow one to access vast Internet resources of geographic data and offer various capabilities of data analysis as well as their visualisation. These tools

[77] There are several textbooks on geographic information, and even the best ones tackle a paper map among other considerations of representing geographic phenomena and geographic data models. For instance, a textbook cited many times in this book, *Geographic Information Systems & Science*, by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (2011, 3rd edition) discusses paper maps in Chapter 3 *Representing Geography*, together with vector and raster geographic data models (in addition, Chapter 12 of the textbook *Cartography and Map Production* is also dedicated to maps). In *A Primer of GIS. Fundamental Geographic and Cartographic Concepts* written by F. Harvey (2008), a paper map is referred to frequently in the first chapters that deal with geographic phenomena and their representation. In turn, the textbook *An Introduction to Geographical Information Systems* by I. Heywood, S. Cornelius and S. Carver (2006) presents a paper map and some related problems like scale and generalisation relatively early, in Chapter 2 that deals with geographic data. While I may agree with the statement of the authors that maps have shaped how we conceptualise space, it is much more difficult to accept that an in-depth study of various properties of geographic data has to start with a paper map (p. 35). One of the figures (Figure 3.1, p. 72) is not convincing in how it situates a paper map in between the real world and raster and vector geographic data models. In my opinion combining the problems related to paper maps with the digital representation of geographic data leads to confusion about the appropriate sequence of translating real world phenomena and their properties first to their digital representation, and only later to a visual one.

allow users to immerse themselves in the wealth of these data and to interact with an increasing realism with the real world representation that they form. Craglia et al. (2008) envisage a future as an integration of spatial information infrastructures and digital globes leading to data rich platforms, offering high quality and up-to-date data easily accessible through intuitive visualisation tools and allowing analysis with the full scope of complexity. An example of these future directions is the European programme *Destination Earth*, aiming at creating an accurate digital model of the Earth – a digital twin – that will allow us to analyse the interactions of human – environmental systems.[78] The future, then, is an intricate, multi-faceted and dynamic digital representation of the Earth which will allow us to easily extract desired information. Whether this future will materialise, is something we will discover in the next few years.

[78] *Destination Earth*, accessible at https://digital-strategy.ec.europa.eu/en/policies/destination-earth, accessed: October 2021.

References

Adamczyk J., Będkowski K., 2007, Metody cyfrowe w teledetekcji. Wydawnictwo SGGW, Warszawa

Baddeley A., 2010, Analysing spatial point patterns in R. Accessible at https://research.csiro.au/software/r-workshop-notes/, accessed July 2020

Bian L., 2007, Object-oriented representation of environmental phenomena: Is everything best represented as an object? Annals of the Association of American Geographers 97, 267–281, https://doi.org/10.1111/j.1467-8306.2007.00535.x

Bińczyk E., 2018, Epoka człowieka. Retoryka i marazm antropocenu. Wydawnictwo Naukowe PWN, Warszawa

Bogen J., 2017, Theory and Observation in Science. In: Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy. Accessible at https://plato.stanford.edu/archives/sum2017/entries/science-theory-observation/, accessed: April 2020

Bunge W., 1962, Theoretical Geography. The Royal University of Lund, Lund, Lund Studies in Geography. Ser. C, General and Mathematical Geography 1

Casati R., Varzi A., 2017 Events. In: Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy. Accessible at https://plato.stanford.edu/archives/sum2020/entries/events/, accessed: March 2020

Chakravartty A., 2017, Scientific Realism. In: Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy. Accessible at https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/, accessed: March 2020

Chazdon R. L., Brancalion P. H. S., Laestadius L., Bennett-Curry A., Buckingham K., Kumar C., Moll-Rocek J., Guimarães Vieira I. C., Wilson S. J., 2016, When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. Ambio 45(5), 538–550, https://doi.org/10.1007/s13280-016-0772-y

Chorley R., Kennedy B., 1971, Physical Geography. A Systems Approach. Prentice-Hall International Inc., London

Corbett J., 2001, Torsten Hägerstrand, Time Geography. In: CSISS Classics. Accessible at https://escholarship.org/uc/item/2t75b8sj, accessed: June 2020

Couclelis H., 1992, People Manipulate Objects but Cultivate Fields. Beyond the Raster-Vector debate in GIS. Accessible at http://clamsitel.pbworks.com/w/file/fetch/49252349/Couclelis%201992%20Objects%20Fields.pdf, accessed: September 2020

Couclelis H, 2010, Ontologies of geographic information. International Journal of Geographical Information Science 24(12), 1785–1809. https://doi.org/10.1080/13658816.2010.484392

Craglia M., Goodchild M. F., Annoni A., Camara G., Gould M., Kuhn W., Mark D., Masser I., Maguire D., Liang S., Parsons E., 2008, A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. International Journal of Spatial Data Infrastructures Research 3, 146–167, https://doi.org/10.2902/1725-0463.2008.03.art9

Crosier S., 2001, Johann-Heinrich von Thünen, Balancing Land-Use Allocation with Transport Cost. In: CSISS Classics. Accessible at https://escholarship.org/uc/item/02r7z6qw, accessed: July 2020

Dalton C. M., 2013, Sovereigns, Spooks, and Hackers: An Early History of Google Geo Services and Map Mashups. Cartographica: The International Journal for Geographic Information and Geovisualization 48(4), 261–274, https://doi.org/10.3138/carto.48.4.1621

DiBiase D., DeMers M., Johnson A., Kemp K., Taylor Luck A., Plewe B., Wentz E., 2006, Geographic Information Science and Technology Body of Knowledge. Association of American Geographers, Washington, DC. Accessible at http://www.aag.org/cs/bok, accessed: March 2020

Forman R. T. T., Godron M., 1986, Landscape Ecology. John Wiley & Sons, Inc., New York

Frank A. U., 2001, Tiers of ontology and consistency constraints in geographical information systems. International Journal of Geographical Information Science 15(7), 667–678, https://doi.org/10.1080/13658810110061144

Galton A., 2001, Space, time and the representation of geographical reality. Topoi 20, 173–187, https://doi.org/10.1007/978-3-030-16829-2_4

Gangodagamage C., Zhou X., Lin H., 2008, Autocorrelation, Spatial. In: Shekhar S., Xiong H. (eds.), Encyclopedia of GIS. Springer, Boston, MA

Gaździcki J., 2004, Leksykon Geomatyczny. Accessible at https://www.ptip.info/leksykon, accessed: April 2020

Gaździcki J., 2009, Druga faza programu prac INSPIRE: stan w Europie i wyzwania dla Polski. Roczniki Geomatyki 7(6), 61–72

Goodchild M. F., 1992, Geographical information science. International Journal of Geographical Information Systems 6(1), 31–45, https://doi.org/10.1080/02693799208901893

Goodchild M. F., 2007, Citizens as sensors: The world of volunteered geography. GeoJournal 69, 211–221, https://doi.org/10.1007/s10708-007-9111-y

Goodchild M. F., Yuan M., Cova T. J., 2007, Towards a general theory of geographic representation in GIS. International Journal of Geographical Information Science 21(3), 239–260, https://doi.org/10.1080/13658810600965271

Haggett P., 2001, Geography: a modern synthesis. Harper & Row, New York, 3rd edition

Hansen M. C., Potapov P. V., Moore R., Hancher M., Turubanova S. A., Tyukavina A., Thau D., Stehman S. V., Goetz S. J., Loveland T. R., Kommareddy A., Egorov A., Chini L., Justice C. O., Townshend J. R. G., 2013, High-resolution global maps of 21st-century forest cover change. Science 342(6160), 850–853, https://doi.org/10.1126/science.1244693

Harvey F., 2008, A Primer of GIS. Fundamental Geographic and Cartographic Concepts. The Guilford Press, New York, London

Harvey F., Kozak J., 2011, Digital Earth – i co dalej? Roczniki Geomatyki 11(1), 7–18

Heywood I., Cornelius S., Carver S., 2006, An Introduction to Geographical Information Systems. Pearson Education Limited, Harlow, 3rd edition

Jezioro P., Kozak J., 2004. Wprowadzenie do GIS. Ćwiczenia. Wyd. 2, zmienione i poszerzone. Wydawnictwo TEXT.

Kadaj R., 2002, Polskie układy współrzędnych – formuły transformacyjne, algorytmy i programy. Accessible at http://www.geonet.net.pl/?informacje/publikacje.html, accessed November 2021

Kaim D., 2017, Land cover changes in the Polish Carpathians based on repeat photography. Carpathian Journal of Earth and Environmental Sciences 12(2), 485–498

Kozak J., 2004, Globalne cyfrowe modele wysokości – stan obecny i perspektywy. Teledetekcja środowiska 33, 119–124

Klimaszewski M., 1988, Rzeźba Tatr Polskich. PWN, Warszawa

Kwan M-P., 2004, GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. Geografiska Annaler: Series B, Human Geography 86(4), 267–280. http://dx.doi.org/10.1111/j.0435-3684.2004.00167.x

Langran G., 1993, Time in Geographic Information Systems, Taylor & Francis, London – Washington, DC

Lillesand T. M., Kiefer R. W., 1994, Remote Sensing and Image Interpretation. John Wiley & Sons, Inc., New York – Chichester – Brisbane – Toronto – Singapore, 3rd edition

Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W., 2011, Geographic Information Systems & Science. John Wiley & Sons, Inc., 3rd edition

Lund H. G., 1999, A `forest' by any other name. Environmental Science & Policy 2, 125–133

Maguire D. J., 1992, An overview and definition of GIS. In: D.J. Maguire, M.F. Goodchild, D.W. Rhind (eds.), Geographic Information Systems. Volume 1: Principles. John Wiley & Sons, 9–20. Accessible at http://www.wiley.com/legacy/wileychi/gis/volumes.html, accessed: September 2020

Micklin P., 2010, The past, present, and future Aral Sea. Lakes and Reservoirs: Research and Management 15(3), 193–213, https://doi.org/10.1111/j.1440-1770.2010.00437.x

Miller H. J., 2005, A measurement theory for time geography. Geographical Analysis 37(1), 17–45, https://doi.org/10.1111/j.1538-4632.2005.00575.x

Miller H. J., Dodge S., Miller J., Bohrer G., 2019, Towards an integrated science of movement: converging research on animal movement ecology and human mobility science. International Journal of Geographical Information Science 33(5), 855–876, https://doi.org/10.1080/13658816.2018.1564317

OGC City Geography Markup Language (CityGML) Encoding Standard, 2012, Open Geospatial Consortium, OGC 12-019. Accessible at https://www.ogc.org/standards/citygml, accessed: September 2020

Openshaw S., 1983, The modifiable areal unit problem. Concepts and Techniques in Modern Geography, 38 Geo Books, Norwich

Pasławski J. (ed.), 2006, Wprowadzenie do kartografii i topografii. Wydawnictwo Nowa Era, Warszawa

Peuquet D. J., 1988, Representations of Geographic Space: Toward a Conceptual Synthesis. Annals of the Association of American Geographers 78(3), 375–394

Peuquet D. J., 1990, A conceptual framework and comparison of spatial data models. In: Peuquet D. J., Marble D. F. (eds.), Introductory readings in Geographic Information Systems. Taylor & Francis, 250–285

Peuquet D. J., 2001, Making space for time: Issues in space-time data representation. GeoInformatica 5(1), 11–32, https://doi.org/10.1023/A:1011455820644

Rettler B., Bailey A.M., 2017, Object. In: Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy. Accessible at: https://plato.stanford.edu/archives/win2017/entries/object/, accessed: March 2020

Rozporządzenie Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych. Dz.U. 2012, poz. 1247

Schaefer F. K., 1953, Exceptionalism in Geography: A Methodological Examination. Annals of the Association of American Geographers 43(3), 226–249, https://doi.org/10.1080/00045605309352114

Schumm S. A, Lichty R. W., 1965, Time, space, and Causality in Geomorphology. American Journal of Science 263, 110–119

Sinton D. F., 1978, The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study. In: Dutton G. (ed.), Harvard Papers on Geographic Information Systems, First International Advanced Study Symposium on topological Data Structures for Geographic Information Systems

Smith B., Mark D. M., 2001, Geographical categories: An ontological investigation. International Journal of Geographical Information Science 15(7), 591–612, https://doi.org/10.1080/13658810110061199

Sui D. Z., 2004, Tobler's first law of geography: A big idea for a small world? Annals of the Association of American Geographers 94, 269–277, https://doi.org/10.1111/j.1467-8306.2004.09402003.x

Śledziński J., not dated, Krótki wykład. Alfabet GPS. Accessible at https://geoforum.pl/gnss/alfabet-gps, accessed: June 2020

Tal E., 2017, Measurement in Science. In: Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy. Accessible at https://plato.stanford.edu/archives/fall2017/entries/measurement-science/, accessed: April 2020

Terms and Definitions FRA 2020, 2018, Forest Resources Assessment Working Paper 188, Rome. Accessible at http://www.fao.org/forest-resources-assessment/en/, accessed: September 2020

Theobald D. M., 2001, Topology revisited: Representing spatial relations. International Journal of Geographical Information Science 15(8), 689–705, https://doi.org/10.1080/13658810110074519

Thomasson A. L., 2001, Geographic objects and the science of geography. Topoi 20, 149–159, https://doi.org/10.1007/978-3-030-16829-2_8

Tomlin C. D., 1991, Cartographic Modelling. In: D.J. Maguire, M.F. Goodchild, D.W. Rhind (eds.), Geographic Information Systems. Volume 1: Principles. John Wiley & Sons, 361–374. Accessible at http://www.wiley.com/legacy/wileychi/gis/volumes.html, accessed September 2020.

Turner B. L., 2002, Contested identities: Human-environment geography and disciplinary implications in a restructuring academy. Annals of the Association of American Geographers 92(1), 52–74, https://doi.org/10.1111/1467-8306.00279

Ustawa z dnia 28 września 1991 r. o lasach. Dz.U. 1991 nr 101, poz. 444.

Weisstein E.W., 2020, Delaunay Triangulation. Accessible at MathWorld – A Wolfram Web Resource. https://mathworld.wolfram.com/DelaunayTriangulation.html, accessed: May 2020

Woleński J., 2007, Epistemologia. Wydawnictwo Naukowe PWN, Warszawa.

Worboys M. F., 1999, Relational database and beyond. In: Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W (eds.), Geographical Information Systems. Principles, Techniques, Management and Applications, John Wiley & Sons, 373–384

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES