

---

# **Principles of GIS and Remote Sensing**

***Release 4.0.0***

**ITC - Faculty of Geo-information Science and Earth Observation**

**Dec 06, 2021**



# INTRODUCTION

<b>1</b>	<b>About this repository</b>	<b>1</b>
1.1	How to use these exercises . . . . .	1
1.2	Structure . . . . .	2
1.3	Conventions . . . . .	2
1.4	Reporting Issues . . . . .	2
1.5	License . . . . .	3
1.6	Citation . . . . .	3
1.7	Inquires . . . . .	3
1.8	References . . . . .	3
<b>2</b>	<b>Spatial Data Models</b>	<b>5</b>
2.1	Geographic Phenomena . . . . .	5
2.2	Computer Representations . . . . .	6
<b>3</b>	<b>Data Entry</b>	<b>17</b>
3.1	Direct Spatial Data Acquisition . . . . .	17
3.2	Indirect Spatial Data Acquisition . . . . .	17
<b>4</b>	<b>Spatial Reference Systems</b>	<b>23</b>
4.1	Reference Surfaces . . . . .	23
4.2	Coordinate Systems . . . . .	24
4.3	Map Projections . . . . .	25
4.4	Coordinate Transformation . . . . .	26
4.5	Ground Control Points . . . . .	26
4.6	Additional Questions . . . . .	27
<b>5</b>	<b>Databases</b>	<b>29</b>
5.1	Database Management Systems . . . . .	29
5.2	Relational Databases . . . . .	31
5.3	Data Retrieval . . . . .	33
<b>6</b>	<b>Vector Analysis</b>	<b>43</b>
6.1	Geometric Measures . . . . .	43
6.2	Overlays . . . . .	46
6.3	Proximity Operators . . . . .	47
6.4	Networks . . . . .	49
<b>7</b>	<b>Network Analysis</b>	<b>53</b>
7.1	Optimal path analysis . . . . .	54
7.2	Advanced optimal path analysis . . . . .	58
7.3	Network allocation . . . . .	64

7.4	Trace analysis . . . . .	68
7.5	Challenge . . . . .	70
<b>8</b>	<b>Raster Analysis</b>	<b>71</b>
8.1	Reclassification . . . . .	71
8.2	Automatic Classification . . . . .	71
8.3	Manual Classification . . . . .	73
8.4	Raster Overlay . . . . .	77
8.5	Raster Measurements & Computations . . . . .	81
8.6	Surface Analysis . . . . .	86
<b>9</b>	<b>Data Quality</b>	<b>89</b>
9.1	Accuracy . . . . .	89
9.2	Data Acquisition Errors . . . . .	93
9.3	Error Propagation . . . . .	96
9.4	Data Management and Metadata . . . . .	100
<b>10</b>	<b>Visualisation</b>	<b>103</b>
10.1	Cartographic Grammar . . . . .	103
10.2	Critical Map Viewing . . . . .	104
10.3	Creating a Topographic Map . . . . .	105
<b>11</b>	<b>Radiometric Operations</b>	<b>109</b>
11.1	Image Display . . . . .	109
11.2	Image Enhancement by Histogram Operations . . . . .	117
11.3	Image Enhancement by Filter Operations . . . . .	123
<b>12</b>	<b>Radiometric Corrections</b>	<b>131</b>
12.1	Haze correction . . . . .	132
12.2	Sun Angle Correction . . . . .	134
<b>13</b>	<b>Digital Image Classification</b>	<b>135</b>
13.1	Inspection of Image characteristics . . . . .	137
13.2	Legend & Visual Interpretation . . . . .	137
13.3	Field Survey . . . . .	138
13.4	Organising Field data . . . . .	138
13.5	Defining Mapping Classes . . . . .	139
13.6	Defining Spectral Signatures (ROIs) . . . . .	139
13.7	Evaluation of Spectral Signatures . . . . .	140
13.8	Supervised Image Classification . . . . .	140
13.9	Accuracy assessment . . . . .	142
<b>14</b>	<b>Systems and Models</b>	<b>143</b>
14.1	Systems . . . . .	143
14.2	Models . . . . .	144
14.3	The Purpose of a Model . . . . .	145
14.4	Modelling Processes . . . . .	146
14.5	Evaluating Systems Models . . . . .	147
<b>15</b>	<b>Spatial Data Infrastructure</b>	<b>155</b>
15.1	The PDOK Geoportal . . . . .	155
<b>16</b>	<b>Use and Users</b>	<b>161</b>
<b>17</b>	<b>Resampling</b>	<b>163</b>
17.1	Basic Resampling . . . . .	164




17.2	Advanced Resampling . . . . .	169
<b>18</b>	<b>Mono and Multi-Temporal Classification</b>	<b>171</b>
18.1	Setting Up the Analysis Environment . . . . .	172
18.2	Data Preparation . . . . .	173
18.3	Classification . . . . .	179
18.4	Analysis . . . . .	180
<b>19</b>	<b>Change Detection</b>	<b>183</b>
19.1	The Sistan Basin . . . . .	183
19.2	Understanding the Data . . . . .	185
19.3	Overall Concentration of Environmental Variables . . . . .	186
19.4	Identification and Comparison of Seasonal Changes . . . . .	190
19.5	Quantifying Differences in the Presence of Water . . . . .	192
19.6	Quantifying Changes in Water and Vegetation . . . . .	196
<b>20</b>	<b>Common GIS File Formats</b>	<b>201</b>
20.1	Vector GIS File Extensions . . . . .	201
20.2	Raster GIS file extensions . . . . .	203
<b>21</b>	<b>Shaded Terrain Models</b>	<b>205</b>
21.1	Depicting a DEM as Layer Tints . . . . .	205
21.2	Using a DEM to Create a Hillshade . . . . .	206
<b>22</b>	<b>Sistan Basin Dataset</b>	<b>211</b>
	<b>Bibliography</b>	<b>213</b>



## ABOUT THIS REPOSITORY

Welcome to the **Principles and Applications of Geographic Information Systems and Earth Observation open courseware**. These exercises are developed and maintained by the [Faculty of Geo-Information Science and Earth Observation \(ITC\)](#), University of Twente, The Netherlands. All the exercises were originally developed for the [Master of Science in Geo-Information Science and Earth Observation](#), and are still used in our courses. They are made available to the geo community as a statement of ITC's commitment towards education and capacity development. Anyone can use them provided the license terms are respected (see the License section). The exercises assume the use of the free and open-source software [QGIS](#) but are not meant to be a software training. Instead, the activities are designed to teach key concepts of Geographic Information Science and Earth Observation. You are free to use them in your own classes (provided you include the attribution credits) or on your own as an independent learner.

### 1.1 How to use these exercises

On the left pane, there is a list of topics; each of those topics contains (primarily) software-based exercises designed to demonstrate related concepts. Although the arrangement of the topics reflects the order we recommend the exercises be taken; they do not build on each other and therefore can be taken in any order. Along with the exercise description, you may be suggested to take readings explaining a concept. The  icon signals a link to the [Living Textbook](#). The Living Textbook consists of a conceptual map built on top of a wiki that allows the user to read about a certain concept and explore how that concept relates to other concepts. The exercises might also contain questions. In a supervised environment, we use those questions to stimulate the class and assess how the learning process is going. We do not provide answer sheets - you will have to verify your answers independently if you choose to use this repository without support. Since we assume the use of [QGIS](#), consider keeping at hand the following resources:

- The [QGIS manual](#) can help you to find answers on how to use the tools and interfaces, and it is the first source you should consult when you have a question on how to use [QGIS](#).
- If you have more advanced questions, such as how to implement a rather specific procedure or data analysis related to GIS and [QGIS](#), then [GIS Stack exchange](#) might help you with that.
- You could also subscribe to the [QGIS-Users mailing list](#). The mailing list contains an extensive archive of Q&As from the [QGIS](#) user community. You can also post your questions, and someone from the community might answer.

## 1.2 Structure

This repository is divided into five parts:

**Part One**, the Introduction, contains general information about the exercises. This is the part you are reading right now.

**Part two** contains exercises related to Geographic Information Science where you can learn about Geographic Information Systems and Spatial Analysis.

**Part three** contains exercises related to Earth Observation, where you can learn about Remote Sensing and Image Interpretation.

**Part four** contains exercises related to Data Integration, where you can learn about system dynamics and approaches to integrate data from multiple sources and perform advanced analysis.

**Part five**, the Appendices, contain additional information that will be useful for some of the exercises.

## 1.3 Conventions

Along with this document, we use some formatting conventions to highlight content that requires special attention. These are as follows:

---

**Note: QGIS.** This will provide additional explanations specific to how QGIS works.

---

---

**Note: Reflection.** This will describe situations or post questions that require a deep level of reasoning. You should see this content as *mental puzzles* that will help to broaden your understanding of certain topics.

---

---

**Important: Resources.** This will describe the software and datasets required for completing a certain exercise. Not all exercises include the use of data; therefore, this will appear only in the sections that need it.

---

**Attention: Question.** This will post questions that you have to answer during the exercises. Write down your answer and take them to the virtual classroom.

## 1.4 Reporting Issues

If you find an error or inconsistency in this website, you can report it by opening an issue in this [GitHub repository](#). Alternatively, you can also notify us by email to [a.dasilvamano@utwente.nl](mailto:a.dasilvamano@utwente.nl).

**Include the following:**

1. A short but precise description of the issue,
2. A link to the title of the specific section that contains the issues. *You can find the URL of a title's section by clicking on the **chain icon** on the right side of the title's section.*
3. Optionally, include a screenshot of the part of the section containing the issue.

**Relevant issues to report:**

- Broken links.
- Wrong referencing, e.g., links do not refer to the correct resources.
- Missing or incomplete content. Text or figures.
- Grammatical errors. We use British spelling.

## 1.5 License

## 1.6 Citation

You can cite this repository using this reference: Mano, Andre (ed); Augustjin, Ellen-Wien; Bakx, Wan; Garcia, Manuel; & Knippers, Richard. (2021, August 17). andremano/online-core: Principles and Applications of GIS and RS. Zenodo. <https://doi.org/10.5281/zenodo.5213896>

## 1.7 Inquires

If you want to have this or other courses taught at your organization under the guidance of a teacher from our faculty, you can contact the editor and maintainer of this repository, Mr. André Mano ([a.dasilvamano@utwente.nl](mailto:a.dasilvamano@utwente.nl))

## 1.8 References


Andre Mano, Ellen-Wien Augustjin (2021): Teaching Geographic Information Science concepts with QGIS and the Living Textbook. Towards a sustainable and inclusive Distance Education. FOSS4G Asia. Kathmandu, Nepal.

*Section author: Andre Mano and Manuel G. Garcia*







## SPATIAL DATA MODELS

### 2.1 Geographic Phenomena

 **Geographic phenomena** is a term for all phenomena with a spatial dimension, such as landcover or air pollution. Phenomena are geographic (spatial) when they:

- Can be named or described;
- Can be georeferenced (have a location on the Earth's surface);
- Can be assigned a time interval.

Geographic Phenomena can be divided into  **geographic objects** and  **geographic fields**, and Geographic Fields can represent geographic phenomena as  **discrete fields** or  **continuous fields**.

**Attention: Question.** Give some examples of phenomena for which one of the three elements above is missing. Can they be used in a GIS?

#### 2.1.1 Objects and Fields

**Task 1** Complete the table below by indicating if the phenomenon on the left is an object or field and if it is discrete or continuous.

Phenomena	Field or Object	Discrete or Continuous
Wind direction		
Temperature		
Roads		
Buildings		
Field sample points		
Landuse		
Soil Type		

## 2.1.2 Object's Parameters

**Task 2** Which locational parameters can be used to describe objects? Not all the parameters that you identified are essential for all objects. Which of the locational parameters are important for the objects below?


Object	Parameter 1	Parameter 2	Parameter 3	Parameter n
House				
Road				
Well				
Lakes				
Rivers				

## 2.1.3 Crispy and Fuzzy Boundaries

Another notion that is important for describing Geographic Phenomena is  **boundaries**. We distinguish two types of boundaries: Crisp and Fuzzy.

**Attention: Question.** Can you give one example of a crisp boundary and one of a fuzzy boundary for any geographic phenomena?

## 2.1.4 Autocorrelation

There is one more concept that requires an introduction at this point, and that is  **Spatial autocorrelation**. Spatial autocorrelation is based on Tobler's first law of geography.

**Task 3** Define in your own words what spatial autocorrelation is.

Although all computer representations store data as finite representations, it is important that you realise that some phenomena show autocorrelation.

---

## 2.2 Computer Representations

The previous tasks did not use any software. This should not have surprised you because the focus, so far, was on understanding geographic phenomena and their characteristics. In this section, we will focus on showing how computers represent geographic phenomena. *A good understanding of geographic phenomena will help you in choosing a suitable computer representation.*

Here, you will explore different types of computer representations for geographic phenomena available in a GIS, and how to select the most suitable computer representation for a specific phenomenon. To get an overview of the different ways we can store geographic phenomena in a GIS, we are going to use the concept map of your Living textbook.

**Attention: Question.** Look for the term “Geographical Representation” in the Living textbook, and explore all related concepts two links apart. How many different types of representations of spatial data did you find?




**Important: Resources** You will require the latest LTR version of LTR version of QGIS, plus the dataset `data_modeling.zip`. When you unzip the dataset, you will find the following files inside:

- `Cities.csv` – a comma-separated values file with city names;
- `Spatial_data_modelling.qgs` – A QGIS project preloaded with datasets
- `elevation.tif` – a digital elevation model;
- `points.gpkg` – a vector dataset representing elevation points.

## 2.2.1 Tessellations

Tessellations are a way to represent geographic phenomena in a GIS. A tessellation splits the geographic space into small cells in such a way that the complete area is covered by them. They are like tiles on a floor or a wall. In most cases, such tiles are square cells, and when all cells are equal in size, we call this a regular tessellation.

In a  **Regular tessellation**, all cell have the same size and each cell is associated with a value (cell value). This value has a data type, such as integer or floating-point.

An integer data type is a number that does not contain any decimals. They are often used to indicate codes in a discrete field (e.g. a land use class). A float or floating-point data type is a number that may contain decimals. A floating-point data type that can store very big numbers (64bit) is known to have ‘double precision’, and it is often called “Double”. The table below shows a list of very common data types used in a GIS.

SHORT INTEGER	Numeric values without decimals within a specific range. Application: store coded values.
LONG INTEGER	Numeric values without decimals within a specific range. The range is larger than a short integer.
FLOAT	Numeric values with decimals within a specific range. Single precision (32bits).
DOUBLE	Numeric values with decimal within a specific range. Double precision (64bits).
TEXT	Names or other textual qualities.
DATE	Dates and times.

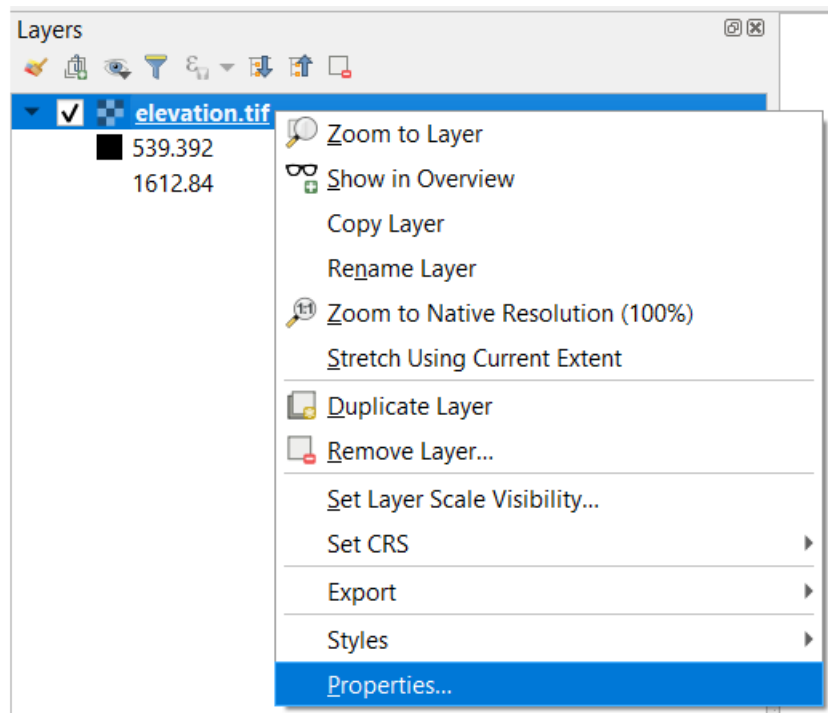
**Task 4** What data type would you use to represent a discrete field when you store it as a tessellation? And for a continuous field?

**Task 5** Boundaries in raster layers are both artificial and fixed. This has advantages and disadvantages. Can you give some examples of the advantages and disadvantages of artificial and fixed boundaries in raster layers?


**Task 6** Open the ‘*spatial\_data\_modelling*’ QGIS project we provided in the dataset, and explore the properties of the tessellation representing elevation (`elevation.tif`).

How many rows/columns do the `elevation.tif` data has? Are the values of type integer or floating-point? What are the minimum and maximum values?

*Hint: from the layers panel, right-click on the layer to access its Properties... Once in the properties dialog, look into the Information dialog.*



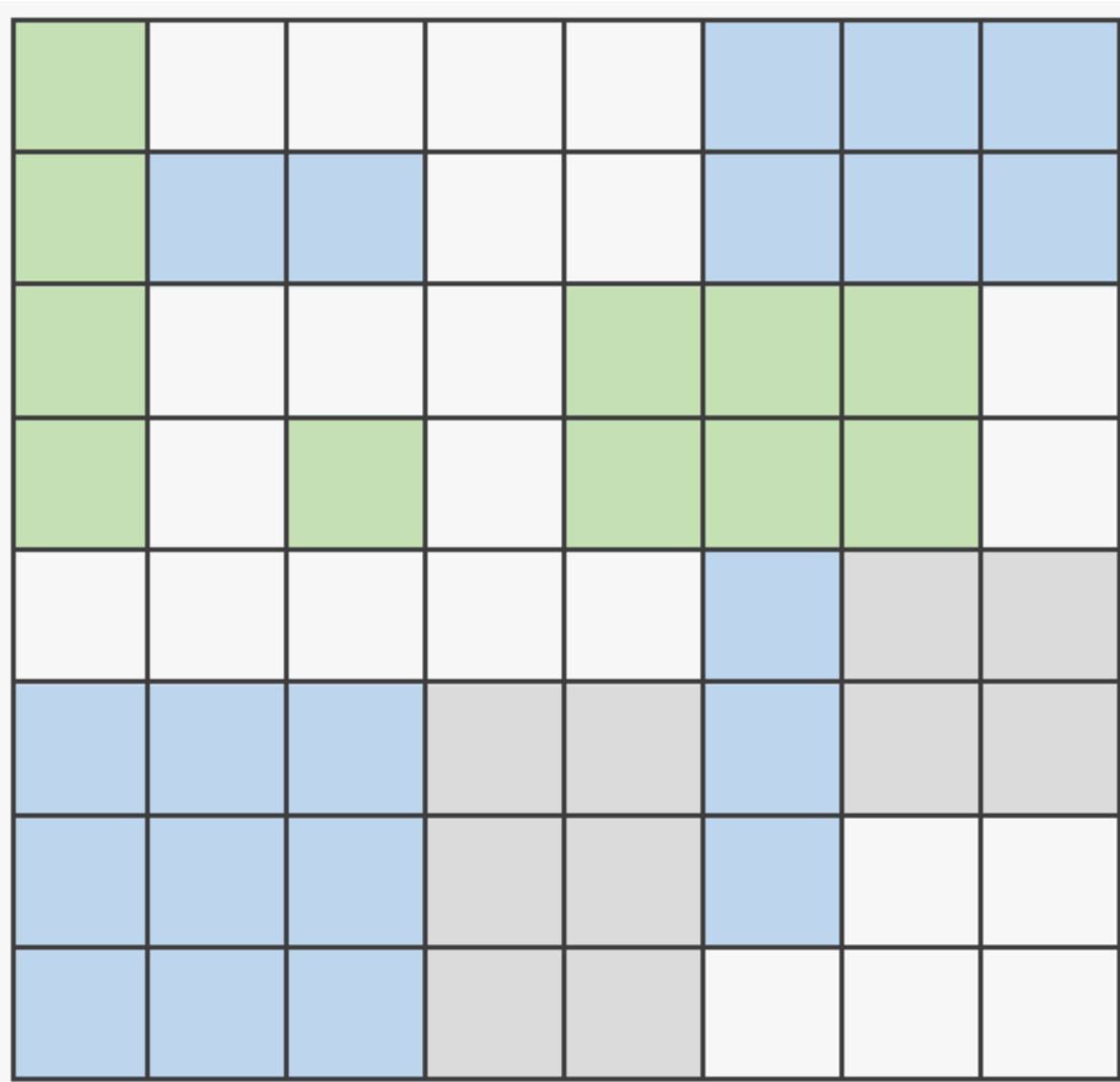
**Attention: Question.** What is the difference between a raster and a grid?

There are also  **Irregular tessellations**. In irregular tessellations a geographic area is partitioned into cells which are **not equal in size**.

**Attention: Question.** It is often stated that irregular tessellations are more adaptive compared to regular tessellations. What exactly is meant by this?

**Task 7** Although there are multiple examples of irregular tessellations, you only have to study one example: “the Quadtree”. If you are not familiar with Quadtrees yet, refer to your Living Textbook for more information. The best way to learn how Quadtrees work is to construct one manually.

Construct the Quadtree for the raster layer shown below.




**Attention: Question.** Using a Quadtree to represent a geographic phenomenon improves computation performance (computations are faster). Do you understand how this works?

**Task 8** Calculate the area of the green, blue and white cells in the Quadtree in each level of the Quadtree. Assume the size of each original cell is  $100 \times 100 \text{ m}$ .

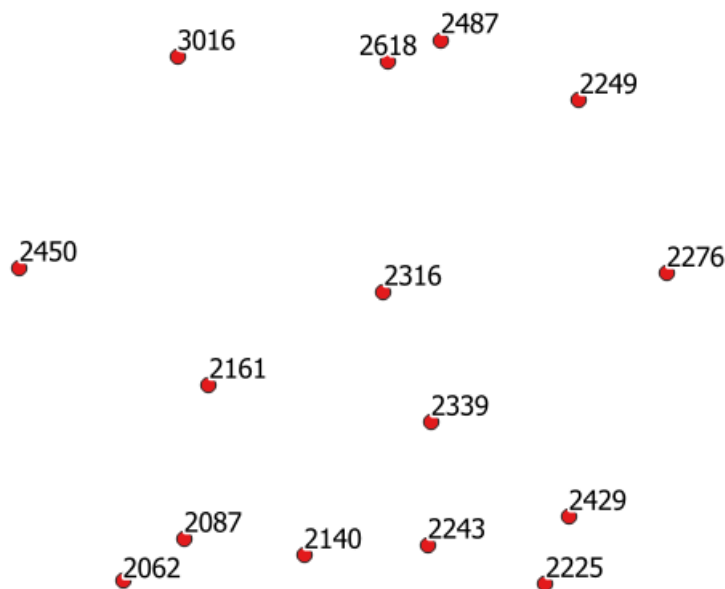
## 2.2.2 Vector Data Model

The main difference between our first data model (tessellation) and the vector data model is that tessellations do not explicitly store the georeference of the phenomena, but the vector data model does. This means that with every feature, coordinates are stored. In this section, we will discuss four examples of vector data representations: Triangulated Irregular Networks (TIN), Polygons, Lines and Points.

We start with the  **Triangulated Irregular Networks**. (TINs) because they have some characteristics in common with tessellations.

**Attention: Question.** Which characteristics have in common TINs and tessellations?

**Task 9** Using the picture below, manually create a TIN from the given points.



**Attention: Question.** You may be surprised, but not all triangulations are equally good. The standard triangulation is a Delaunay triangulation. Was your triangulation Delaunay?

**Task 10** In your QGIS project, you find a layer with points. Generate a Delaunay triangulation and compare the result with the tessellation you made.

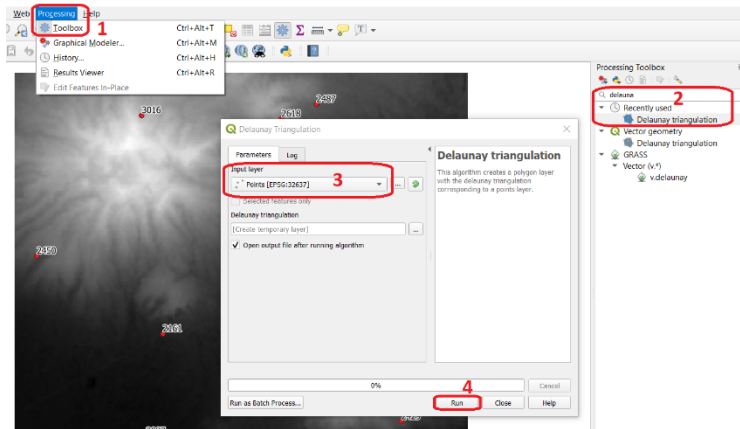


Fig. 2.1: Steps to create a Delaunay triangulation in QGIS

A triangulation can also be used to generate a continuous tessellated surface by means of interpolation. In those cases, each cell is assigned the value that is related to how far that cell is from the anchor points.

**Task 11** Create a tessellation using the **TIN interpolation** tool; use as input the anchor points you have in your QGIS project. Then, use the **Identify tool** to inspect the cell values.

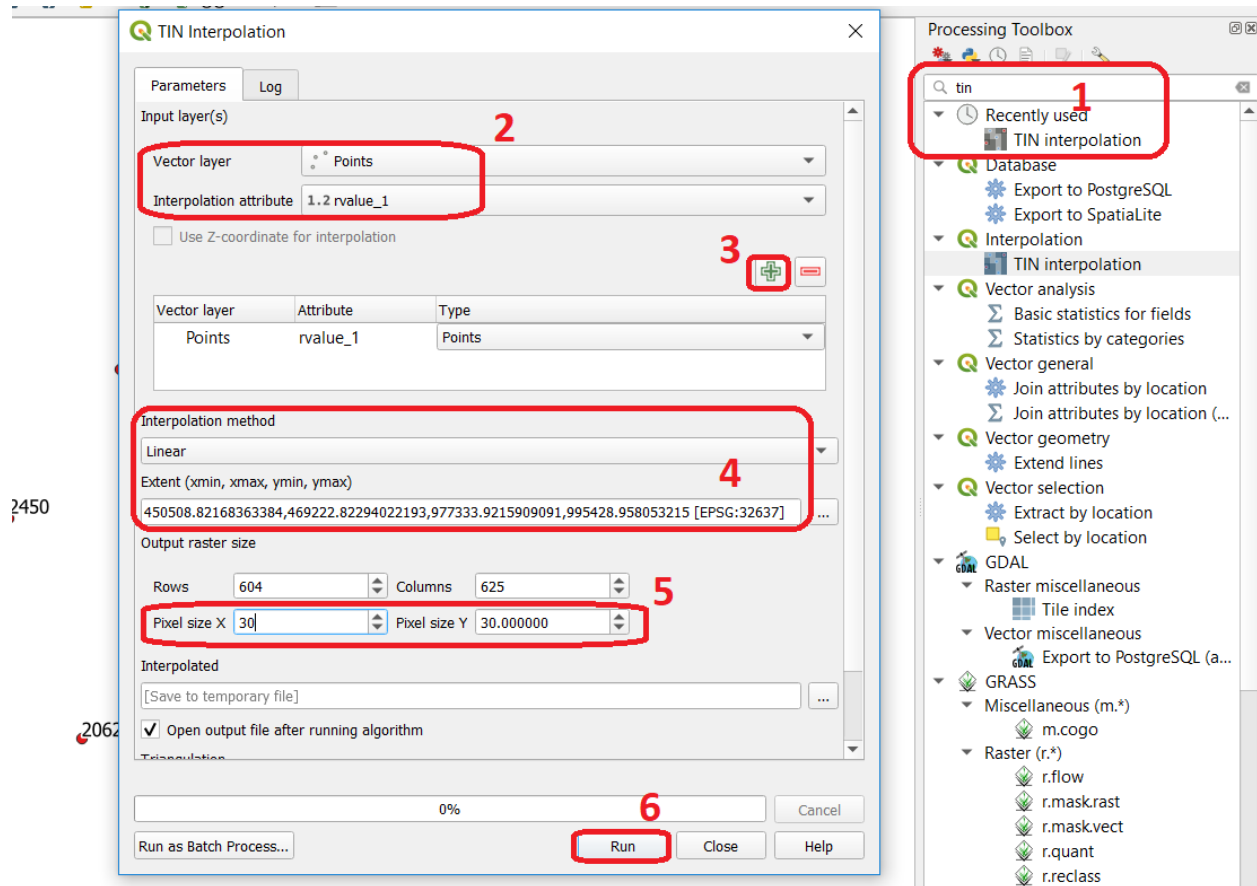


Fig. 2.2: Steps to create a tessellation from a TIN in QGIS

**Important: QGIS.** QGIS does not perform ‘on the fly interpolation’ – meaning that any point you click within your interpolated surface will have its value calculated on the spot. Instead, what QGIS does is to generate a tessellation of predefined cell size where each cell has a fixed value. ‘On the fly’ interpolations are supported in ArcGIS, for example; however, it is a functionality that will only exist within ArcGIS – the resulting data structure cannot be exported and used in other software packages.

**Attention: Question.** What exactly are the advantages of a TIN over a tessellation?

**Task 12** In your data, you find a table `Cities2.csv`. Try to use this table to create a point layer in QGIS. Start a new QGIS project and add the layer to QGIS using the *Delimited Text* option.

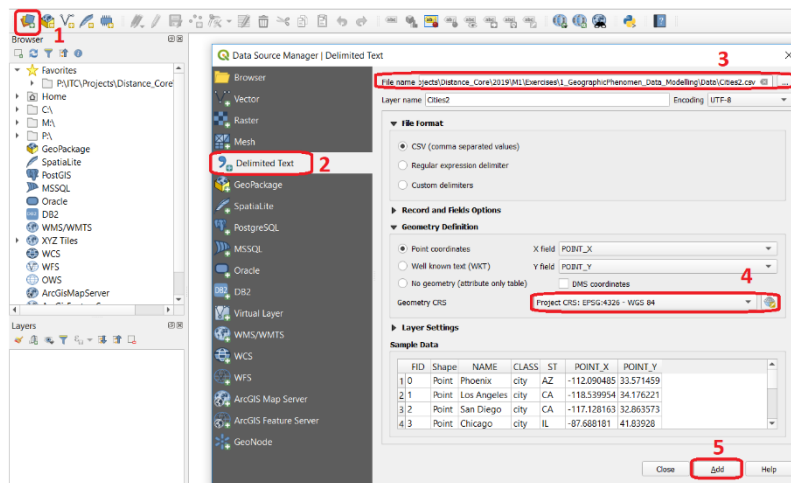



Fig. 2.3: Steps to create a point layer using a CSV file in QGIS

From the previous task, you should have clear that **points** are the simplest of geometries – they have a Y and X coordinates that anchors them to the spatial frame you are working on.


Another way of representing geographic phenomenon in the vector data model is using a  **Line representation**. A **line** is nothing more than two or more connected points.

**Attention: Question.** What is the difference between nodes and vertices, and how can we know the direction of a line?


The last representation in the vector data model is **polygons**. Polygons are one of the most well-known and commonly used vector data models. There are two important parts when using a polygon data model: the boundary model and the Topological model.

The boundary model explains how areas are represented and by storing the closed boundary that defines an area. A closed boundary is defined by a closed line (consisting of nodes and vertices, where the start and end vertices intersect). When representing the footprints of houses or the borders between countries, the boundary of each feature (house/country) is stored individually.

The Topological model is discussed in the the following section on *Topology*.


**Task 13** Read the section  **Area representation** and describe in your own words the problems that may arise when using the boundary model without topology.

2.2.3 Topology

The third topic in this exercise is  **Topology**. You will first have to understand what topology is before learning different ways to use it. Topological properties are geometric properties and spatial relations that are not affected by the continuous change of shape and size of a vector data layer (points, lines, or polygons).

**Task 14** Imagine you are looking at a map (take any map you like). Make a small list containing at least five examples of spatial topology that are visible in your map and five examples of properties that are not topological (use the table below).

Example	Topological	Non-topological
1		
2		
3		
4		
5		

When looking at two polygons, we can define all their possible topological relationships. To do so, we must describe each polygon in terms of its boundary and its interior (the area inside the boundary). Read  **Topological relationship**.

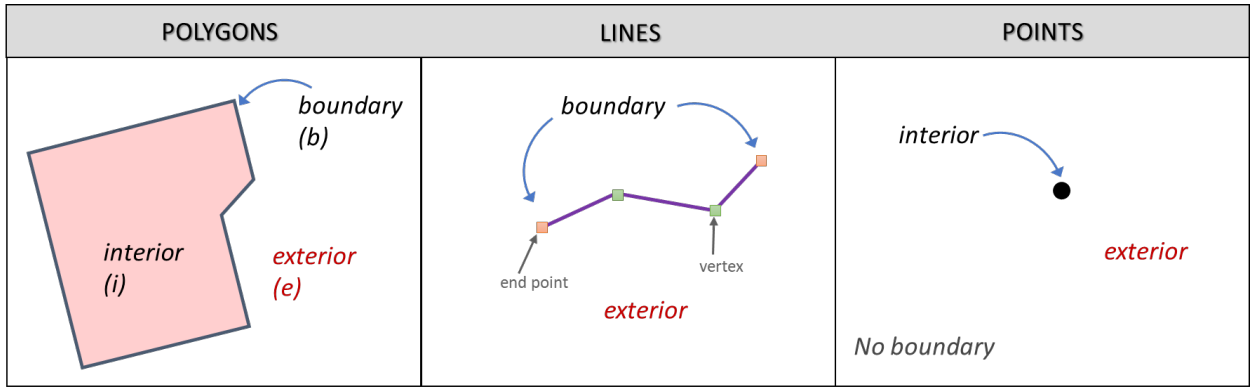

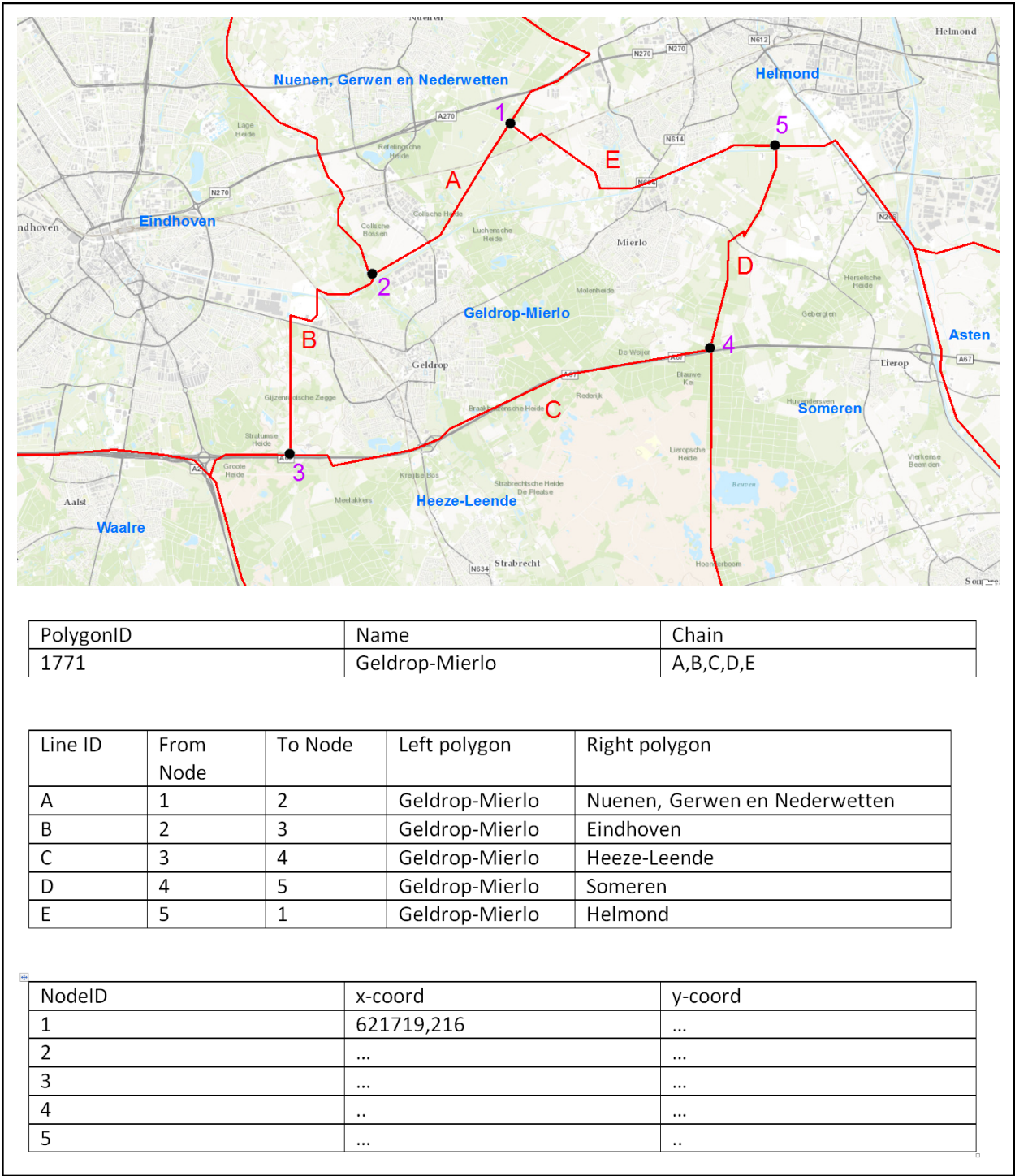


Fig. 2.4: The boundary, interior and exterior of polygons, lines and points

**Attention: Question.** What is the correct mathematical (set theory) expression that describes the *covers* relationship? How does this expression differ from the covered by relationship?

By now, you should understand what topology is, but you may wonder how it can be used. During the coming exercises, you will see many different uses, but for now, focus on an example given in the  **Topological data model**.

**Attention: Question.** For the map below, can you complete its corresponding attribute table following the topological data structure? The map below shows a polygon layer based on administrative units (municipalities). Focus your attention on the **Geldrop-Mierlo** municipality and its adjacent municipalities. The table below shows an example of the topological data structure for Geldrop-Mierlo.



Topology can also be used to ensure consistency of the geometries in a vector layer. There are five rules of [Topological consistency](#), which you should know about.

**Task 15** Identify for every example below which rule of topological consistency is violated.




Example	Rule(s)
The boundary of a polygon is not closed.	
Two lines cross each other without an intersection.	
There is a gap between two polygons.	
Two polygons overlap.	

Additional uses of topology will be discussed in the sections: *Data Entry*, *Using Spatial SQL* and *Networks*. In this course, **you only need to understand Topology in a conceptual level.**

**Attention: Question.** The following statements are made about time. What is your opinion on them? Are they true or false?

## 2.2.4 Temporal Dimension

In many situations, it is not enough to describe geographic phenomena only in terms of space, but also in terms of time because such phenomena change over time. The change may be relatively fast (like the clouds in the sky, hurricanes, and traffic) or slow (like the movement of a glacier).

To including time in the representation of spatial data, we talk about the  **Spatial-temporal data model**. This model defines different types of change: *change in attributes*, *change in location* (movement) and *change in shape* (growth) or combinations of these three.

**Task 16** Below you see a list of different types of change and some combinations. Can you write down an example for each type?

Type of Change	Examples
Attribute	
Attribute and Location	
Attribute and Shape	
Location	
Location and Shape	
Attribute, Location and Shape	

**Attention: Question.** The following statements are made about time. What is your opinion on them? Are they true or false?

- Although time is continuous in nature, in a GIS it is always represented in a discrete manner.
- There are many examples of spatial phenomena for which valid time is simply unknown.
- Branching time should be looked at into the future, as the past is already known and has only one branch.
- Time granularity is comparable to the spatial concept of resolution.

### Note: Reflection.



1. So far, you used vector representation of area features stored as Shapefiles. Are these shapefiles storing topology? In other words, do Shapefiles use a topological data model?
2. In this exercise, we have mainly focussed on 2-D data modelling examples. Yet, the real world is 3D. Do you know any examples in which a real 3-D data model would be needed? Is there also a 3-D topology?




3. Which other compression techniques exist besides Quadrees?
  4. Besides rectangular cells, other shapes can be used. What are the advantages of using Hexagonal cells?
  5. Make a comparison between raster and vector data models and list the advantages and disadvantages of each one.
- 

*Section author: Ellen-Wien Augustijn, André da Silva Mano, Manuel Garcia Alvarez & Amy Corbin*

## DATA ENTRY

### 3.1 Direct Spatial Data Acquisition

The  acquisition of spatial data can be done from several sources. There has been an increase in data acquired (or produced) using  remotely sensed sources, such as satellite imagery.



Other sources of spatial data include  Aerial survey,  Terrestrial survey and  Crowdsourcing. The use of different spatial data sources implies that how suitable (strengths and weaknesses) each particular source is for a particular analysis, depends on the acquisition methods. In this section we focus on methods for direct or **primary** data acquisition.

**Attention: Question.** What are possible the advantages and disadvantage of using the data sources listed in the table below? Fill in at least one advantage and one disadvantage.

Data Source	Advantage	Disadvantage
Remote sensing		
Aerial survey <sup>1</sup>		
Terrestrial survey		
Crowdsourcing		

### 3.2 Indirect Spatial Data Acquisition

Although spatial data can be acquired from third-party sources like government agencies or specialised companies, there will always be the need to acquire your own data. This usually means ‘digitising’ also known as ‘vectorisation’ – the process of capturing objects from a raster base layer like a map or an aerial photograph as points, lines and polygons. In this section, we focus on methods for indirect or **secondary** data acquisition. Specifically, the main techniques used for *vectorisation*.

**Attention: Question.** Read and observe the relation between  Digitizing and  Scanning. Is Digitising the only way to turn a scanning into Spatial data?

---

<sup>1</sup> It should be noted that aerial surveys are a form of remote sensing, but not from space.


**Important: Resources.** You will require the latest LTR version of [QGIS](#), plus the dataset [data\\_entry.zip](#). When you unzip the dataset, you will find the following files inside:

- `data_entry.qgs` – a QGIS project file;
  - `checking_errors.qgs` – a QGIS project file;
  - `Pearl_Harbour_topographic_map_(1999).tif` – a raster map;
  - `Educational_facilities.csv` – tabular data;
  - `Polygons.gpk` – a polygon vector layer.
- 

### 3.2.1 Digitising

Extracting the data, you need from a raster base map to a vector layer starts with creating a new dataset (i.e. layer), where the features that are about to be created will be stored. Technically speaking, it is a simple task; however, you should always take a moment to assess the requirements before proceeding with the actual software operation.

Capturing elements from a base map is an abstraction exercise; this abstraction depends on the scale and purpose for which the data will be used. For example, think of airports; will you represent them (abstract them) as points or as polygons? The answer to this question will depend on how you are going to use the data. If you want to publish a world map of the major airports, probably you could depict them as points. But if you're going to map the accessibilities to a given airport, a larger scale will be needed; therefore, polygons might be better.

The  [attributes associated](#) with the geometries are another important aspect to consider. The choice of attributes depends not only on the scale and intended use, but it also depends on the availability of the data (e.g. what is the capacity of the airport? How does it rank on security? How many international connections does it serve? – would these be information you need to have? And if so, do you have access to this data?)

**Task 1** Start QGIS and open the `data_entry.qgs` project. Among others, you will see a layer named `Pearl_Harbour_topographic_map_(1999).tif` Observe the map and complete the table below, considering the following requirements:

- Think of at least three vector layers that can be acquired from the raster base map;
- Make sure all geometric types – Polygon, Line, Point are represented;
- For each layer think of at least two attributes.

LayerName	Geometric Type	Attribute 1	Attribute 2
water_lines	line	Id	length

**Task 2** Now that you know what you want to extract and how are you are going to abstract it, proceed with the creation of the new layers. Digitise at least three features per layer.




For this task, you may want to watch the video tutorial on [basic digitizing](#):

---

**Note: QGIS.** Refer to [Editing](#) for a detailed description of vector editing with QGIS.

---

### 3.2.2 Topological Consistency

 **Topology** refers to the spatial relationships that should exist among the geometries of a vector dataset, and it is based on the  **Topological data model**. Topology can be a complex subject, but we will take a very pragmatic approach and show you how to maintain the most common topological relationships  **topological relationships**: adjacency in polygons and connectivity of lines.

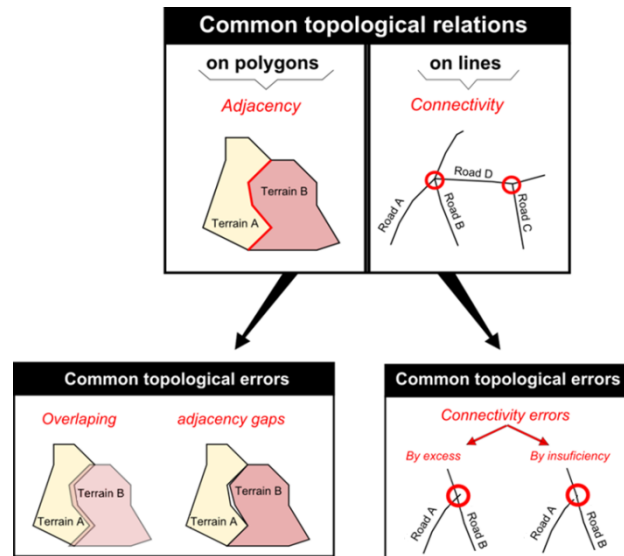


Fig. 3.1: Common topological relations on polygons, lines, and points

In the previous task, for the layer of geometry type 'Line' you probably digitised something that is supposed to be a network like roads or water lines. The key characteristic of a network is *connectivity*. However, if you happen to have digitised lines that are supposed to be connected and you zoom in to the point where the intersection is supposed to be, you will see that lines are not connected. Instead, you will see connectivity issues either by excess or by insufficiency (also known as *overshoots* and *undershoots* respectively).

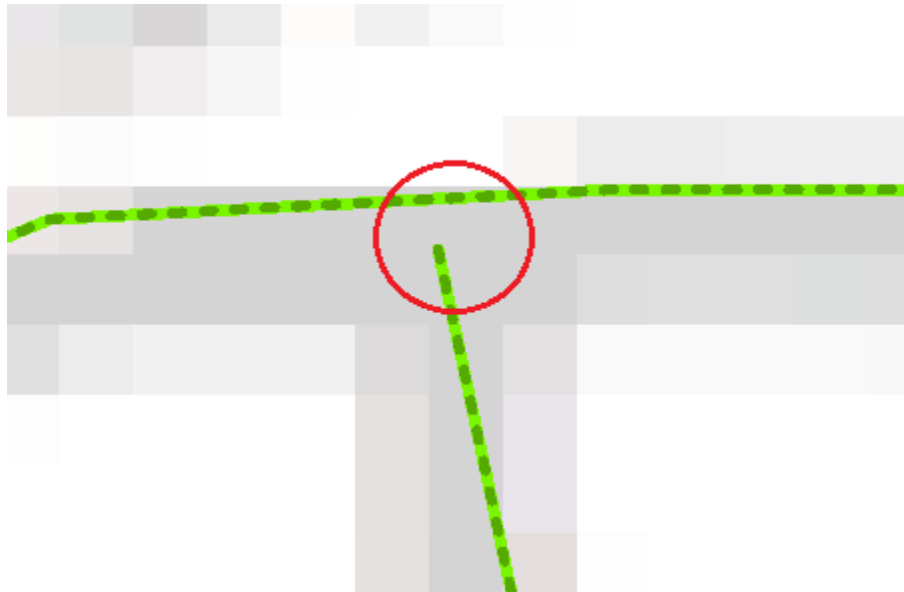



Fig. 3.2: Connectivity issues between lines. The case of undershooting

To ensure  **Topological consistency** between geometries, e.g., that line segments get properly connected while digitising, we have to set a snapping tolerance, which tells the GIS software to connect lines that are within certain distance automatically. Otherwise, it will be challenging to ensure that our lines are connected.

**Task 3** In QGIS, go to *Project > Spaning Options* and enable *Snapping mode*. Enter a tolerance of *20px* for every layer of lines that you may have.

If you may want to watch the video tutorial on [advance editing](#) :

**Task 4** Digitise some new lines making sure they are topologically connected. You will notice during digitising; if you go closer than a certain distance of an existing feature; the line would be automatically ‘pulled’ towards the nearest vertex or segment of the closest feature. You are thus ensuring connectivity.

In the case of polygons, it is also possible to ensure that adjacent polygons do not overlap.

**Attention: Question.**

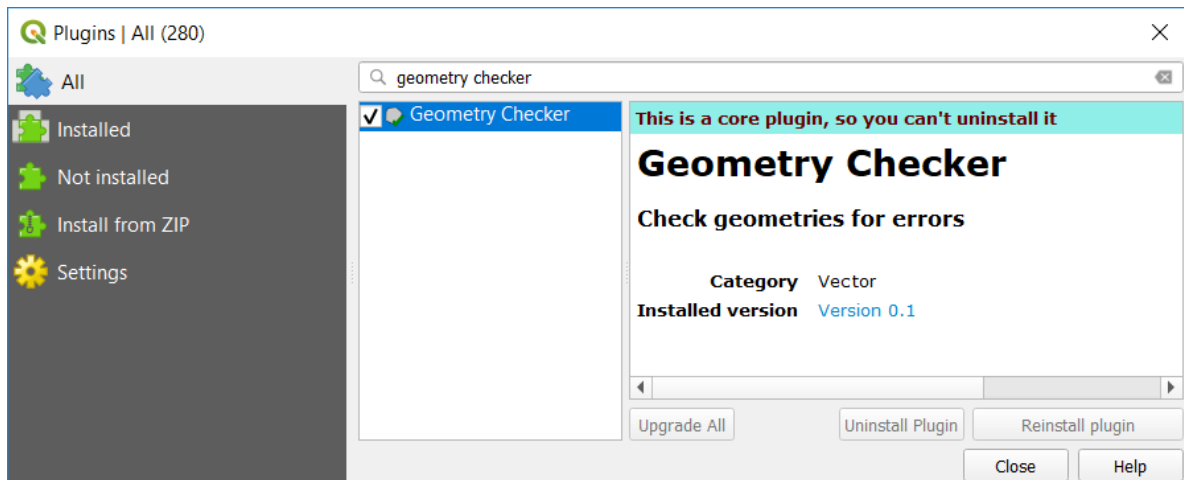
- How to define a snapping tolerance?
- What do the options ‘Enable topological editing’ and ‘Enable snapping on intersection’ allow you to do? Try to think of situations where these options might be useful.

---

**Note: Reflection.** Ensuring the topological consistency of your vector data is usually not that difficult if you are in control of the data acquisition technique (vectorisation) from the moment the dataset is created. Problems often arise when you receive datasets from third parties. When that happens, you should always do check to make sure the dataset maintains the basic topological relations.

---

**Task 5** Start QGIS and open the `checking_errors.qgs` project. You will see one layer (polygons). Find a way to check if there are overlapping or adjacency errors automatically. Tip: you may want to install and activate the **Geometry Checker** plugin. Once it is activated, it should be reachable from the Vector menu.



In some cases, detecting and fixing topological errors is not that simple. Just keep in mind that you should always check the integrity of the data you receive, especially if you do not know the source and lineage of the data.

### 3.2.3 Spatialising Data

Another way to acquire spatial data is by means of spatialising data. In other words, associate a geographic location with objects. This is a very common procedure when you get, for example, a spreadsheet or some sort of tabular data.

You can spatialise your data in two ways. By means of a *join* (a concept that will be explored later ahead in the course), or by means of building point geometries given that the tabular data contains X and Y coordinates.

**Task 6** Spatialising data. Open the `data_entry.qgs` project and create a point layer using the `educational_facilities.csv` file. Follow the steps depicted in the screenshot below.

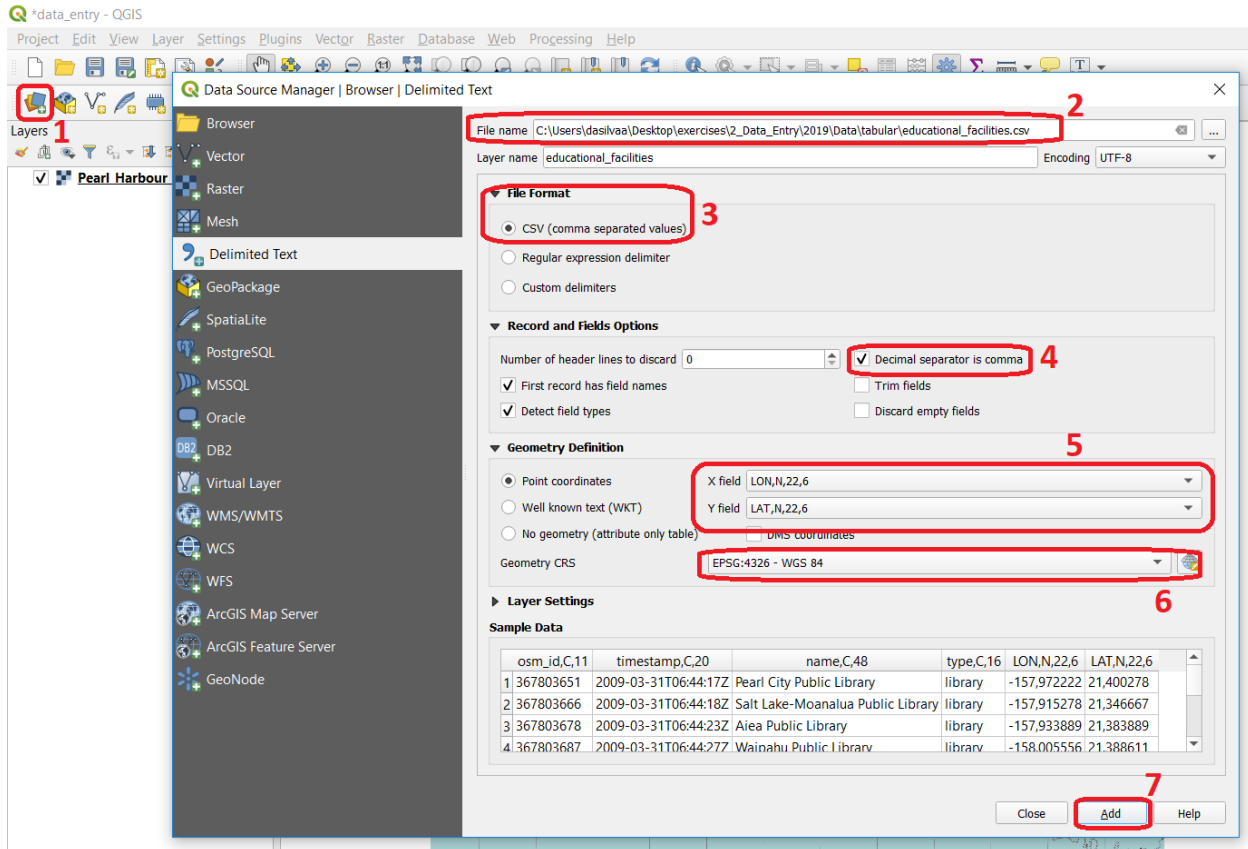


Fig. 3.3: Step to create a point layer from the 'educational\_facilities.csv' file

**Attention: Question.** If all went well, you should have ended up with a layer of points in your project. Does that mean that the `educational_facilities.csv` is spatial data?





In the Appendices section, you find a list of *Common GIS File Formats*.

Section author: André da Silva Mano & Manuel Garcia Alvarez



## SPATIAL REFERENCE SYSTEMS

### 4.1 Reference Surfaces

We use a  **Reference surface** (i.e. datum system) to approximate the shape of the Earth. A horizontal datum (also called geodetic datum) is a model used to measure positions on the Earth. A  **Vertical datum** is a model used to measure elevations, which is related to the  **Geoid**. The elevations are usually related to a mean sea level, and the horizontal position is measured by using an  **Ellipsoid** as reference.

**Attention: Question.** In the Netherlands, the elevation is measured using as a reference the water level in the canals of Amsterdam. This vertical datum is called *Normal Amsterdams Peil (NAP)*. A large part of the Netherlands would flood if we would not have any dunes or dikes, see Fig. 4.1. In fact, all land below 0 meters NAP would be the sea. Different countries measure height by referencing to different mean sea levels. Why do we not use one measure for all countries?




Fig. 4.1: Left: The Netherlands compared to sea level, according to Jan Arkesteijn. Right: The current coastline of the Netherlands.

**Attention: Question.**

- Why we use different reference surfaces for horizontal and vertical positioning?
- Define in your own words the difference between a local and global horizontal datum?
- What is the horizontal datum for your country?
- How are the horizontal and vertical datums connected?

---

## 4.2 Coordinate Systems



In mapping, a  **Coordinate system** is used to uniquely determine the position of a place with respect to the Earth's surface.


---

**Important: Resources.** You will require the latest LTR version of QGIS, plus the dataset [georeferencing.zip](#). When you unzip the dataset, you will find the following files inside:

- `DEM10.tif` (and auxiliary files) – a digital elevation model in raster format;
- `Topographical_map_dominica.tif` – a (ungeoreferenced) raster map;
- `Topographic_map_hengelo.tif` – a (ungeoreferenced) raster map;
- `Control_points.csv` – a table with points collected via GPS;
- `Floodzones.gpkg` – vector data (polygons) of floodable areas;
- `Highways.gpk` – line vector layer;
- `Parish.gpkg` – vectors representing administrative boundaries (parish level);
- `Rivers.gpkg` – line vector layer representing rivers.

---

**Task 1** Define in your own words the difference between  **coordinate systems** and  **planar coordinate systems**.

The most widely used global coordinate system is the  **Geographic coordinate system**. It consists of lines of geographic latitude and longitude. Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid. Lines of equal longitude are called meridians and form ellipses (meridian ellipses) on the ellipsoid.

Geographic coordinate systems use latitude and longitude to describe a position on the Earth's surface. *Latitude* is zero on the Equator and increases towards both poles to  $90^{\circ}N$  and  $90^{\circ}S$ . Latitude is positive on the northern hemisphere and negative on the South hemisphere. *Longitude* is measured from the meridian of Greenwich where it is zero, and it increases eastwards up to  $180^{\circ}$  ( $180^{\circ}E$ ), and westwards up to  $180^{\circ}$  ( $180^{\circ}W$ ). Longitude is positive towards the East and negative towards the West.

Geographic coordinates are typically expressed in two formats. In *Degrees, Minutes Seconds* or in *Decimal* formats. However, from a computational point of view, the decimal format should be preferred because it makes it easier for the computer to parse the values. If you use the decimal format, the computer will read it as a decimal number, but if you use the DMS format, the computer will most likely read it as a string of text with no numeric or mathematical meaning.

For example, Enschede is located at:

- Latitude:  $52^{\circ}13'05''N$  (DMS format) or  $52.21806$  (decimal format)

- Longitude: 65344E (DMS format) or 6.895556 (decimal format)

**Task 2** Use Google Maps to find what is in the following locations.

Latitude	Longitude	What
52°1305 N	6°5344 W	
52°1305 S	6°5344 E	
52°1305 S	6°5344 W	

**Task 3** Start a new project in QGIS and map the locations contained in the `control_points.csv` file. This file contains geographic coordinates captured using a GPS device. You may need to refer to the section *Spatialising Data* for this.



**Attention: Question.** What was the geographic coordinate system (geodetic datum) used in the `control_points.xlsx` file?

**Task 4** Open each of the remainder datasets in QGIS and check whether each dataset is spatially referenced or not.

**Attention: Question.**

1. Do you have any data in a geographic coordinate system?
2. What was the geographic coordinate system used?

## 4.3 Map Projections

A  **Map projection** is a mathematically described technique for representing the Earth's curved surface on a flat map. Flattening out a spherical surface is an imperfect task, as you can experience yourself when you are peeling an orange and try to lay flat the skin. To represent the Earth's surface on a map, we use a map projection. Map projections are developed for specific purposes, and each one has some distortions. Therefore, understanding the  **classification of map projections** is essential when choosing a suitable map projection.

No matter which map projection you choose, it always comes with certain distortions. You can experience this yourself for the case of the widely used Mercator Projection by using this link: <http://hive.sewanee.edu/pridepj0/286/mercatorMap.html>

**Attention: Question.** Suppose you wish to produce a small-scale map of your country. The map should show the population densities for the different regions (or provinces). What type of map projection would you suggest (consider projection class, property and other projection parameters)?

This interactive Map Projection Selection Tool can help to select a map projection <http://projectionwizard.org>


**Task 5** Load the vector and raster datasets you downloaded into a new QGIS project and answer the following questions:

- Are there any datasets that use a map projection?
- What geographic coordinate system is used by the projected datasets?

For this task, you might want first to watch the video [managing coordinate systems](#).

---

## 4.4 Coordinate Transformation

Users of geo-information often need to transform datasets from a particular coordinate system to another.  [Coordinate transformations](#) are necessary because some analysis requires planar measurement units. In other cases, the map-user have especial requirements on how to represent spatial data on a map.

**Attention: Question.** You should have noticed that the following datasets have different coordinate systems and datums: ‘DEM’, ‘floodzones’, ‘highways’, ‘parish’, and ‘rivers’.

- What problems can result out of leaving these layers with their current coordinate systems?
- If you had to transform all these datasets to a single coordinate system, which coordinate system will you choose as a target? What criteria determine your choice? Give examples for these specific datasets.

**Task 6** After choosing a projected coordinate system, and datum for your datasets, use QGIS to do the transformations. Put attention on which layer requires a transformation of projections or a transformation of projection and datum, because **the order in which you do transformation matters**.



**Note: QGIS.** You may need to refer to QGIS manual, section [projecting and transforming data](#).

---

**Task 7** Sometimes, when visualising data in QGIS, some of the layers do not align with each other. This may be because a data layer has an **unknown or incorrect coordinate** system defined. Or a data layer has no *georeferencing* ( no reference to any coordinate system at all). Do you have any data layer that does not align with the rest of the data layers in your current project?

---

## 4.5 Ground Control Points

You can load data into a GIS software which has not been georeferenced to any coordinate system—for example, a scanned topographic map or an image. In the next exercise, we will  [georeference](#) a scanned topographic map using  [Ground control points](#) to match the spatial reference of other layers in the project.

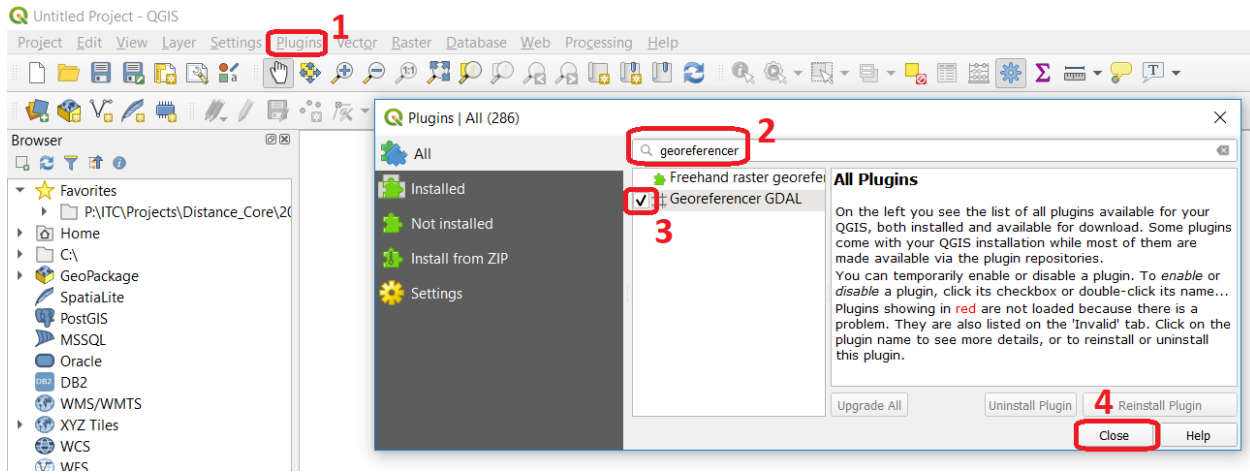
**Task 8** Load the `topographic_map_dominica.jpg` into a QGIS project and identify points on the topographic map that can be used as control points.

**Attention: Question.** Why are these points your best choice? How many points are you going to use?

**Task 9** In GIS, open the Georeferencer (from the Raster menu tools) and load the file `topographic_map_dominica.jpg`. Use the points you chose above to georeference the scanned map. Be sure to pay close attention to the coordinate system on the scanned map and your existing data layers. Attempt to match the map as best you can to your current vector layers.

For this task, you might want first to watch the video tutorial [georeferencing an image](#).

**Note: QGIS.** If the Georeferencer is not visible from the raster menu, you have to go to the plugins menu and make sure it is active.



**Attention: Question.**

- The quality of the georeferencing can be assessed by computing the  $RMSE$ . Do you understand how it works?
- What type of transformation did you apply in the georeferencing task?
- What can you say about the accuracy of the transformation?

## 4.6 Additional Questions

**Attention:**

1. Open the `topographic_map_hengelo.tif` of Enschede.
  - a. Is the image is georeferenced?
  - b. Zoom in to the legend at the center-bottom of the screen. How many coordinate systems are available in the topographic map?
  - c. Which of the three systems has a grid on the topographic map?
  - d. Which of the three systems you would use as coordinate system to georeference the digital version of the map?
2. Imagine the case where you import a digital map into your GIS system and after the import you realize that the digital map has coordinates but the coordinate system is unknown.
  - a. Do you think this case is uncommon?
  - b. What to do in this case? Do you have to georeference the map again or you just assign the Coordinate System in case you know it?

3. You receive 4 projected topographic maps printed on paper, the maps cover the same area:
- Map 1: has information on the ellipsoid, datum and projection
  - Map 2: has information on the ellipsoid and datum but not projection
  - Map 3: has information on ellipsoid and projection but not datum
  - Map 4: has information on datum and projection but not ellipsoid
- a. Out of these 4 maps how many maps have enough information to build a complete spatial reference system in order to digitize correctly the topographic maps?
- b. Can a topographic map with coordinates be digitized without knowing the coordinate system?

*Section author: Richard Knippers, André da Silva Mano & Manuel Garcia Alvarez*

## DATABASES

## 5.1 Database Management Systems


Begin by reading about  Database and  Database management system (DBMS) in the Living Textbook.

**Task 1** In the following table, you find the definitions for some of the most frequently used terms in database technology. Such as Database, Data, Dataset, Database management system (DBMS) and a Database system. Fill in the terminology column with the term that matches each definition.

Terminology	Definition
	It is a resource held on paper or in digital format that serves to record or administrate some facts and description of phenomena of interest
	A homogeneous collection of data usually describing a single kind of phenomenon.
	A collection of interrelated data sets properly structured by means of, and stored through a DBMS. (A file containing datasets)
	A software package that is designed to manage databases. This means, it allows to set-up, maintain and exploit one or more databases.
	Combination of a database and its DBMS.

**Attention: Question.** What would be the benefits of having a DBMS in your professional field? Can you also think of any disadvantages?

### 5.1.1 GIS and Spatial Databases

A  Geographic Information System should not be called GIS if any of these four components are missing:

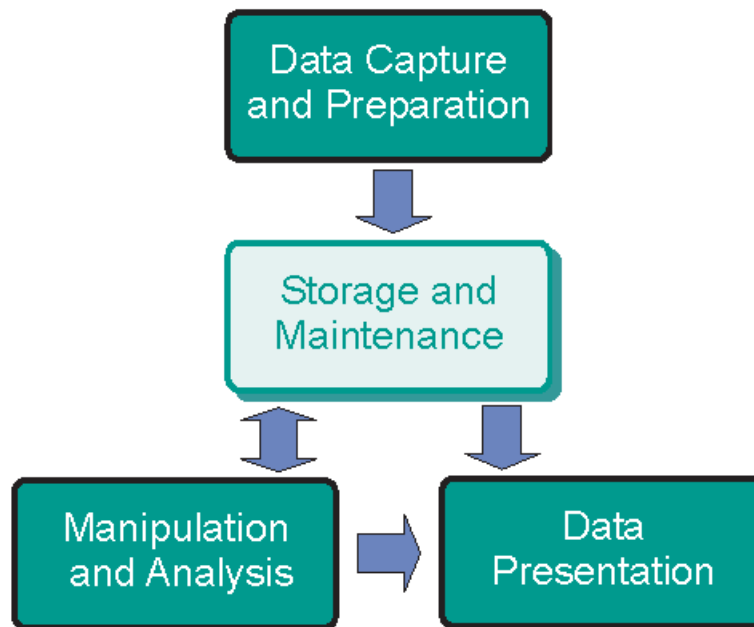





Fig. 5.1: The four components of a Geographic Information System

Behind the user interface of a GIS Software, there are things like application logic and data storage. Data can be stored as GIS files or in an (external) database. GIS software can use its DBMS or be connected to another DBMS such as PostgreSQL, Oracle, etc.

In GIS, in addition to the tabular data, we work with Spatial Objects. This requires a database that supports “spatial data”, which means that a database is capable of storing geometry data type.  Spatial databases provide spatial data handling in terms of storage, manipulation and analysis. A spatial database is not constrained by the need to present data visually.





A Spatial DBMS is a software package that:


- can work with underlying DBMS;
- supports spatial data models (vector, field), spatial data types, e.g. point, line, or polygon;
- provides functionality to  Query and manipulate spatial data using  Structured Query Language (SQL);
- supports spatial indexing (efficient algorithms for processing spatial operations) and domain-specific rules for query optimization.

<b>Attention: Question.</b> Which are some of the most popular Spatial DBMS?
--

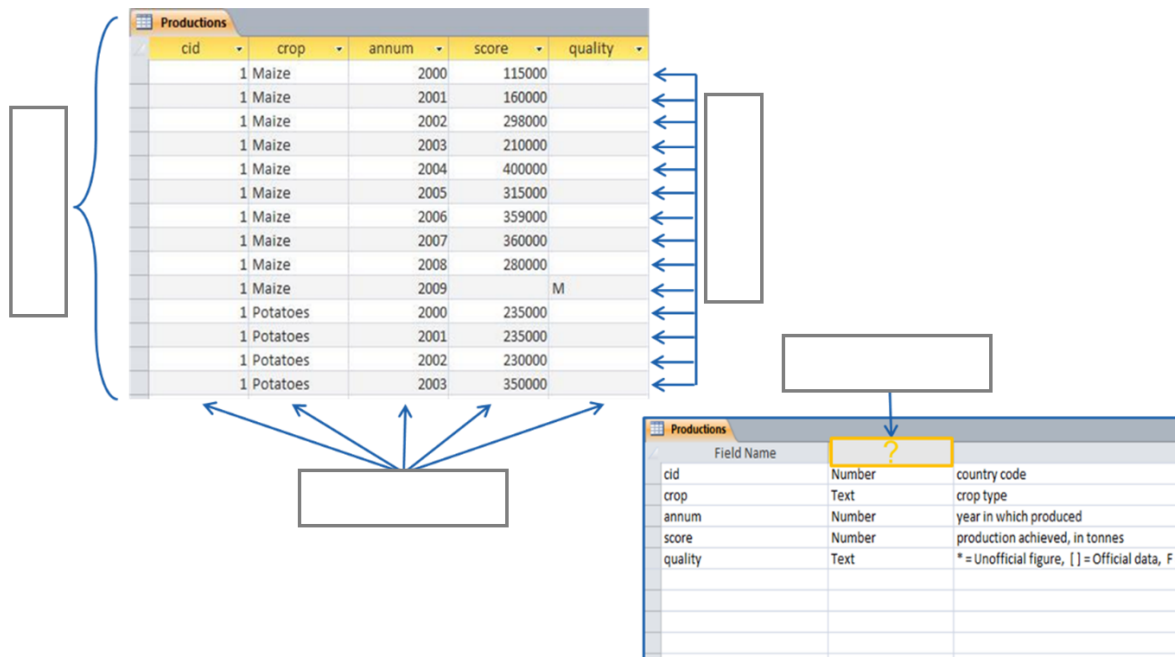


## 5.2 Relational Databases

Databases use data models for handling that data that they store. One of the most common models is the  **Relational data model**. In this section of the exercise, you will explore the main building components of a relational database:  **Relation**,  **Tuple**, and  **Attribute**. Then you will learn how a relational schema is represented in the form of a relation.

**Attention: Question.** Which are some of the most popular Spatial DBMS? Using the concept map in the  **Living Textbook**, reflect on the relationships connecting the concepts associated with Relational database. Can you elaborate a short explanation of that representation?

**Task 2** Below, you find part of the instance of a relation in a database. Fill in the blanks with the names of the components indicated by the arrows. Choose from: attribute, domain, relation, tuple.




**Task 3** Draw a relation based on the relation schema defined below. Fill it with three examples (rows) of data that match the attribute domains.

```
Courses={ID:integer, Title:varchar(50), StartDate:date, Location: varchar(50) }
```

**Attention: Question.** What other relevant attributes can be added to the course relation above? Think about a reason (requirement) to model courses in your database and modify the attributes according to your requirements.

## 5.2.1 Keys

 **Primary keys and foreign keys** enforce database integrity. By having a primary key in each relation, you make sure that there are no duplicated tuples in a relation. A **primary key** can be one attribute or a combination of attributes as far as it is unique and minimal. This means that if a unique combination of values can be achieved with two attributes, more attributes should not be used. You can read more about primary keys in your Living Textbook (page: Keys)

**Task 4** In the instance of the relation below, which attribute(s) can be possible candidates for primary keys? Which attribute(s) would you choose to set a primary key? Explain your reasons or the choice you made.

ID	name	surname	date_of_birth
42	john	smith	15/05/1977
43	ruby	jason	04/08/1987
44	sara	patil	09/12/1984
45	jason	moller	27/07/1981
46	john	alvarez	23/10/1986

Now, imagine that the table (relation) above actually contains hundreds of tuples. *Would your choice for a primary key still be a safe pick to ensure all the tuples have a primary key that it is unique?*

Relations in a relational data model are linked to each other through a common attribute. A foreign key is a set of attributes that are used to refer to tuples in another relation. A **foreign key** must correspond with the values of a primary key in another relation. A foreign key behaves like a 'logical pointer'.

**Task 5** In the figure below, you see the instances of two relations in a database. One holds data about countries; the other holds data about the production of crops productions in different countries and for various years. The countries and production relations should be related through a common attribute. Mark the primary key and the foreign key attributes and draw an arrow from the primary key pointing to the foreign key.

id	cname
1	Afghanistan
2	Albania
3	Angola
4	Argentina
5	Aruba
6	Australia
7	Bahamas

cid	crop	annum	score	quality
1	Maize	2017	115000	
1	Maize	2018	160000	
1	Maize	2019	298000	
1	Potatoes	2018	210000	
2	Potatoes	2018	40000	
2	Potatoes	2019		M
1	Maize	2016	230000	

**Task 6** Think of at least three relations (including courses) that can model a universe of discourse for managing courses in a university. Define the attributes and attribute domains for each of these relations and ensure your database integrity by specifying primary keys, foreign keys and any other given constraint.

- Hint1: Typically, you may think of Students and Courses. You can replace any of these relations and add as many more as you think is necessary for capturing the data you need.
- Hint2: Bear in mind that a student can take multiple courses and a course can have multiple students. How can you ensure the unique combination of students and courses?

A similar example is mentioned in this [website](#).

## 5.3 Data Retrieval

So far, you have got acquainted with the concepts of the relational database and its components. Now, you will interact with a database and perform simple queries to retrieve data stored in the database.

In this exercise, you will not interact with an RDBMS but with a Spatialite database which can be described as a file-based database. This is good enough to show you how a GIS organizes and queries data, without the technicalities behind a specific RDBMS. You will also understand how the data in a GIS can be interrelated by using joins. By learning this, you should not have problems if you need to interact with data in a fully-fledged RDMS.

---

**Note: Reflection.** Even though it is a good practice to store and organize your data in a database, it is not always necessary. In personal projects, you can use a file structure in your GIS or a file-based database instead of an external database. A database is not necessary when you do not have multiple users, and the amount of data is not very large or the queries you will define are not complex. But in large projects, in which various users will work with hundreds or thousands of files and perform advanced queries; you should design and implement a database before starting analyzing and visualizing your data in a GIS.






---

Having the necessary datasets is the starting point to be able to extract relevant information. Querying data is an important skill of a GIS analyst. To accomplish this, we use SQL – Structured Query Language.

---

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset [database\\_queries.zip](#). When you unzip the dataset, you will find the following files inside:

- `Database_queries.qgs` – a QGIS project preloaded with the dataset described below;
  - `country_Data` – a table with additional information on the countries of the world;
  - `database.sqlite` – a SpatiaLite database containing:
    - `countries` – polygons of the boundaries of the countries of the world;
    - `railroads` – lines representing the main railroads of the world;
    - `urban_areas` – polygons of the boundaries of the main urban areas of the world
- 

**Task 7** Read about  [querying a spatial database with SQL](#). You should understand what  [Tuple selection](#) means; what is  [Attribute projection](#); and how can we  [join](#) two or more than two relations. Such understanding is necessary for solving query-formulation problems, using a  [Join condition](#).

### 5.3.1 Selection by Attributes

The first approach we will discuss for querying data is tuple selection based on attributes, (i.e. ignoring spatial relationships). Tuple selection works like a filter; it retrieves tuples that meet a particular condition.

A tuple selection written in SQL has the following structure:

```

1 SELECT *                /* attributes to be projected */
2 FROM Parcel             /* relation(s) to be queried */
3 WHERE area_m2 > 1000    /* condition(s) */

```

SQL has a very clear structure. Line 1 (**select**) defines the attributes to be projected (shown in the result of the query), Line 2 (**from**) defines the relations that will be used. Line 3 (**where**) contains the condition(s) that will be applied to filter the data.

In SQL, you can use comparison operators (>, <, =, !=, >=, <=) and logical operators (AND, OR, Like, NOT) to define conditions for filtering data.

When you query data in a GIS (and not a DBMS), the way you express your attribute selection may be slightly different, but it usually follows the SQL syntax. The following task shows you how to query a data layer.

**Task 8** Open the QGIS project *'database\_queries'*. Right-click on the layer *countries* > *Filter..*, define the simple query below, and click OK. See Fig. 5.2

`"sovereignty" = 'Russia'`

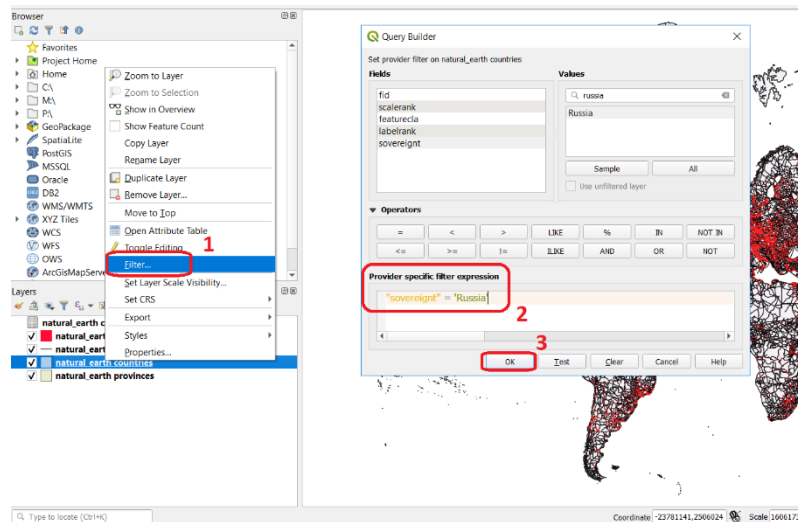


Fig. 5.2: Filtering data using a simple query in the query builder

You should see now that your dataset only shows the countries that belong to Russia.

### 5.3.2 Joins

Often the information you need to retrieve is stored in two different relations. When this happens, you have to use joins. The pre-requisite to joining relations is to declare a join condition. The condition has to be met by both relations in the same manner. This is usually a common attribute that relates tuples in both relations.

In SQL, a join can be defined by structuring statements in the following way:

```
1 SELECT *
2 FROM TitleDeed, Parcel
3 WHERE TitleDeed.Plot = Parcel.ID
```

**Attention: Question.** In the SQL statement above, you immediately notice that it is using two relations instead of one. Can you tell which common attributes are being used in this example?

**Task 9** Examine the attribute tables of *countries.shp* and *countries\_data.shp*. You will certainly note that the data contained in the attribute table of *'countries\_data'* data layer complements the data provided by *'countries'* data layer.

**Task 10** Write a joining condition for the datasets above. The following SQL statement will help get you started.

```

1 SELECT *
2 FROM countries, countries_Data
3 WHERE /* Here goes your condition */

```

QGIS (and other GIS packages) provides a graphical interface from where you can define a join without having to type the SQL statements.

**Task 11** Use QGIS to apply the join defined in the *previous task*. Fig. 5.3 shows how it is done.

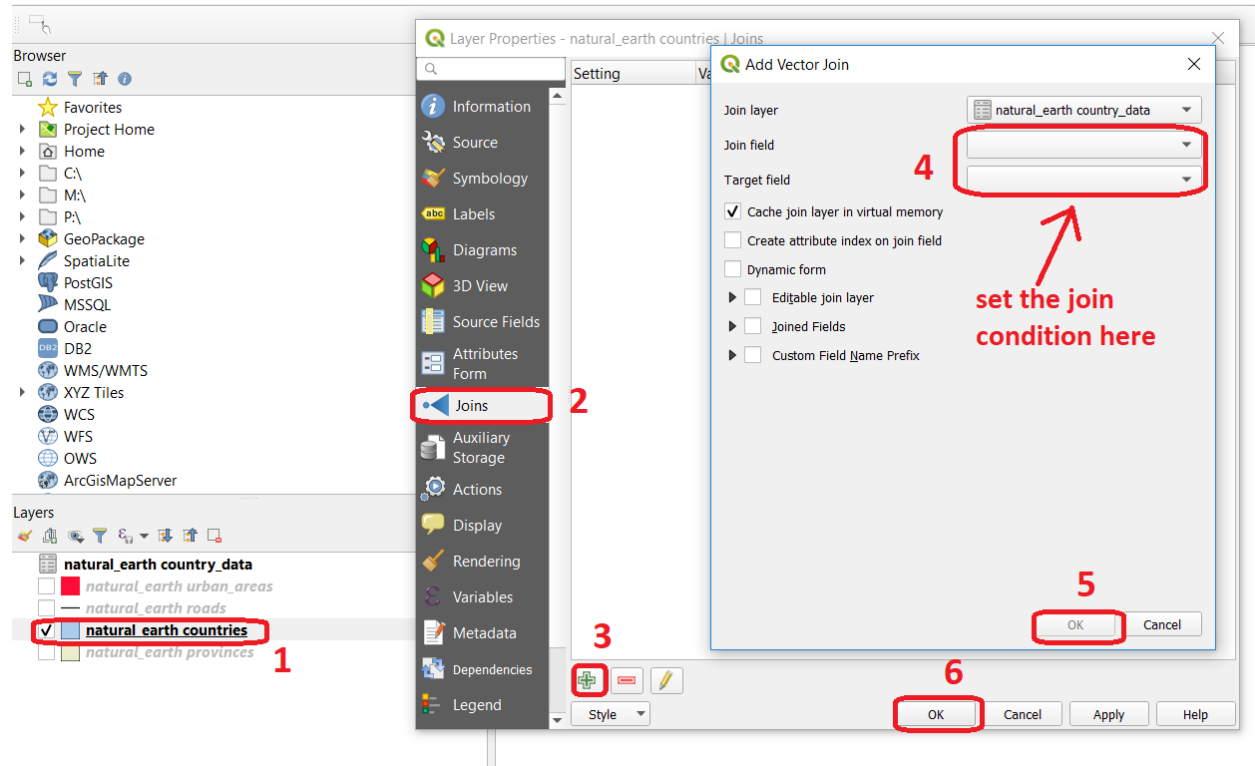
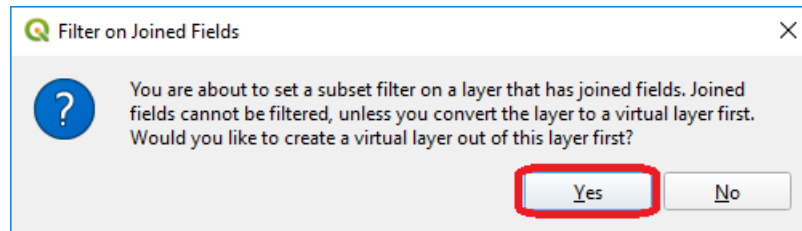


Fig. 5.3: The definition of a join in QGIS

**Note: QGIS.** Joins only exist in the scope of a project. Notice that the countries dataset is now richer in attributes as long as the table 'countries\_data' is in the same project, and as long as the join condition remains active. If you open the countries in a different QGIS project, you will see that the attribute table does not include the attributes from the 'countries' table. To make the results of a join permanent, you have to create a dataset by exporting the joined dataset to a new file. This procedure is common to any GIS software. Also, be aware that filtering will not take into account the joined attributes, unless you the results to a new dataset, or create a virtual field (check the next task).

Now that the attribute table of the 'countries' layer is extended, we can revisit **Task 8** and make more interesting queries.

**Task 12** Right-click on the layer countries > Filter... The message depicted in below will pop up. Click Yes.



Once you are in the filter menu, formulate queries to answer the following questions. Remember to click the *Clear* button before proceeding to the next query problem.

**Attention: Question.**

1. How many countries start with the letter 'E'?
2. List all the countries belonging Africa and Europe that have more than 100.000.000 inhabitants.
3. How many countries whose official name includes 'Republic' are classified as '6. Developing region'?

### 5.3.3 Selection and Joins by Spatial Relationship

GIS handles spatial data, and this means that:

*"Everything is related to everything else, but near things are more related than distant things."*

The implications of Tobler's first law of Geography are what allow us to search for information based on spatial relationships while ignoring the semantics associated with the features. The key idea to remember is that all the objects represented in the same coordinate space have at least one thing in common - space itself! This means that we can filter our data based on several Topological relationships like disjoint, meets, equal to, inside, covers etc.

**Task 13** suppose we want to find out how many urban areas are connected to railroads. Use the *Select by location* tool from the **Processing toolbox** to find out. Figure [Fig. 5.4](#)

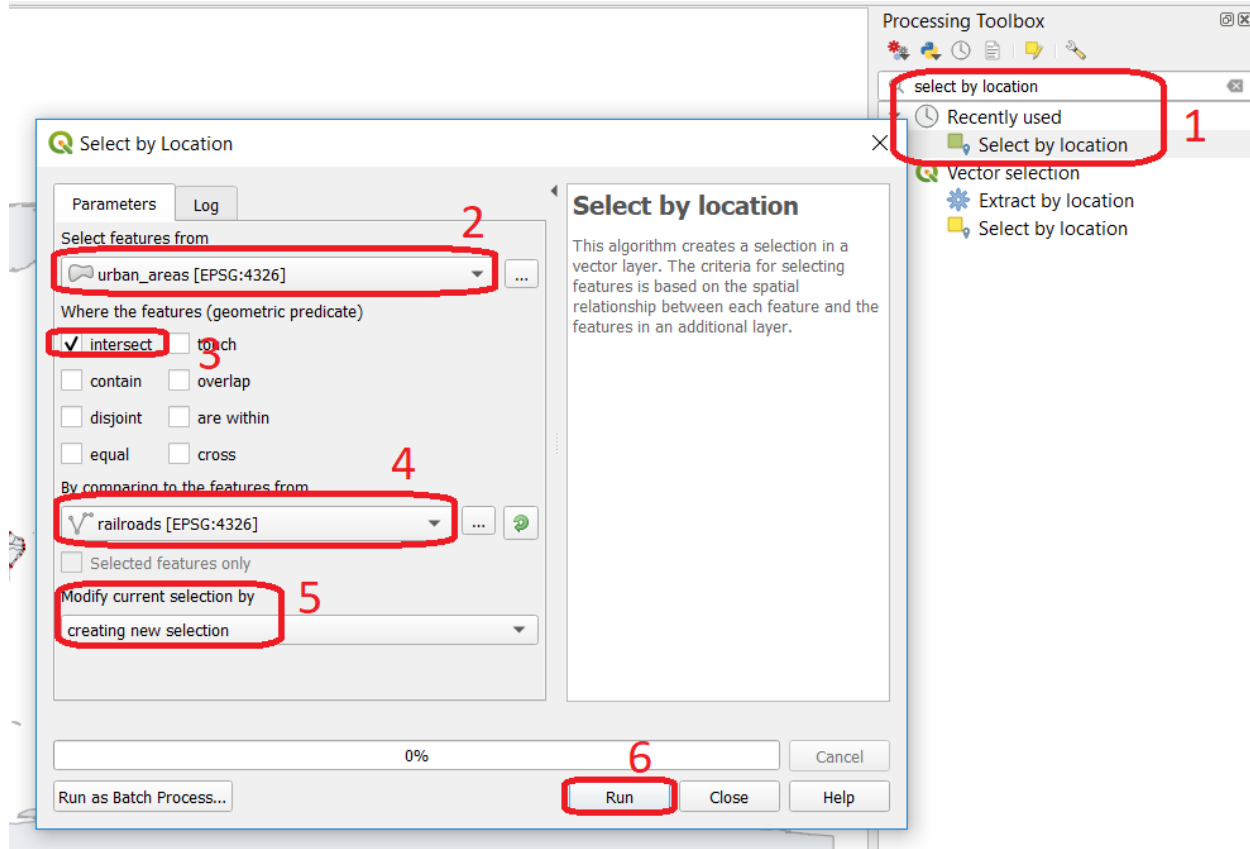


Fig. 5.4: The 'select by location' tool in the Toolbox

If all went well, you should see some features in the 'urban\_areas' layer highlighted with yellow. This means that such features intersect a railroad.

**Note: Reflection.** What your GIS did was to perform a join on the fly based on a spatial predicate – Intersects. The difference is that attributes were not joined; instead, features were selected, but the principle is the same.

In SQL terms, what happened was something like this:

```

1 SELECT u.*                                /* project only the attributes of u */
   ↪ relation u */
2 FROM urban_areas as u, railroads as r      /* relation aliases */
3 WHERE st_intersects(u.geom, r.geom)        /* spatial relation: intersection */

```

### 5.3.4 Using Spatial SQL

If our data is in a spatial database, we can access all sort of spatial functions using SQL, and obtain answers to simple or complicated questions in a straightforward manner.

To explore spatial functions using SQL, we will use the capabilities of *Spatialite*, a file-based database engine. This means you don't need to install any database software. All data is contained in a single file that can be copied from one computer to another without losing information.

**Task 14** Connect to the Spatialite database.

If your `database.sqlite` is not listed under the **Spatialite** section, you have to connect to it first. Simply *right-click* over *Spatialite* > *New Connection...*, then point to the folder where you have the `database.sqlite` file. Fig. 5.5

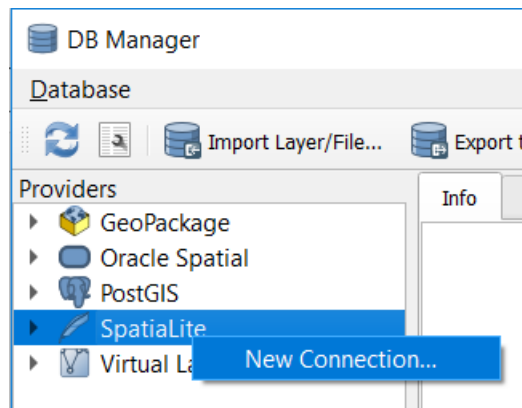


Fig. 5.5: Connecting to a Spatialite database

Then, go to *DB Manager* > *Spatialite* > *database.sqlite* to get a list of the relations inside the `database.sqlite`. Note that layers in the database are already loaded into your project, *do not load them again*.



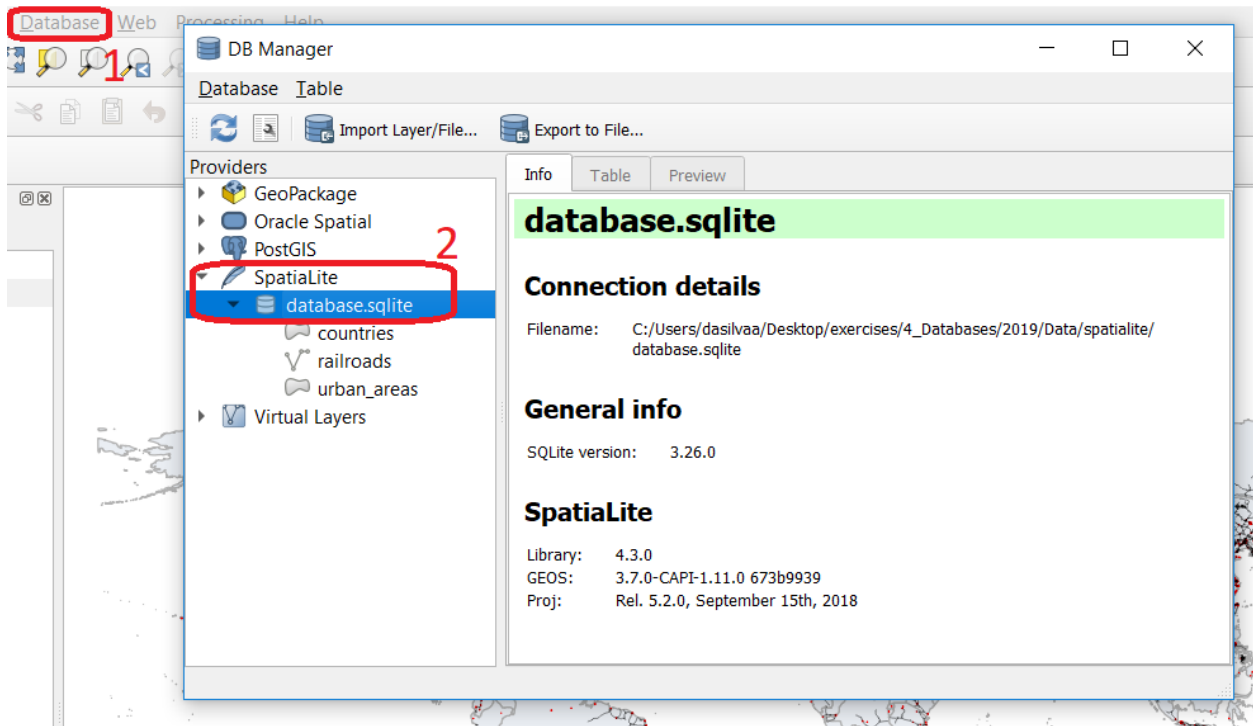


Fig. 5.6: Browsing a Spatialite database in the DB Manager

Now open *Query* window, copy the SQL code below, click *Execute* Fig. 5.7

```

1 SELECT sovereign AS country, st_area(geom) AS area
2 FROM countries
3 ORDER BY area DESC
4 LIMIT 10

```

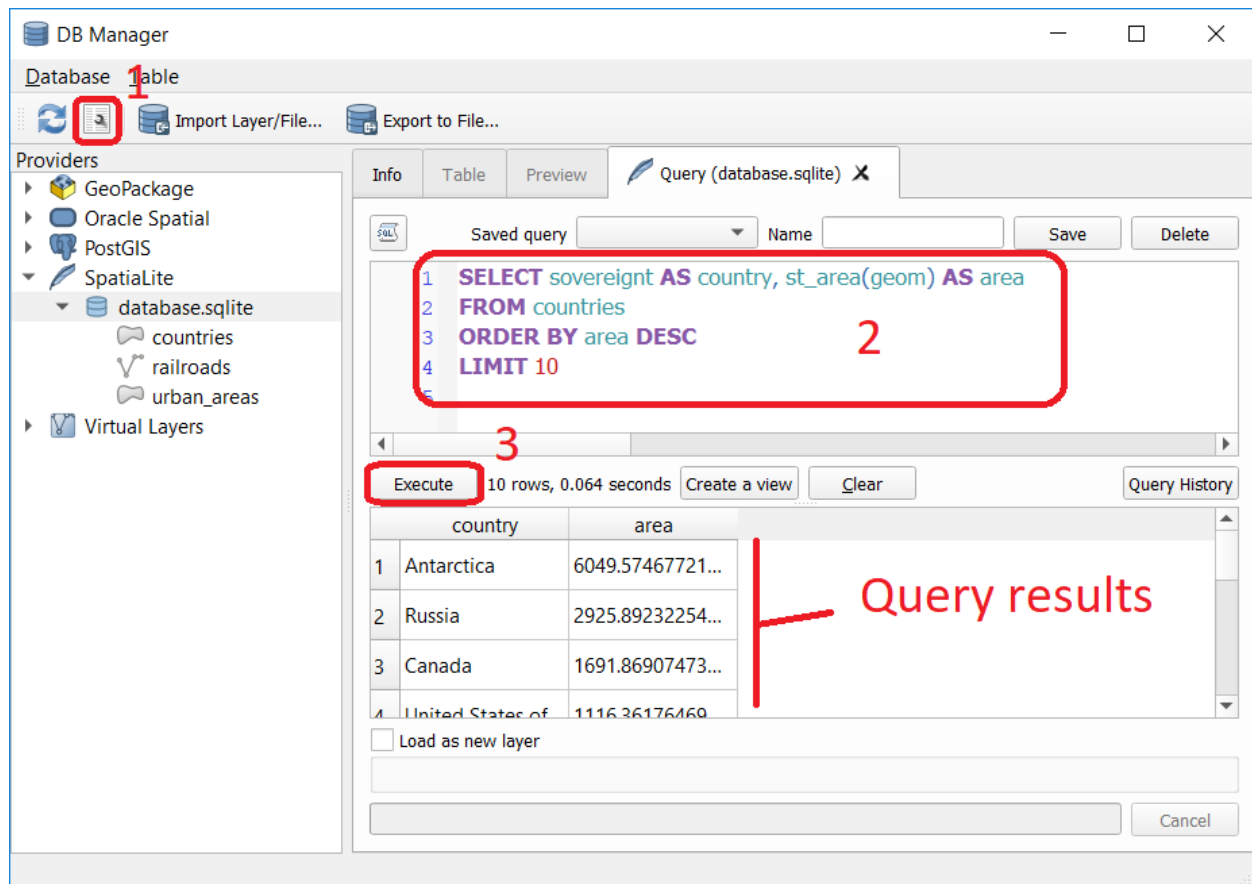


Fig. 5.7: Executing a query in the DB Manager

**Attention: Question.** Analyze the SQL code and try to understand how it works. Although syntactically correct, the results produced by the query above cannot be trusted. Can you explain why?

**Task 15** To finish this section, we will give you another demonstration of the capabilities of a spatial database. Suppose you want to know what is the total area next to the railroads of Australia given a certain distance. The distance could represent a buffer zone restricted to the public for security reasons.

Here is a way to do it using SQL:

```
1 SELECT r.id, st_area(st_buffer(r.geom, 0.1)) AS area
2 FROM countries AS c, railroads AS r
3 WHERE c.sovereign = 'Australia' AND st_intersects (r.geom, c.geom)
```

**Attention: Question.** This is a variation of the previous SQL query. Can you tell what is the difference and why the query below is better?

```
1 SELECT r.id, st_area(st_union((st_buffer(st_transform(r.geom, 3577), 50)))) /
   ↳ 1000000
   AS area_km2
2 FROM countries AS c, railroads AS r
3 WHERE c.sovereign = 'Australia' AND st_intersects (r.geom, c.geom)
```

---

**Important:** This exercise is not meant to provide training on SQL, and we do not expect you to become proficient in it. The main point is that spatial databases are powerful tools and that if you want to take advantage of it, you will have to learn a bit of SQL. If you never had contact with SQL, it might seem intimidating at first, but it is not very hard to learn the basics. A good resource for learning is <https://www.w3schools.com/sql>

---

*Section author: Andre da Silva Mano, Parya Pasha & Manuel Garcia Alvarez*




## VECTOR ANALYSIS

### 6.1 Geometric Measures

The term “geometric measurements” describes measurements on the spatial features themselves, and not on their attribute values. Geometric Measurements include:

- location.
- distance (distance between two features)
- length (length of a line segment or boundary of a polygon)
- area (of polygon features)

In this exercise, we will discuss all these measurements. A general overview of  [measurement operations](#) on vector data is provided in your Living Textbook.

---

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset [vector-analysis.zip](#). When you unzip the dataset, you will find the following files inside:

- `Vector_analysis.qgs` – a QGIS project preloaded with the datasets below;
    - `Centroids.gpkg`
    - `DistancePoints.gpkg`
    - `Linebuf.gpkg`
    - `Thiessenpoints.gpkg`
    - `Overlay1.gpkg`
    - `Overlay2.gpkg`
- 

#### 6.1.1 Location

The GIS always stores the location of the vector features. For point features the  $x$  and  $y$  coordinates are stored. For lines, the start node, end node and internal vertices are stored, and sometimes the length of the line segment. For polygons, it stores the line segments that define the boundaries, including the perimeter, and the area of the polygon. However, sometimes we also store the centroid of line or polygon features.

**Task 1** In your dataset, you find a Shapefile called ‘*Centroids*’. Compute the centroids for the features in that layer in QGIS. Then, check if the centroids are inside or outside the original polygons. [Fig. 6.1](#)

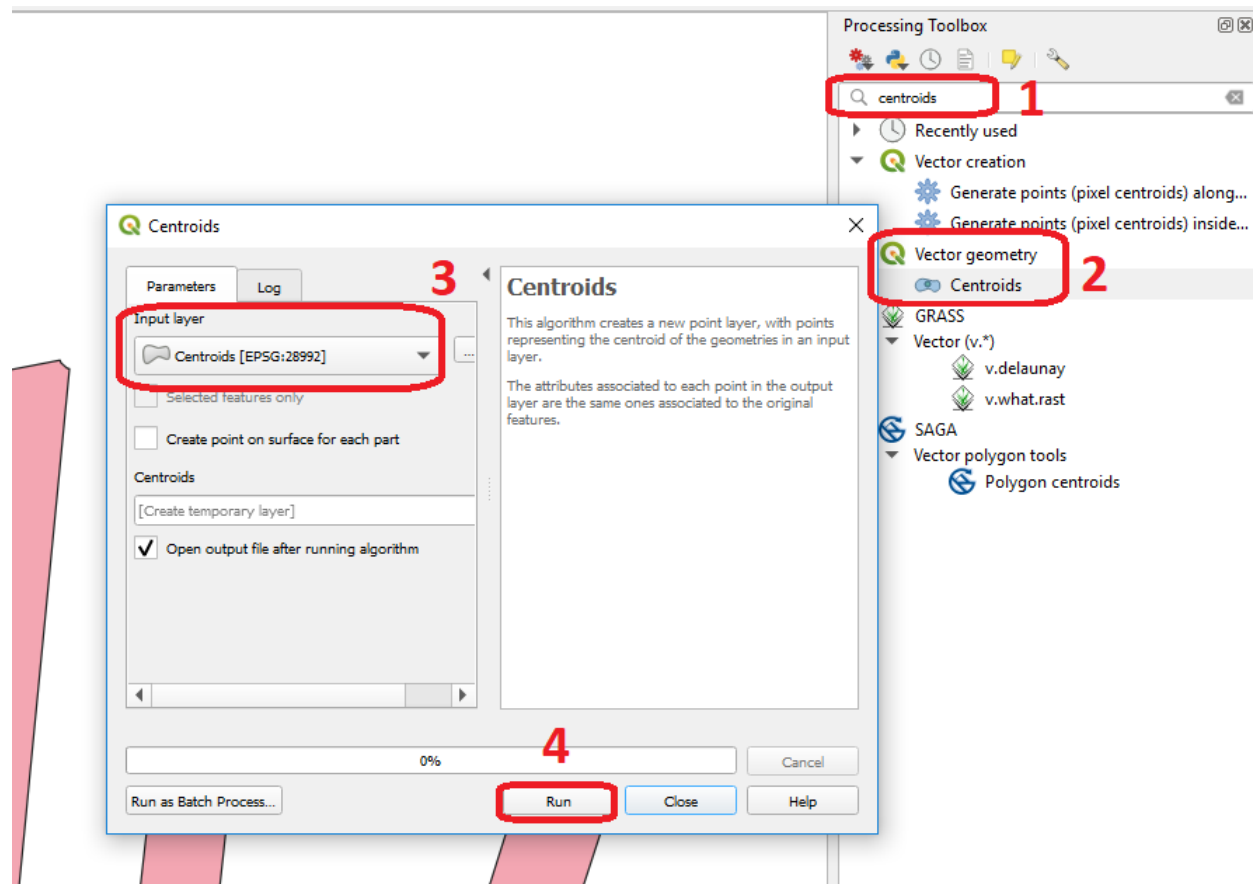


Fig. 6.1: Computing centroids in QGIS

**Attention: Question.** Can you give an example of a situation when computing the centroid is useful?

### 6.1.2 Distance

Another geometric measurement is distance. Computing the distance between two points in a straight line is a basic operation that you can solve using basic math.

**Task 2** Open QGIS and use the **Add Geometry attributes** tool to find the exact coordinates of the poings in the '*DistancePoints*' layer. *Processing toolbox > Vector geometry > Add geometry attributes*. The *x* and *y* coordinates will be added to the attribute table.

**Task 3** Using the *x, y* coordinates from the previous task, calculate **manually** the distance between the two points in meters. See Fig. 6.2

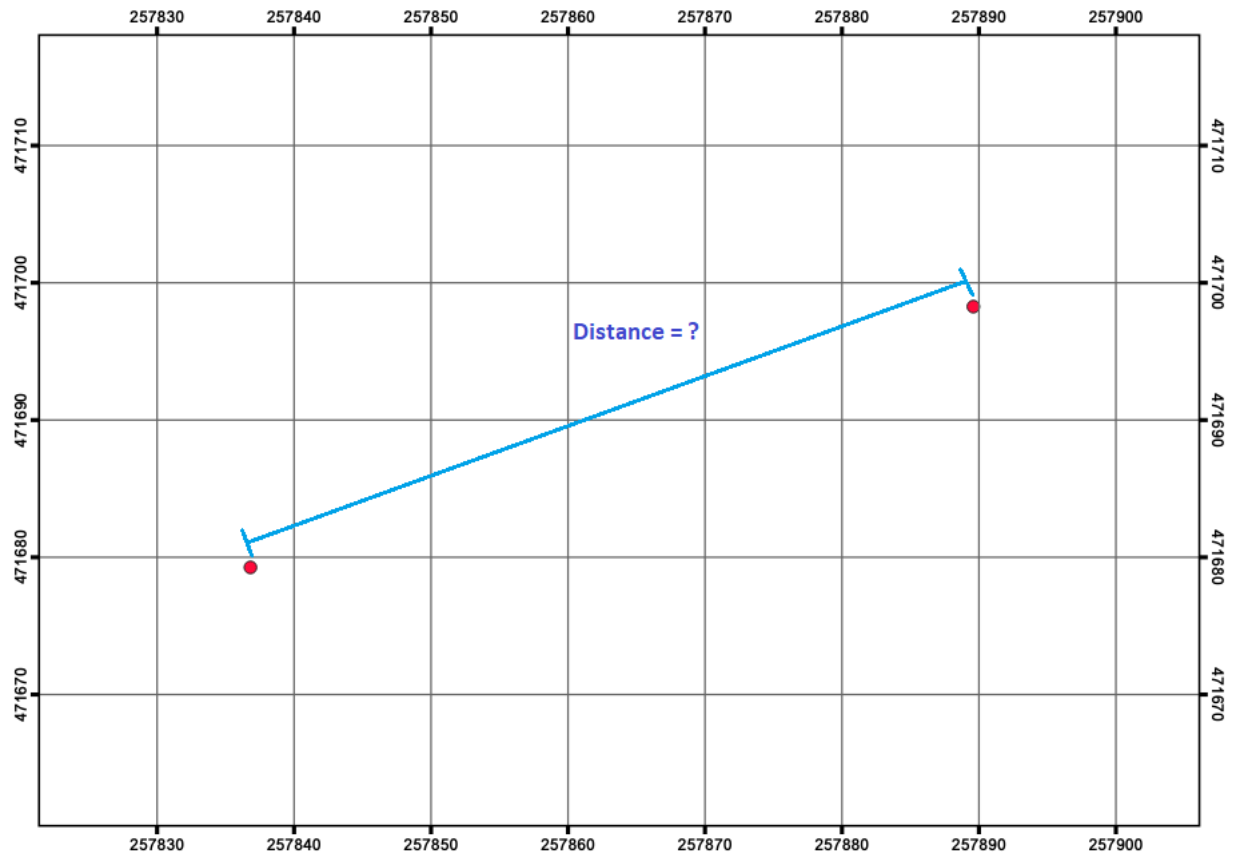



Fig. 6.2: Straight distance between points in the 'DistancePoints' layer

**Task 4** Using the **Measure Line** tool , measure the distance between the points in the 'DistancePoints' layer. Fig. 6.3

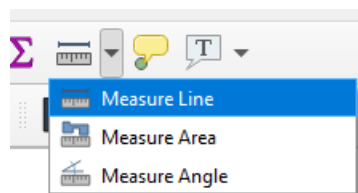


Fig. 6.3: Using the Measure Line Tool

#### Attention: Question.

- Measuring between two points is simple, especially when you use a measurement tool and draw the line you want to measure. However, in GIS software, some tools measure the distance from all features in one layer to the nearest feature in another layer. But, what would be the distance between a point and a line, or between a line and a polygon?
- The minimum distance between the features?
- The distance between the centroids of the features? or


- The distance between the two closest vertices?

Another type of geometric measurement discussed is the *minimal bounding box* of a feature.

**Task 5** Use the **Bounding boxes** tool from the **Processing Toolbox** to visualise the minimal bounding boxes of the features of the ‘*overlay2*’ layer.

---

## 6.2 Overlays

 **Vector Overlay** operations combine two input layers (be it a point, line or polygon layers) into a new data layer. Vector overlay operations apply combinations of the following:

- Intersection of the geometry
- Spatial join of the attribute tables
- Definition of the output map extent

Some overlay operators perform both an intersection of the geometry and a spatial join of the attribute tables in combination with deriving a certain output extent. Still, others only join attribute tables or perform spatial intersections.

**Task 6** Using the three polygons overlay operators discussed in the Living Textbook complete the table below.

Overlay Operator <sup>1</sup>	Intersection of the geometry? (yes/no/partly)	Join attribute tables? (yes/no)	Output extent? (AND/OR)

**Task 7** Find the **Union**, **Intersect** and **Clip** tools in the **Processing Toolbox**, and use them to compute the overlay operations using the ‘*overlay1*’ and ‘*overlay2*’ layers as inputs. Compare the result with the table above.

“*The fundamental operator of all these vector operations is **polygon intersection**. All other operators can be defined in terms of polygon intersection, usually in combination with polygon selection and/or classification*”. Below, you see the result of an overlay operation called: **Symmetrical Difference** between the ‘*overlay1*’ and ‘*overlay2*’ data layers.  
[Fig. 6.4](#)

---

<sup>1</sup> There are many other vector operators besides the operators discussed in the Living Textbook.





Fig. 6.4: Symmetrical difference between 'overlay1' and 'overlay2'

**Attention: Question.** How would you achieve the same results generated by the symmetrical difference tool, using only the *intersect* tool and *selection operators*?

## 6.3 Proximity Operators

We will cover two proximity operations:  **Buffer** and  **Thiessen Polygons**.

You create a buffer using point, line and polygon layers as inputs. Buffers can be created for all the features in a layer or for only a few selected features. We can use a **fixed buffer distance**; in which case, a buffer of the same size will be created for all the features in a data layer. However, we can also use a **variable buffer distance** for each feature; in which case such the buffer distances need to be stored in the attribute table of the layer.

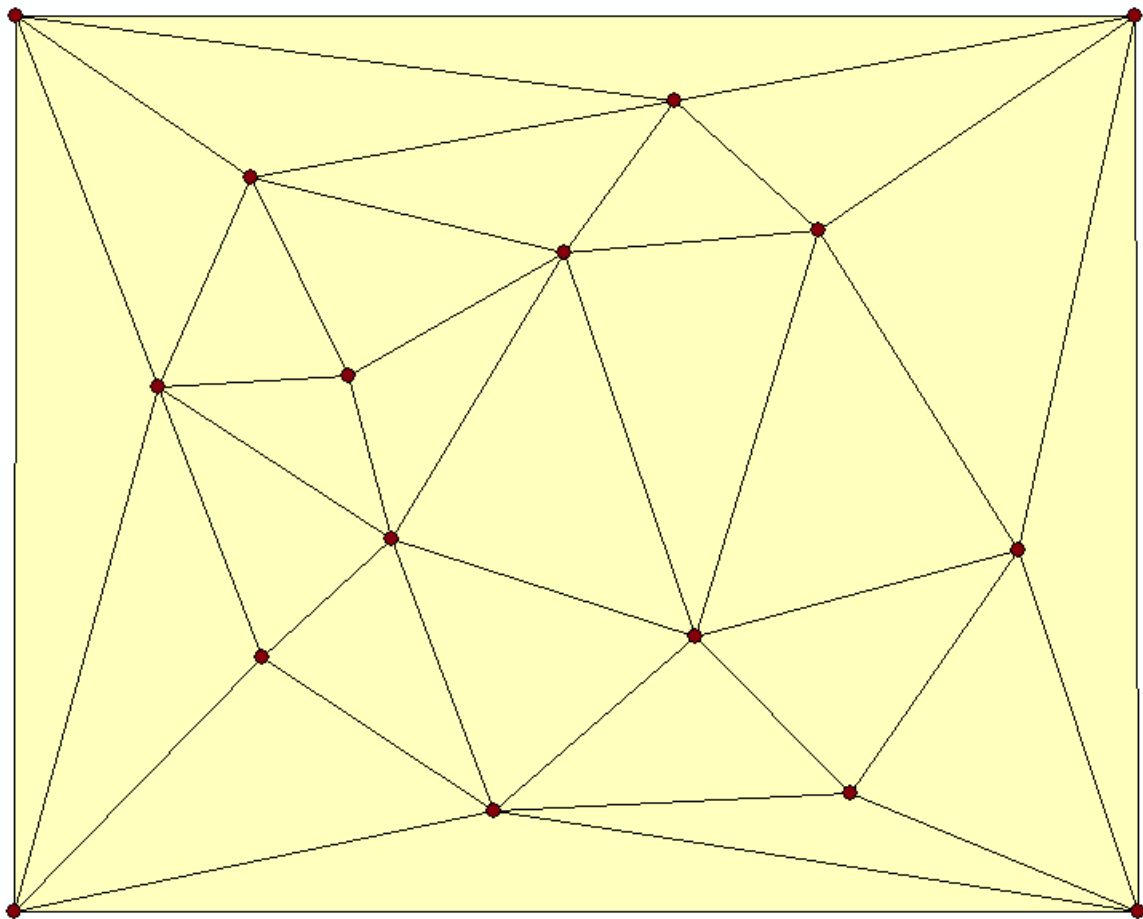
**Task 8** Check the attribute table of the 'linebuf' layer. You will find an attribute called **Bufdist**. Use this attribute to generate buffers with different buffer distances. Go to *Processing Toolbox > Variable distance buffer*.

Then, create a zoned buffer for the 'linebuf' layer using a fix buffer distance. *Processing Toolbox > Multiring buffer (constant distance)*.

**Attention: Question.** One could argue that the problem with buffers is that they are discrete. Can you explain what that means and give an example in which that is a problem?

Another example of proximity operators is Thiessen Polygons. If you are familiar with the concept of *Voronoi Map*, Thiessen polygons are the same. They identify the areas that are closest (in *Euclidean distance*) to each point in a dataset.

**Task 9** Below you see some points and a corresponding TIN (triangulated irregular network). Select 2 or 3 points and draw their corresponding Thiessen polygon.



**Task 10** In the **Processing toolbox** search for a way to generate Thiessen polygons in QGIS. Remember that Thiessen polygons are also called Voronoi Maps and to find the correct tool in QGIS you might search for this term.

---

**Note: Reflection.** This website compares Thiessen Polygons with features in nature like the pattern on a giraffe: <http://forum.woodenboat.com/showthread.php?112363-Voronoi-Diagrams-in-Nature>



**What do Thiessen polygons remind you of?**

---

## 6.4 Networks

Before moving onto network analysis, we have to understand networks a bit better. This means understanding a network's characteristics and data model.

### 6.4.1 Characteristics of Networks


There are two critical aspects in a  **Network**; the directionality of the network and the degree in which the network is planar. When you understand these two concepts you know why different types of networks are modelled in a different way and why not all  **analysis techniques** are relevant for all types of networks.

**Task 11** Complete the table below to create an overview of the different types of networks.

Example	Planar or Non-planar	Directed or Undirected	Type of analysis <sup>2</sup>
River Network			
Road Network			
Electricity Network			
Sewage Network			

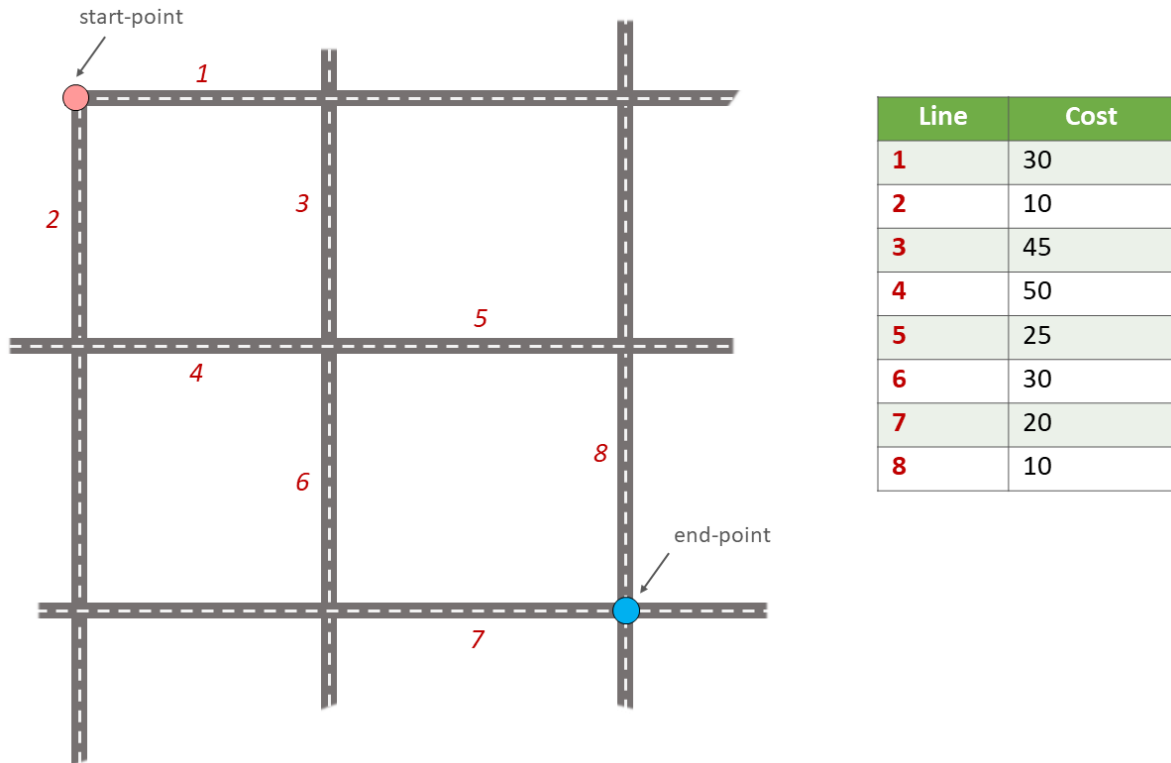
### 6.4.2 The Network Data Model and Analysis

Networks consist of points (nodes) and lines (edges or segments). What is very important for a network is connectivity. Therefore, the smallest gap between the edges stops the flow over the network. We use line topology to ensure that we end with a network with connected points and lines.

In data modelling, we already learned that a line has a '**start node**' and an '**end node**'. Because of this, the network segments have direction. When discussing the directionality in a network, we call the start and end nodes '**from node**' and '**to node**', respectively. In network analysis, we use a **cost function** to represent '*impedance*'; i.e. *a function that determines the cost of moving from one node to another in the network*. Cost functions are stored as an attribute indicating the cost to travel each edge in the network.  **Optimal Path Finding** is an example of network analysis that uses cost functions.

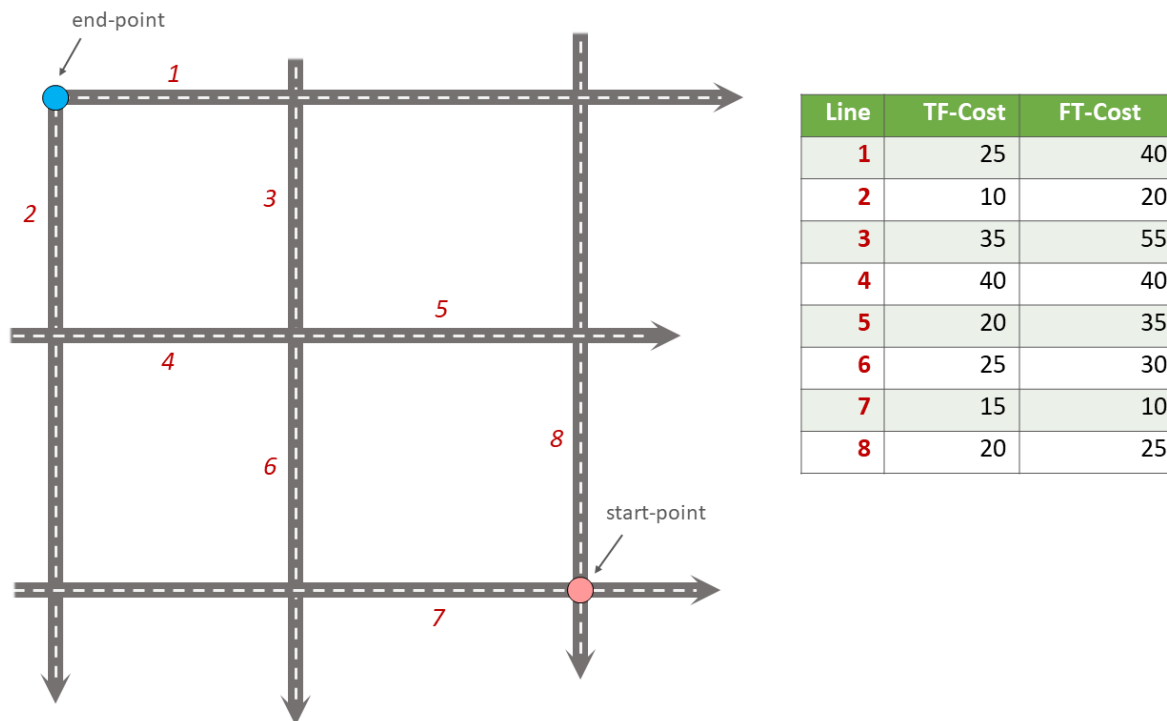
**Task 12** Determine the optimal path of a network. Below you see a road network (left) with the IDs for each line segment. On the left side, you see an (attribute) table with the cost associated with each line segments. **What is the least cost path from the start-point to the end-point?**

<sup>2</sup> Choose from 'optimal pathfinding', 'network allocation', or 'tracing'.



In the previous task, there was only one cost function, and it was applied in any direction. There are many reasons why the cost might be different for different directions—for example, different speed limits, different number of lanes, or less traffic.

**Task 13** Determine the optimal path of the **directed network** below. This time consider two cost functions; a ‘to-from’ cost (TF-Cost) when moving on the direction of the arrows, and a ‘from-to’ cost (FT-Cost) when moving in the opposite direction. *Re-evaluate the route, this time the start and end points are different.* **What is the least cost path from the start-point to the end-point? Is it the same as the previous one?**



**Attention: Question.** The cost function can be associated with lines (as in the previous tasks), and nodes of a network. When would you apply cost on the nodes?

More advance topics on network analysis are [Network Partitioning](#), [Network Allocation](#) and [Trace Analysis](#). Network partitioning is a group of analytical functions that assigns part of a network to predefined target locations. In network allocation parts of a network are assigned to specific locations defined as service areas. In trace analysis, part of the network is also assigned to particular locations, but its use is restricted to directed networks.

**Attention: Question.** In your own words, what are the differences and similarities between Thiessen polygons and Network allocation?

**Task 14** Fig. 6.5 shows you see the results of applying two vector analyses:

1. The result of a zonated (multiring) buffer around a point (yellow dot). Each ring is separated by a distance of 500 m.
2. The result of applying network allocation around the same point as in 1. Each coloured section of the road network is separated by also 500 m.

**Describe the difference between the two analysis and the reasons behind these differences.**

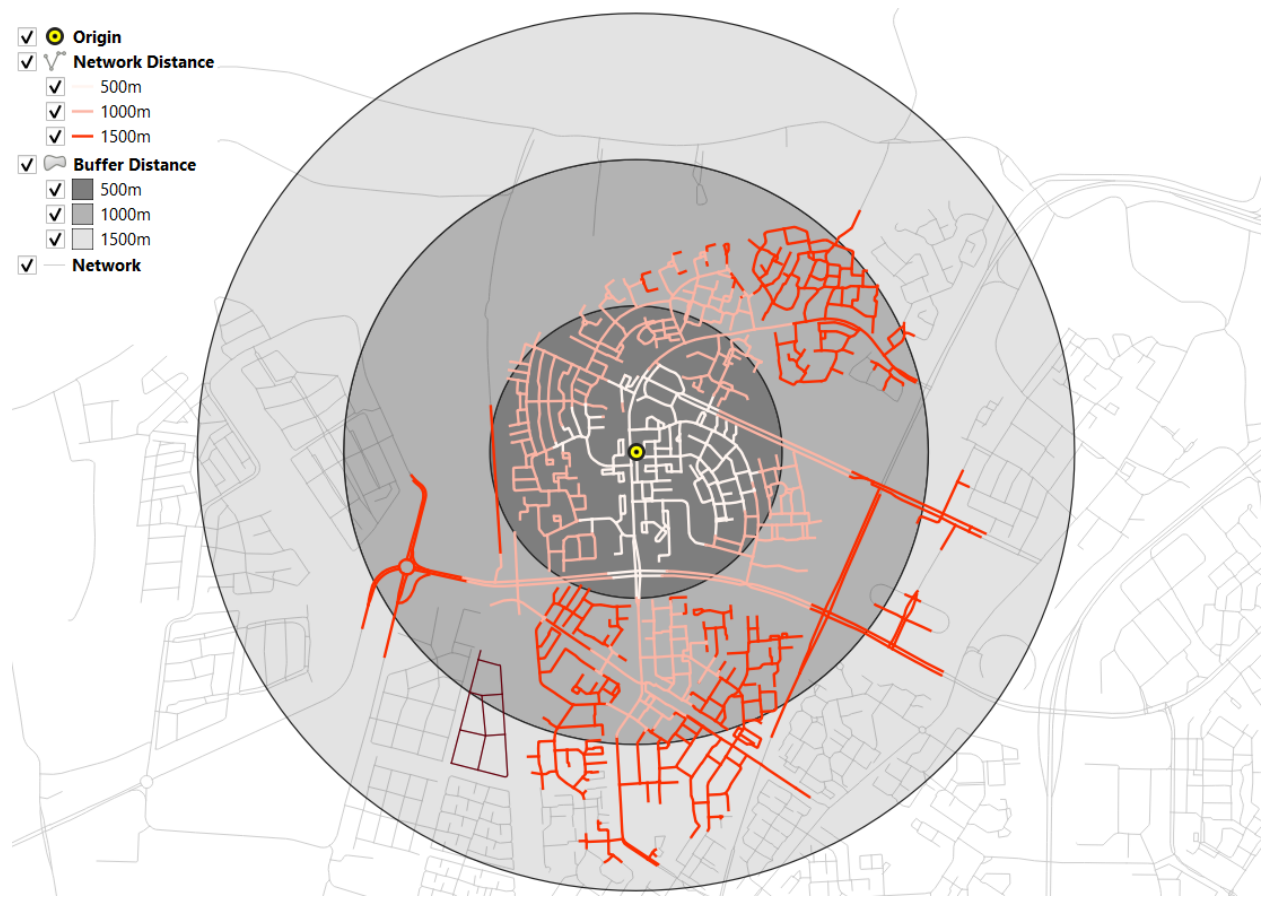








Fig. 6.5: Zonated buffer and network allocation around a point

**Attention: Question.**

- On which types of networks can we apply trace analysis?
- Which are the characteristics that a network must have to apply trace analysis?

Section author: Ellen-Wien Augustijn, Andre da Silva Mano, Manuel Garcia Alvarez

## NETWORK ANALYSIS

This exercise will guide you through examples of  network analysis, specifically  optimal path analysis,  network allocation, and  Trace analysis. Before proceeding with the exercise, it is advisable to read the Living Textbook entries on what is a  network and understand the role  cost functions play when performing any kind of network analysis.

---

**Important: Resources.** You will require the latest LTR version of LTR version of QGIS, plus the dataset `network_analysis.zip`. In it, you will need the following files:

- `Network_analysis.qgs` – a QGIS project preloaded and pre styled with Open Street Map datasets[1] from the central district of Jebel Awlia, in Khartoum, capital of the Republic of Sudan. These include:
  - `school.gpkg` – point features representing schools;
  - `post_box.gpkg` – point features representing post mail boxes;
  - `hospital.gpkg` – point features representing hospitals;
  - `fire_station.gpkg` – point features representing firefighters headquarters;
  - `bus_stop.gpkg` – point features representing bus stops;
  - `building.gpkg` – polygon features representing buildings;
  - `road_network.gpkg` – line features forming a topologically consistent network of roads;
  - `demand_points.gpkg` – point features used for the challenge (optional) exercise.
- `Trace_analysis.qgs` – a QGIS project preloaded and pre styled with the following datasets depicting the county of Armagh, Northern Ireland:
  - `armagh_rivers.gpkg` - line rivers representing the main rivers<sup>2</sup>;
  - `dem.tif` - a digital elevation model extracted from SRTM data<sup>3</sup>;
  - `hillshade.tif` - a hillshade raster used in conjunction with the dem to enhance topographical features.

---

<sup>2</sup> <https://data.gov.uk/dataset/c734ecd3-7603-4397-8da9-57e79e398599/northern-ireland-river-segments>

<sup>3</sup> <https://www2.jpl.nasa.gov/srtm/>

## 7.1 Optimal path analysis

Optimal path, also known as ‘shortest path’ analysis, seeks to find the most efficient route over a network given an origin and a destination. Within this general definition, there are some variants to this problem but at its simplest form the Optimal path analysis operates over a starting point and an ending point.

**Task 1** From the **Processing Toolbox**, filter the tools by **Shortest Path (Point to Point)** and open the tool. Make sure the *path type* to calculate is set to *Shortest* and then select a *Start point* somewhere in the bottom southeast corner of the network and an *End point* somewhere in the northwest corner. Hit the *run* button but **DO NOT** dismiss the tool dialog Fig. 7.1.

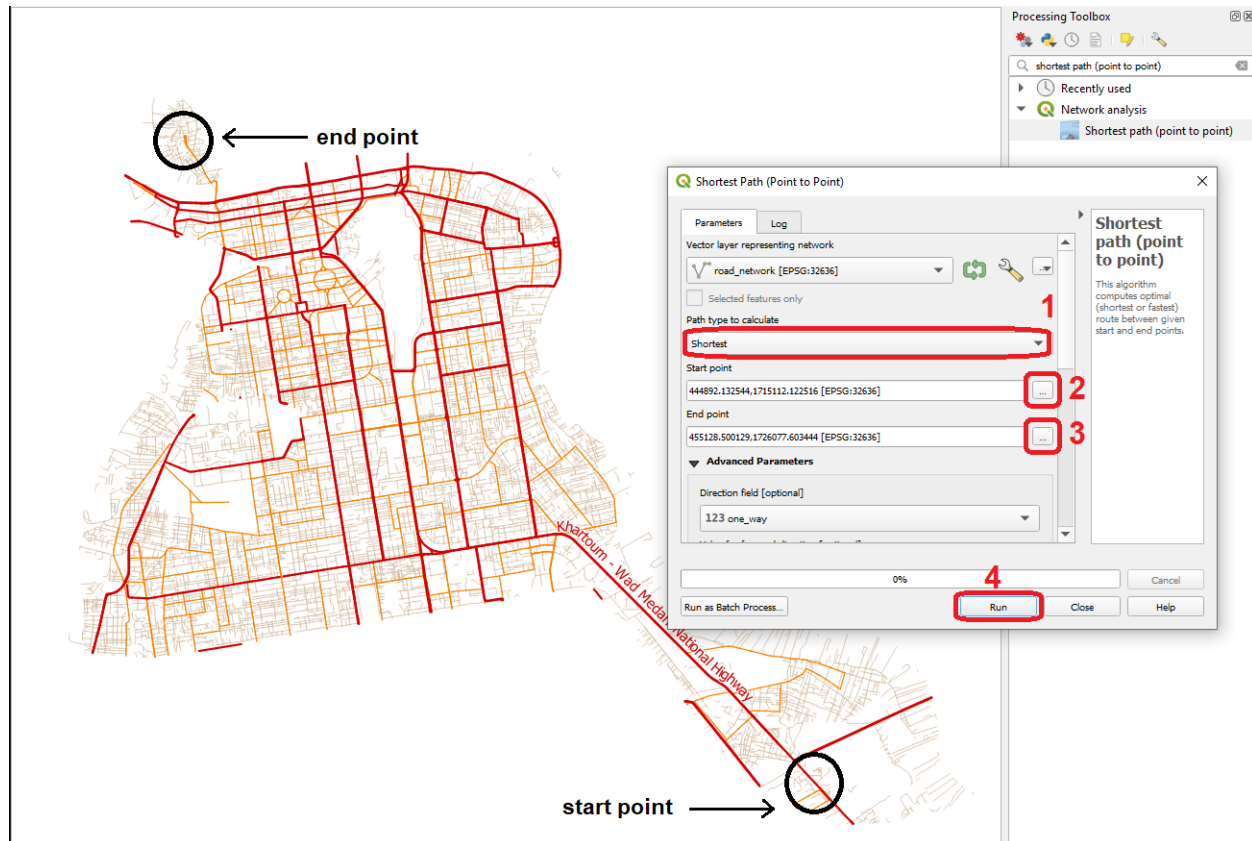


Fig. 7.1: Shortest path (Point to point) using shortest distance

**Task 2 (optional)** You may want to rename the output to something like *shortest\_path\_simple*. Also, consider changing the style (colour and/or width) This may become handy if you want to compare some of the outputs you obtain along this exercise.

**Task 3** Repeat the analysis described in Task 1, but this time using *Fastest* as the *Path type*, and making sure that under the *Advanced settings*, you indicate the field *max\_speed* as the attribute to be used to calculate travel times. Just like before, **DO NOT** dismiss the tool dialog Fig. 7.2.



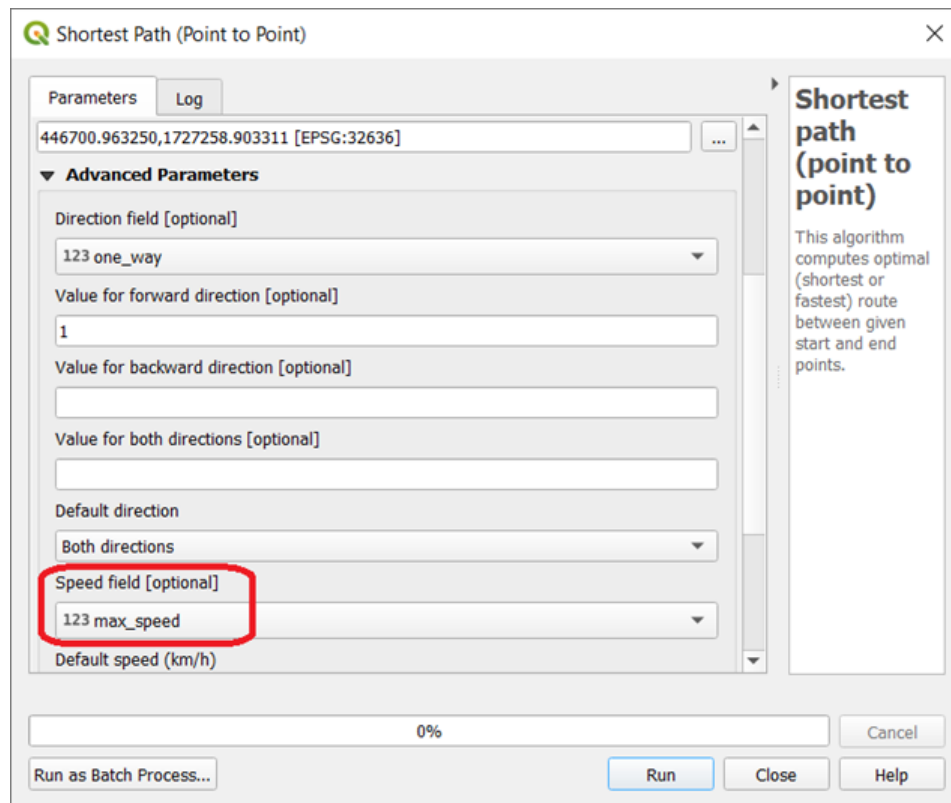


Fig. 7.2: The speed field option

**Task 4 (optional)** Consider to change the name and style of the output. For name, we suggest *fastest\_path\_simple*.

**Attention: Question 1** You should now have two different routes showing the most efficient path Fig. 7.3. You probably realized the cost functions used are different and therefore the results are not quite the same. On what are these cost functions based on? [hint: check the values of the 'cost' attribute of the two outputs]

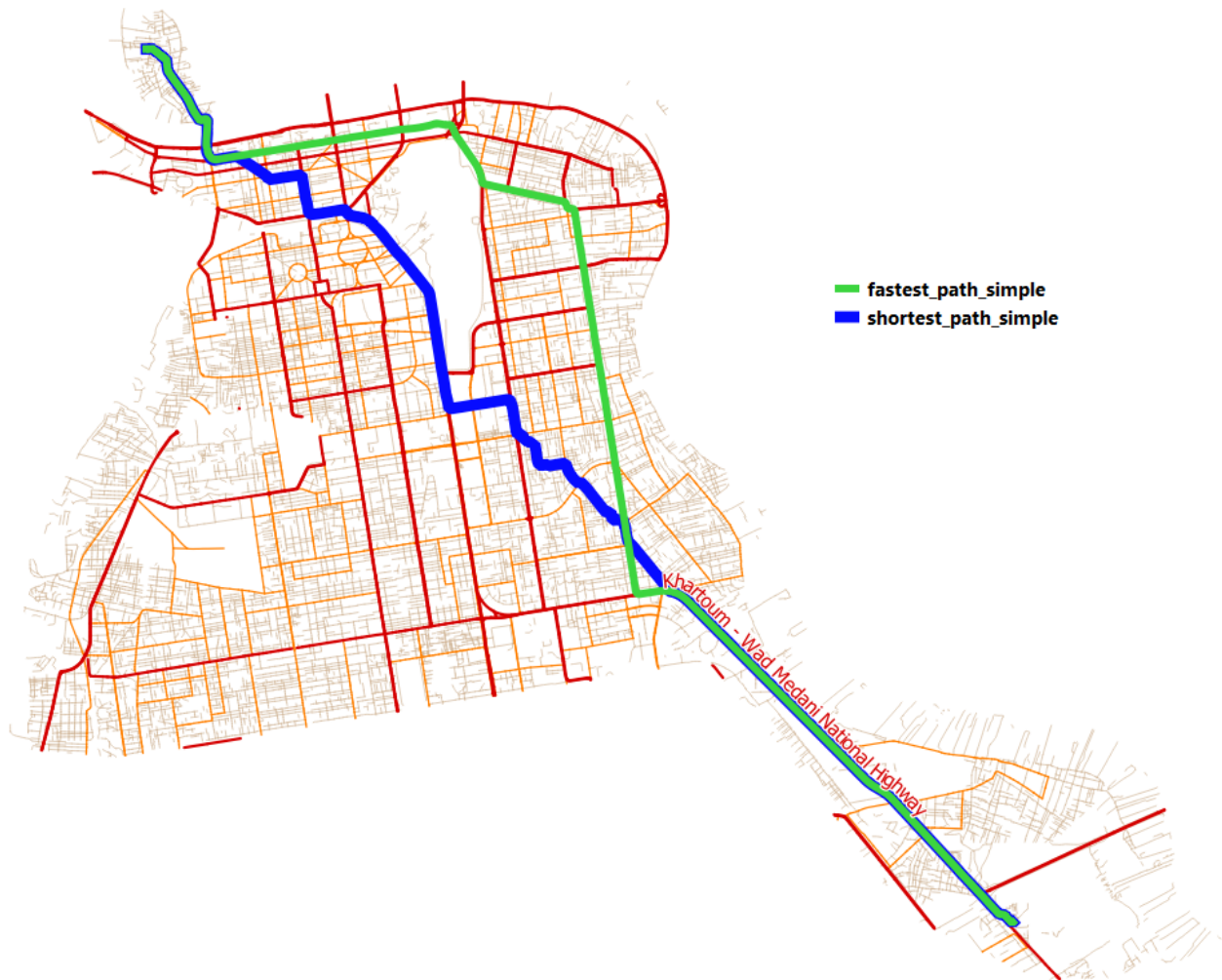


Fig. 7.3: Comparing the shortest with the fastest route

The results of a network analysis depend not only on the cost function but also on any restrictions that may apply. These restrictions take the form of obstacles that force the network analysis algorithm to look for an alternative.

What constitutes an obstacle greatly depends on the use case and travel modality you are analyzing. For example, in the case of road networks, what constitutes an obstacle depends on whether you are traveling by bicycle, car or even public transportation only. The classic example of a restriction are one-way roads. Let us look at an example.

**Task 5** Repeat the analysis described in Task 3. Make sure that under the Advanced settings you keep field *max\_speed* as the attribute to be used to calculate travel times, adding the *one\_way* attribute as the input for the *Direction field* and entering *1* under the *Value for forward direction* parameter Fig. 7.4.

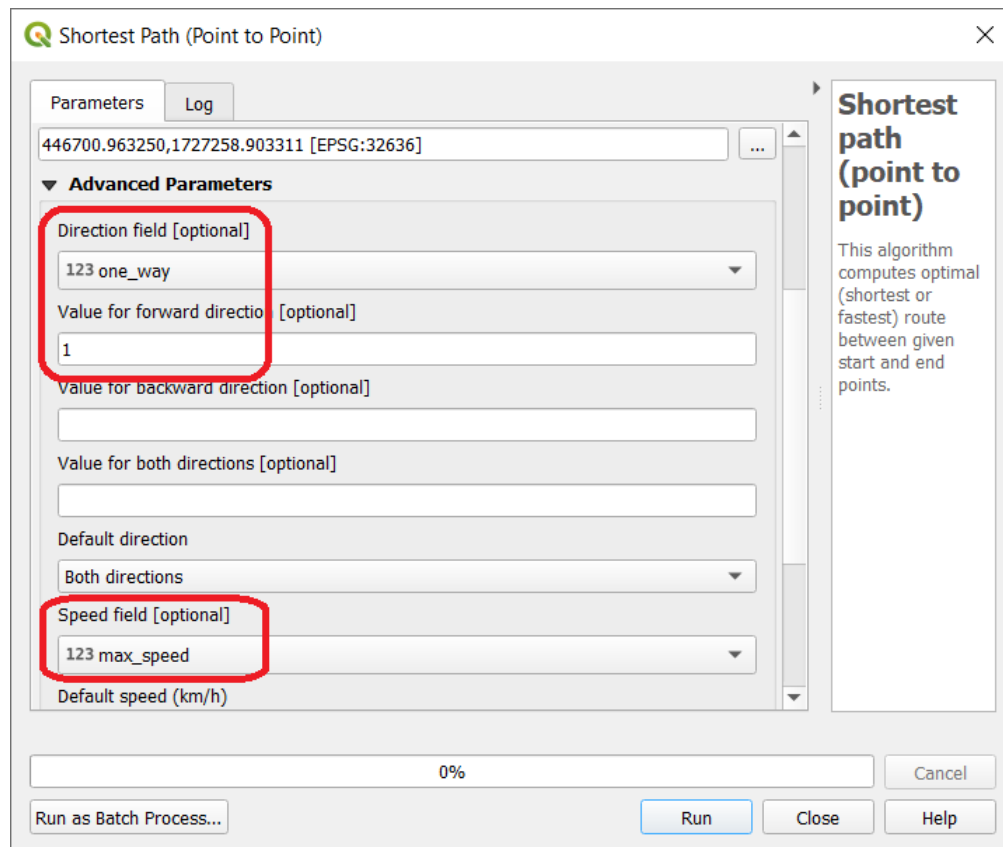


Fig. 7.4: Setting one way restrictions

**Task 6 (optional)** Consider to change the name and style of the output. For name, we suggest *fastest\_path\_restricted*.

Depending on what start and end points you are using, you might have sections of your route that take different roads due to the one way restriction, like in the example below, where the suggested route, when considering the one way restriction, takes a detour [Fig. 7.5](#).

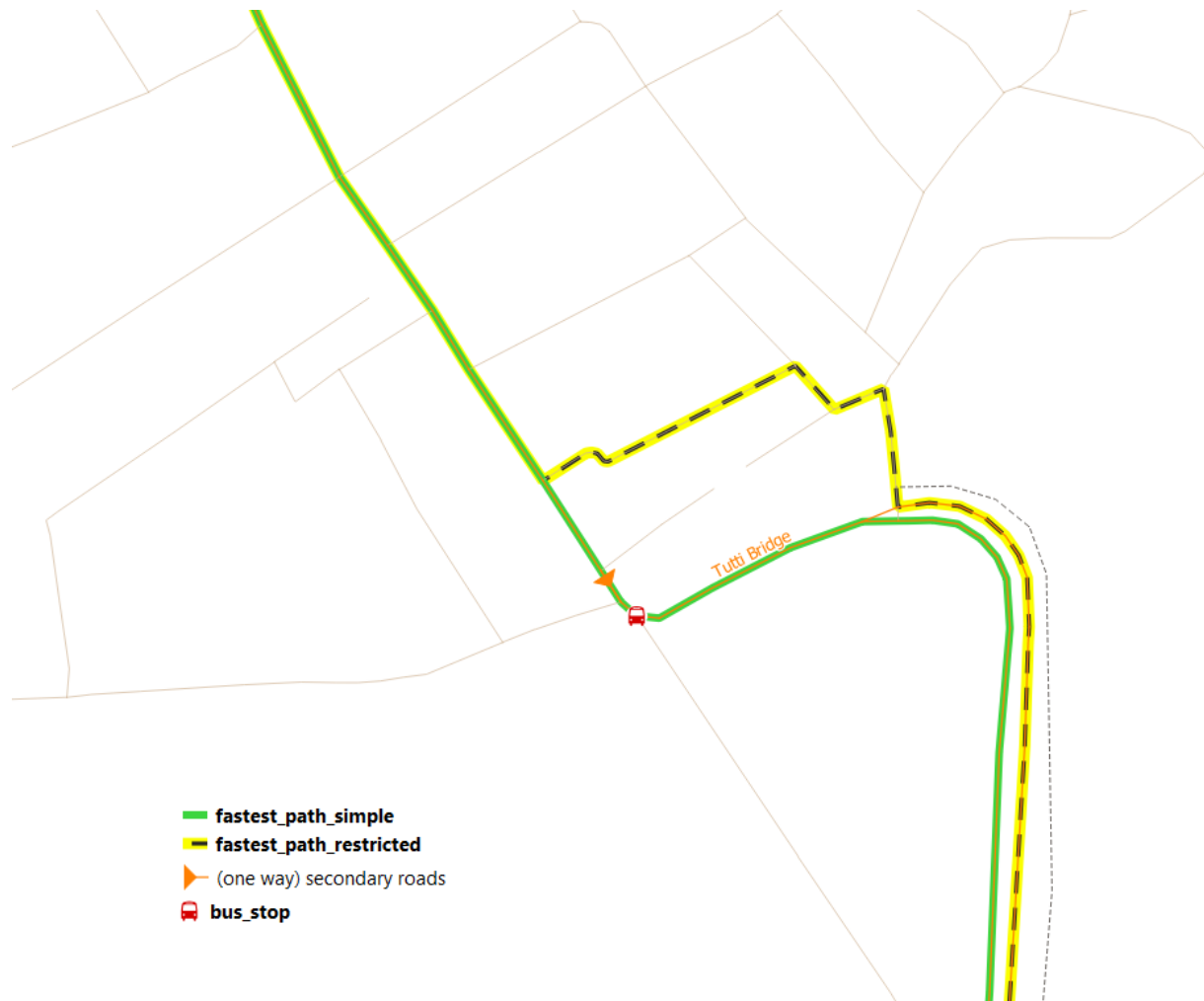


Fig. 7.5: Comparing fastest routes with and without one way restrictions

## 7.2 Advanced optimal path analysis

The previous section introduced the essential elements of optimal path analysis: cost functions and restrictions. We that understanding, we can extend the scope of our analysis and generate a collection of shortest paths from one origin to N destinations (or from N origins to one destination) in one go.

**Task 7** From the **Processing Toolbox**, filter the tools by **Shortest Path (Point to Layer)** and open the tool. For *Start point*, choose any building you wish, and for *Vector layer with end points* choose the layer *hospital*. Choose one of the cost functions and work on the remaining parameters as you prefer and click on **Run** button [Fig. 7.6](#).

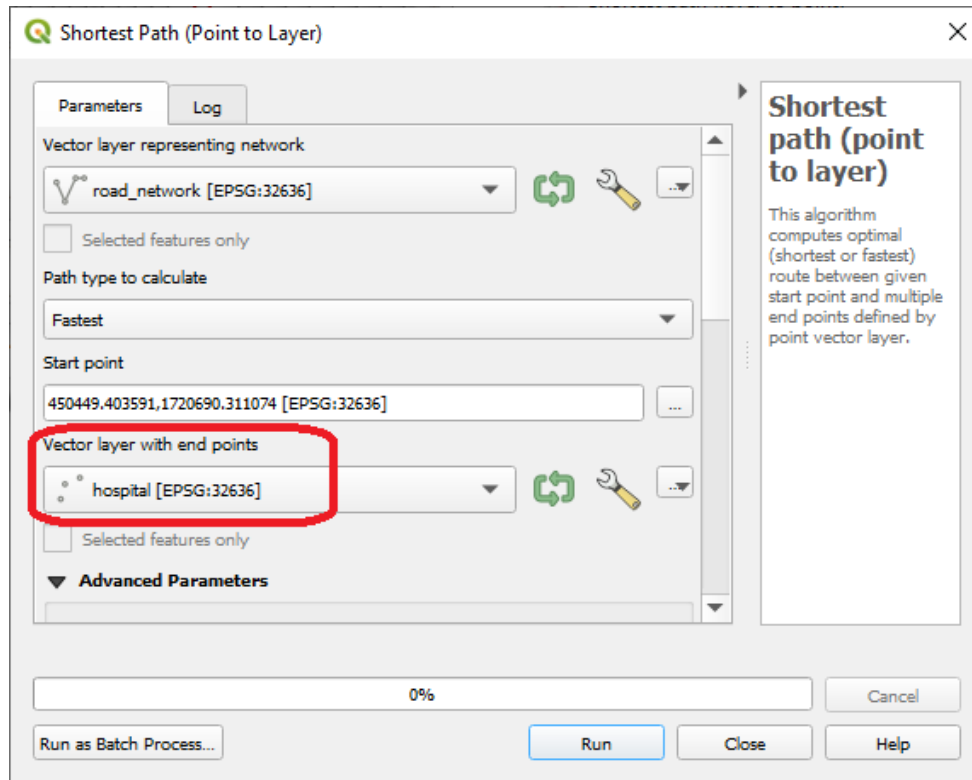


Fig. 7.6: The point to layer tool

The output will consist of a layer with a number of linestrings representing the most efficient route (according to the optimal path settings you chose) from the start point to the N destinations of the layer hospital [Fig. 7.7](#).

**Attention: Question 2** If you invert the operation (i.e. if you use the **Shortest Path (Layer to point)** tool), will you get the same routes and the same costs?

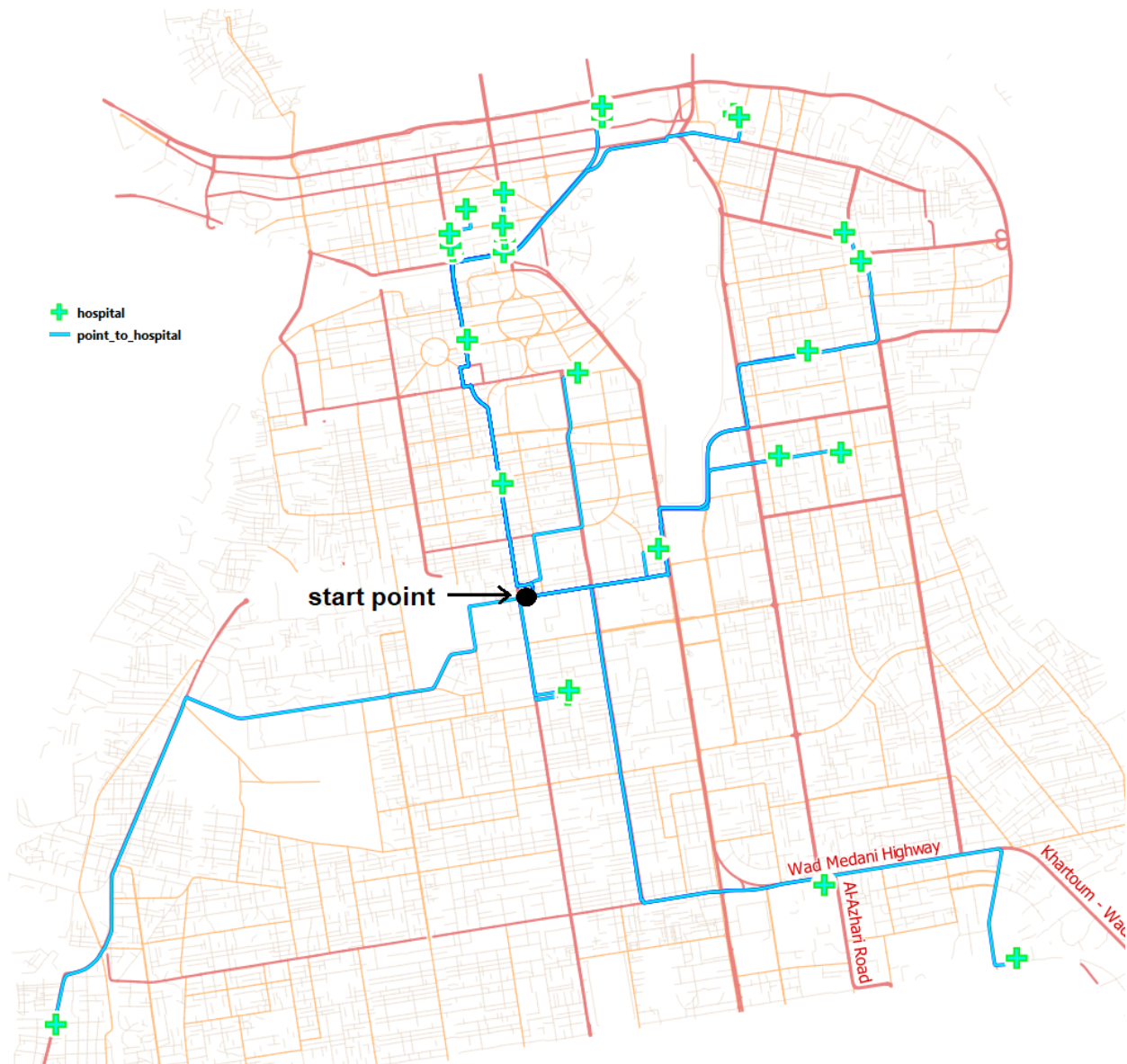


Fig. 7.7: Optimal paths from a start point to every hospital

We can also use the optimal path analysis to determine the best routes between  $N$  origins and  $N$  destinations. To demonstrate this, imagine a situation where for every school you need to determine what the closest bus stop is.

#### Task 8

To conduct this analysis, you will need to install the **QNEAT plugin**. This plugin provides additional network analysis capabilities to QGIS. Once you install it, QNEAT tools will appear under the processing toolbox.

For task 8, you may want to watch the video tutorial on [installing plugins in QGIS](#).

**Task 9** From the **Processing Toolbox**, search for **OD matrix from layer as lines (n:m)** and open the tool. Make

sure you provide the *road\_network* as the *Network*, *school* as the *From-Point Layer* and *bus\_stop* as the *From-Point Layer*. For both point layers, indicate the attribute *fid* as the *Unique Point ID Field*. Finally, as for the *Optimization Criterion* (i.e. the cost function), choose *Shortest Path* Fig. 7.8.

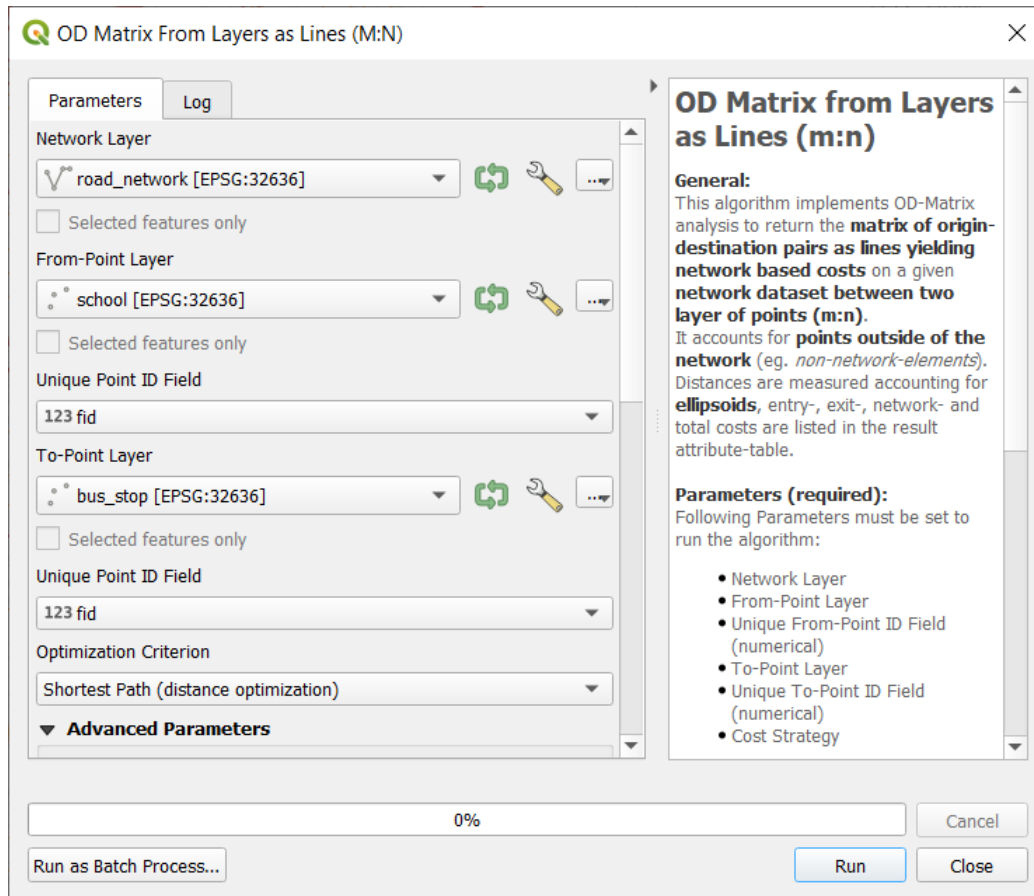


Fig. 7.8: The OD matrix from layer as lines tool dialog

The result might not be what you are expecting. The tool returns a Cartesian product where each combination is represented by a straight line and not the actually route Fig. 7.9. However if you examine the attribute table the cost values you see are calculated over the network. In other words, the costs are not based on Euclidian distances as the straight-line representation might suggest.





Fig. 7.9: The output of the OD matrix between schools and bus stops

If you want to obtain only the lines linking the schools to the closest bus stop, you can build a simple SQL query.

**Task 10 (optional)** From the **Processing Toolbox**, search for the tool **Execute SQL** Fig. 7.10. Indicate the layer over which the SQL query will operate (that will be the layer obtained from Task 9), and use the query below:

---

**Note:** **SELECT** origin\_id, destination\_id, min(total\_cost), geometry **FROM** [name of the layer generated by Task 9] **GROUP BY** origin\_id

---



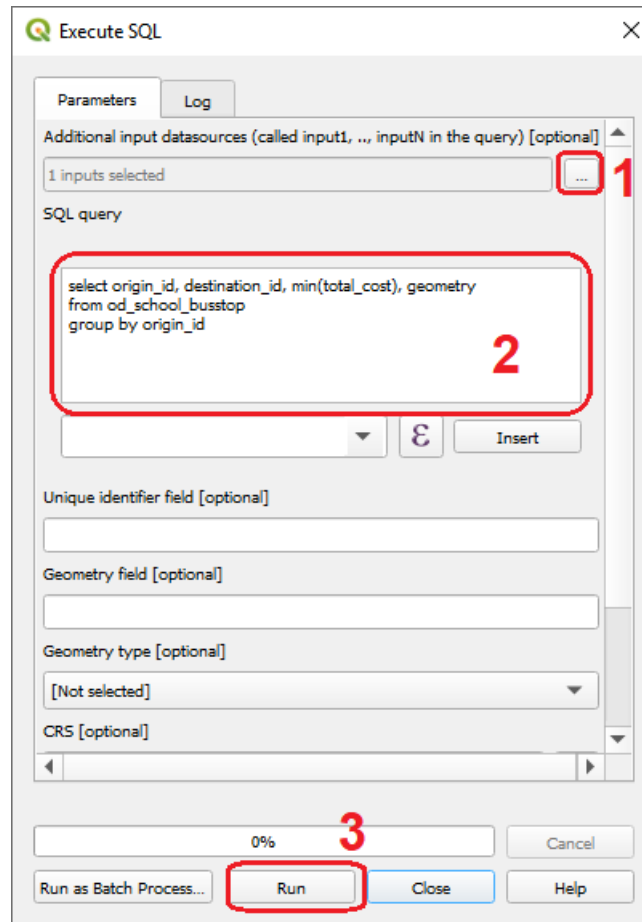


Fig. 7.10: Execute SQL

Depending on the style you apply to the output layer, you may see something like this [Fig. 7.11](#) – a direct link between a schools and the nearest bus stop.

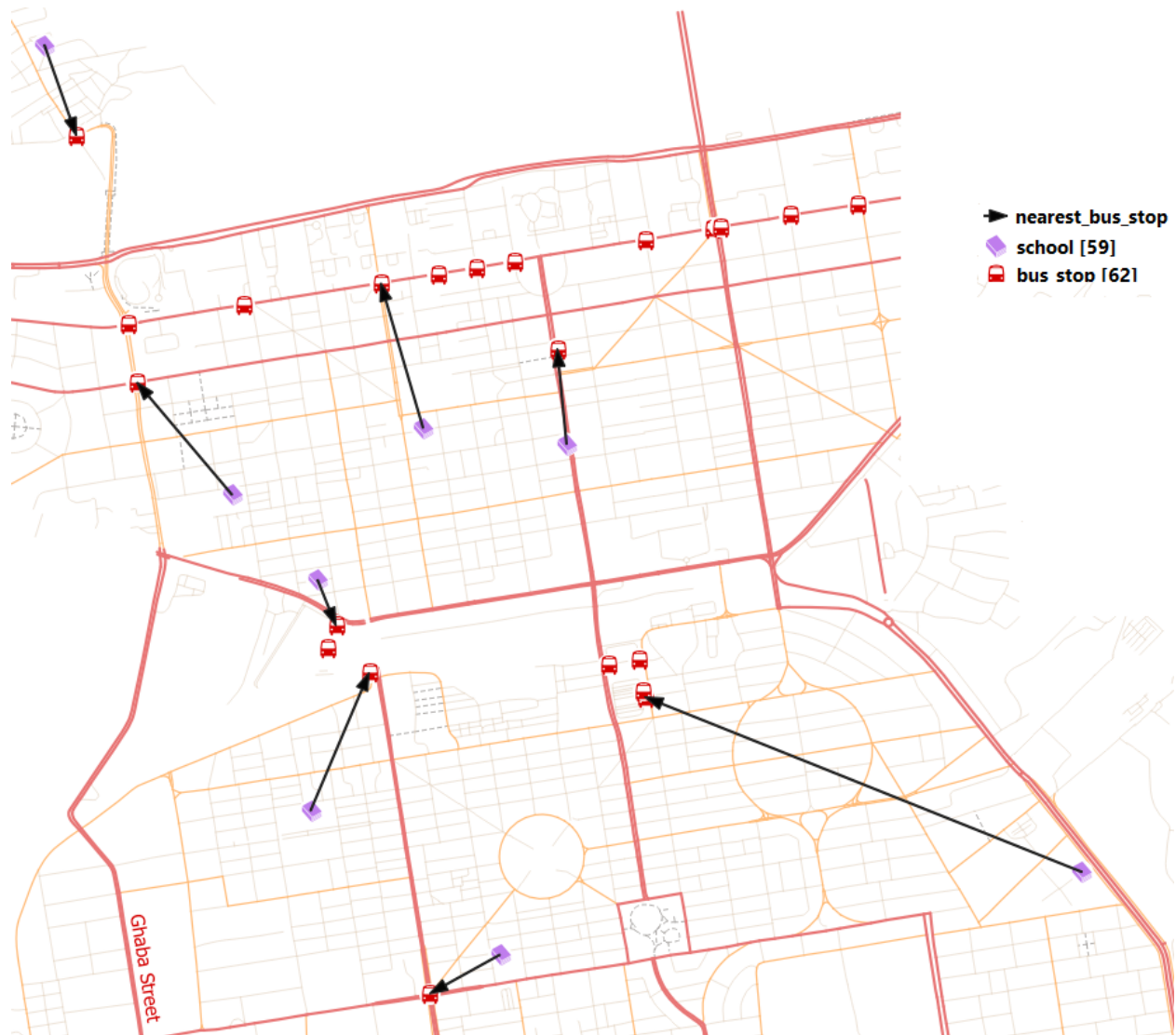


Fig. 7.11: Nearest bus stop

### 7.3 Network allocation

From the Living Textbook, you can read the following definition: In *network allocation*, we have a number of target locations that function as resource centres, and the problem is which part of the network to exclusively assign to which service centre.\*

**Task 11** From the **Processing Toolbox**, search for **v.net.alloc** and open the tool. Make sure you provide the *road\_network* as the *Input vector line layer*, *hospital* as the *Centers Point layer Layer*. Leave the other parameters with their default values and hit **Run** Fig. 7.12.

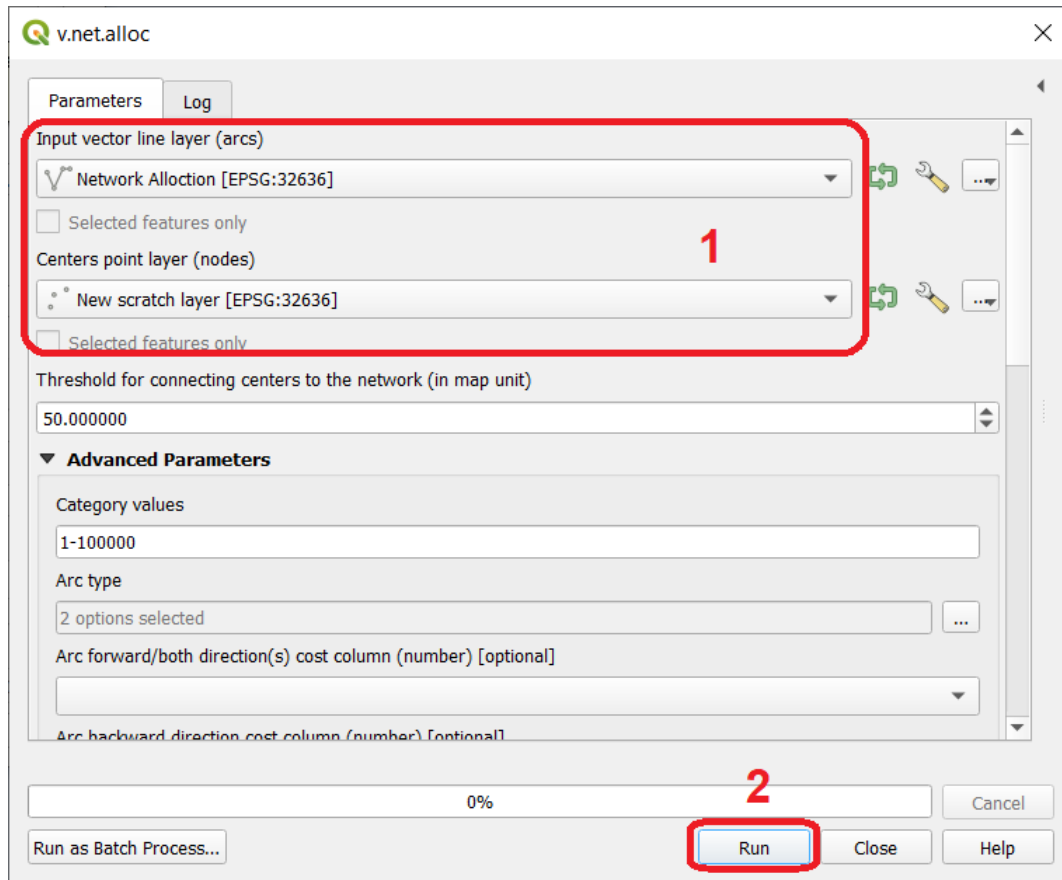


Fig. 7.12: The v.net.alloc dialog

**Task 12** Apply a categorized style to the output to better visualize the output `allocated-network`. Another option is to use the style file `network-allocation.qml` we provide along with the data for this exercise.



Fig. 7.13: The output of the v.net.alloc tool

The result is a partitioned network where each segment/line is assigned to the center whose cost to go to is lower than to any other center.

**Attention: Question 2** What cost function did the v.net.alloc tool use?

A common form of network allocation is a service area. It uses network partitioning to determine which network elements are within a certain threshold from a center in order to determine a service area.

An important characteristic of service area analysis is that the underlying network allocation is not necessarily mutually exclusive. This means the resulting service area may be within the threshold of more than one service center.

To demonstrate this concept we will look at an example of a service area defined as time zones. The situation is as follows: you want to have an estimation of fire emergency response times, and for that, you want to know what areas are within 5, 10 and 15 minutes from a firefighter brigade.

**Task 13** From the **Processing Toolbox**, filter the tools by **Iso-Area as Polygons (From Layer)** and open the tool. Indicate the *road\_network* as the Vector layer representing the network, the layer *fire\_station* as the *Start points*, and the *fid* field as the *Unique Point ID Field*. Then provide the parameters of the analysis: *Fastest Path* as the *Path type to calculate*, 900 seconds (i.e. 15 minutes) will be the maximum *Size of Iso-Area*, with a : guilabel:Contour Interval of 300 seconds (i.e. 5 minutes). Finally, assume a *Default speed* of 15 km/H iso-area-dialog.

The screenshot shows the 'Iso-Area as Polygons (From Layer)' dialog box. It has a 'Parameters' tab and a 'Log' tab. The 'Parameters' tab is active. The dialog is divided into several sections:

- Vector layer representing network:** A dropdown menu showing 'road\_network [EPSG:32636]'. A red box labeled '1' highlights this section.
- Start Points:** A dropdown menu showing 'fire\_station [EPSG:32636]'. A red box labeled '2' highlights this section.
- Unique Point ID Field:** A dropdown menu showing '123 fid'.
- Size of Iso-Area (distance or time value):** A text input field containing '900.000000'. A red box labeled '3' highlights this field.
- Contour Interval (distance or time value):** A text input field containing '300.000000'. A red box labeled '4' highlights this field.
- Cellsize of interpolation raster:** A text input field containing '10'.
- Path type to calculate:** A dropdown menu showing 'Fastest Path (time optimization)'. A red box labeled '5' highlights this section.
- Advanced Parameters:** A section with several options:
  - Entry Cost calculation method:** A dropdown menu showing 'Planar (only use with projected CRS)'.
  - Direction field [optional]:** A dropdown menu.
  - Value for forward direction [optional]:** A text input field.
  - Value for backward direction [optional]:** A text input field.
  - Value for both directions [optional]:** A text input field.
  - Default direction:** A dropdown menu showing 'Both directions'.
  - Speed field [optional]:** A dropdown menu.
  - Default speed (km/h):** A text input field containing '15.000000'. A red box labeled '6' highlights this field.
  - Topology tolerance:** A text input field containing '0.000000'.
- Output Interpolation:** A text input field containing '[Save to temporary file]'.
- Open output file after running algorithm:** A checked checkbox.
- Output Polygon:** A text input field containing '[Create temporary layer]'.
- Open output file after running algorithm:** A checked checkbox.

At the bottom of the dialog, there is a progress bar showing '0%'. A red box labeled '6' highlights the 'Run' button. Other buttons include 'Cancel', 'Close', and 'Run as Batch Process...'.

Fig. 7.14: The Iso-Area as Polygon dialog

As you can observe, the results show the areas that are within a certain distance but there is no indication of what is the closest fire station [Fig. 7.15](#).

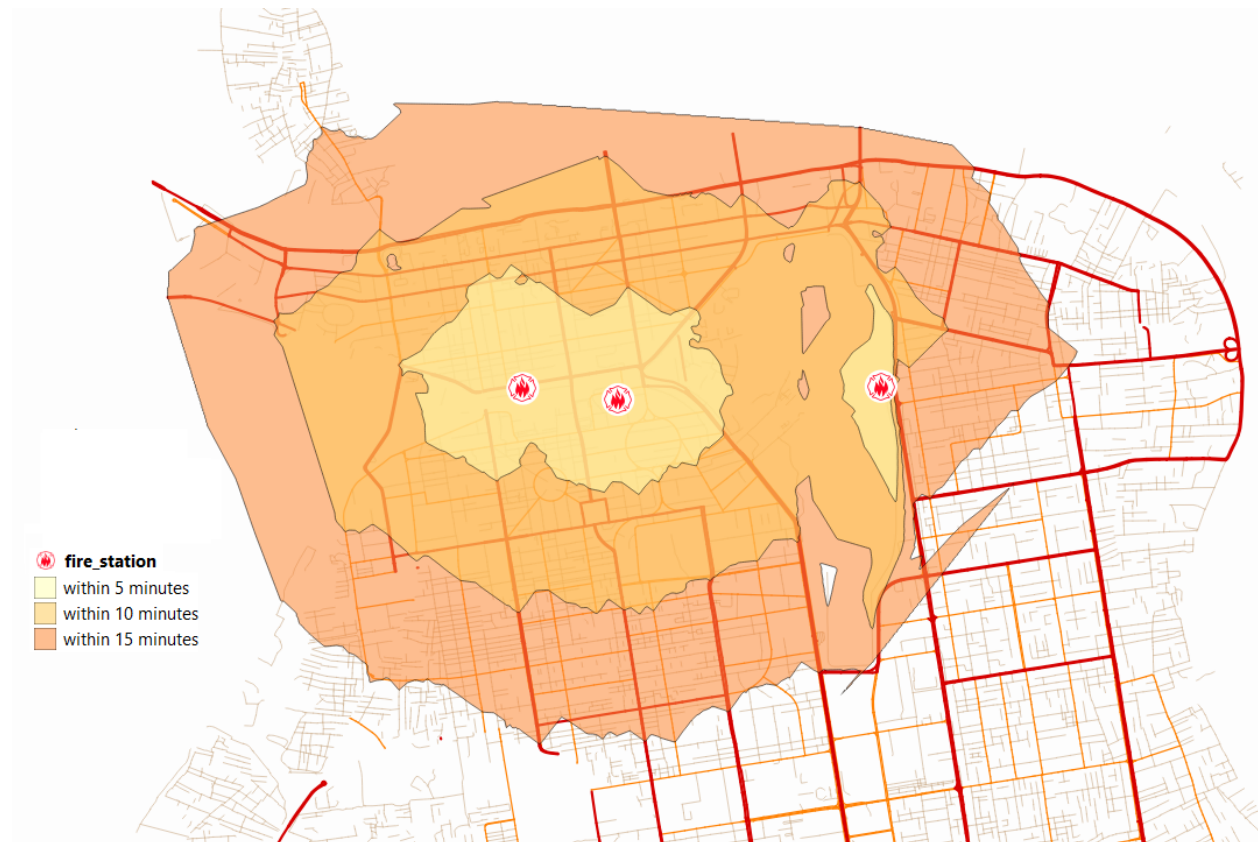



Fig. 7.15: Service area for fire stations based on time intervals

**Attention: Question 2** The analysis you conducted in Task 12 has a few conceptual shortcomings. Can you name a few and discuss their impact in current analysis?

## 7.4 Trace analysis

Trace analysis is a type of network analysis that focus not only on connectivity – an essential topological characteristic of a network, but also on direction. To demonstrate the idea we will use the same type of example depicted in the Living Textbook – the case of hydrological networks.

In this type of networks, and assuming the network is topologically consistent, it is often useful to identify streams upstream or downstream in relation to a given point.

**Task 14** To conduct this analysis, you will need to install the **Flow Trace plugin**. Once you install it, a new button  will be added to the plugin toolbar – simply click on it to use the Flow Trace

**Task 15** Using the *Select feature tool*, select any network segment you wish. Then open the *Flow Trace* plugin and choose the *Direction* you want to do the tracing – *Upstream* or *Downstream* net-flow-analysis.



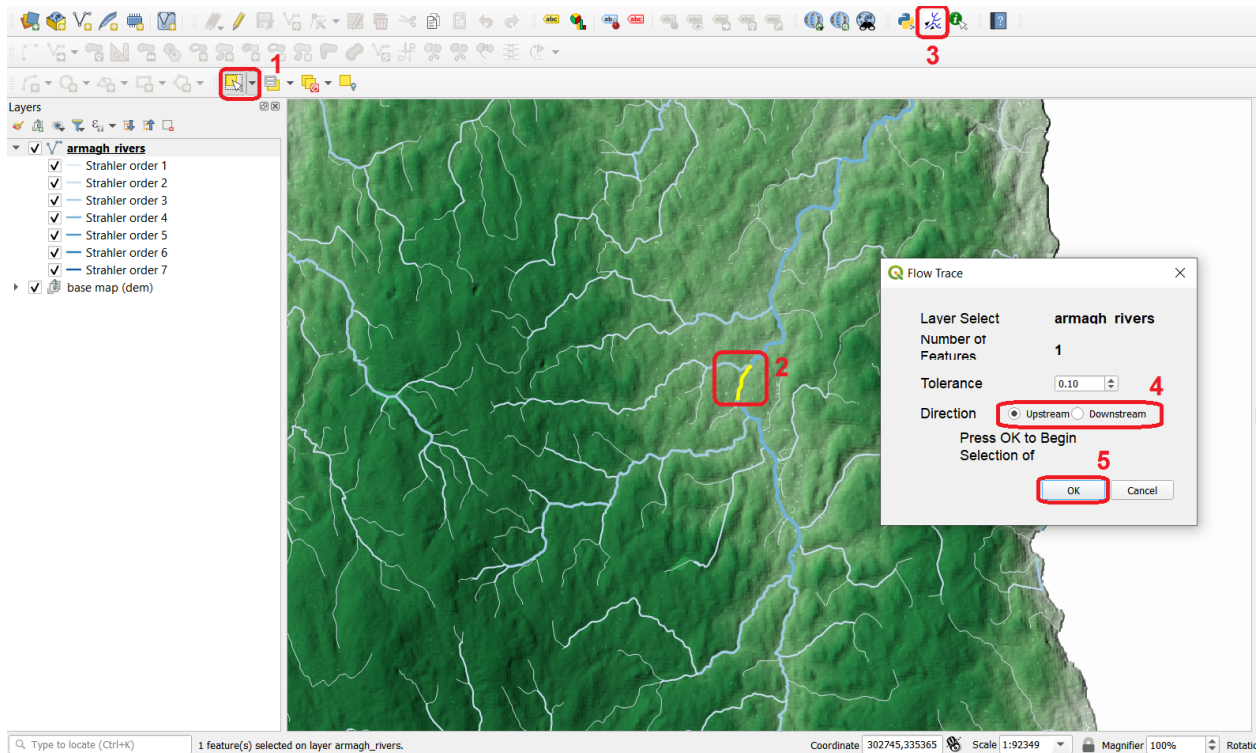


Fig. 7.16: Using the Flow Trace plugin

The result will be the selection of all the lines upstream, downstream or even both depending on what you do. net-flow-analysis-output.

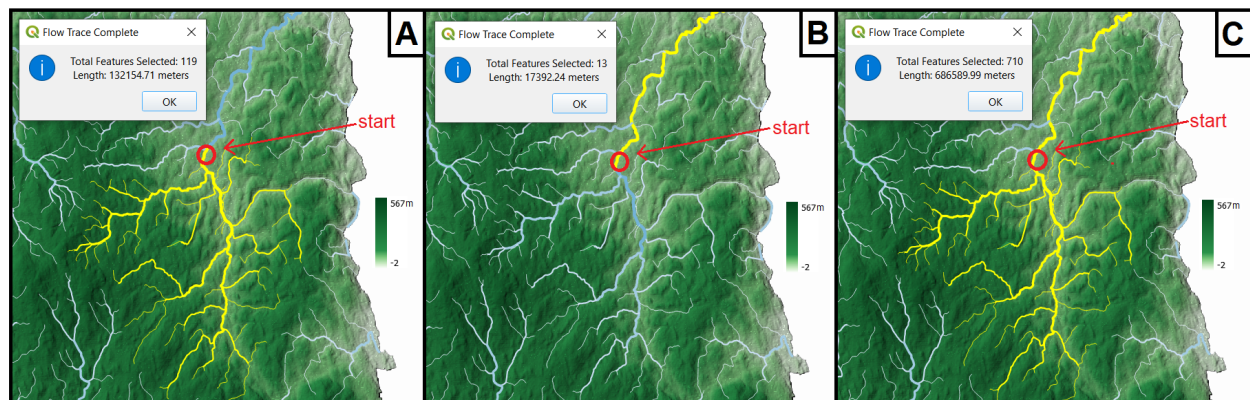


Fig. 7.17: Trace analysis example. [A] – upstream selection; [B] – downstream selection; [C] up and downstream selection

**Attention: Question 2** What is the essential characteristic a network has to have – besides having all the lines connected - in order for trace analysis to be able to operate?

## 7.5 Challenge

In the examples of network allocation, the problem was to know which network segments or which locations within the network, are closer or within a certain distance to certain number of centers.

What if the location of those centers is not yet decided? How to choose the candidate location for a center **in order to maximize the coverage of the service area?**

In your dataset you have three candidate locations for post boxes but they are still under discussion - You can only install one of them. Which of the three locations should be chosen considering:


- A. Demand is assumed to be coming from the buildings – that is where residents, services and industry are located. For the sake of simplification we aggregated buildings by **city blocks** where **one city block = one demand point**; The demand is provided as a dataset named *demand\_points.gpkg* – add it to your QGIS project.
- B. The chosen location should be the one that **serves at least 100 demand points**;
- C. The chosen location should have the **least aggregated travel cost** (the average of all the travel times from the candidate location to the demand points);
- D. People should not travel more than **2000m** to reach the post box;
- E. The analysis should be **based on distance only** as it is assumed most people within the 2000m.

*Section author: André Mano*





## RASTER ANALYSIS

### 8.1 Reclassification

Start this section by reading about the general aspect of  [Reclassification](#).

**Attention: Question.**

- Explain in your own words, which are the differences and similarities between classification and reclassification?
- What is the difference between reclassification of vector data and reclassification of raster data?

When applying a reclassification to a raster layer, you often reduce the number of different cell values and group them into class values. Through reclassification, the level of detail in a raster input layer can be reduced and reveal general patterns in a dataset. There are different classification methods for raster datasets, two of those are  [User controlled classification](#) and  [Automatic reclassification](#).

**Attention: Question.**

- In which type of classification is post-processing relevant?
- How does post-processing relate to the concepts of aggregation, merging, and dissolving?

### 8.2 Automatic Classification

---

**Important: Resources.** You will require the latest LTR version of [QGIS](#), plus the dataset [raster-analysis.zip](#). For this section, you will need the following files:

- `classification.qgs` – a QGIS project preloaded with the datasets described below;
    - `classes.tif`
- 

**Task 1** Below, you see a raster layer represented as an array of values. Classify the values by applying the *equal interval* and *equal frequency* techniques. Use five classes and write down the classified values in the boxes below. A good approach is to draw a raster attribute table first, which indicates the values and their frequencies, and then determine the new class values.

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

Equal Interval


Equal frequency  
(quantiles)


**Task 2** The input layer mentioned in the previous task, it is also provided in the exercise's dataset. Open the project 'Classification', then from the *Layers panel*, right-click over the layer classes, and apply a symbology using an **equal interval** classifier, Fig. 8.1. Repeat the previous steps, and this time apply a **quantile classifier**. Compare the results.

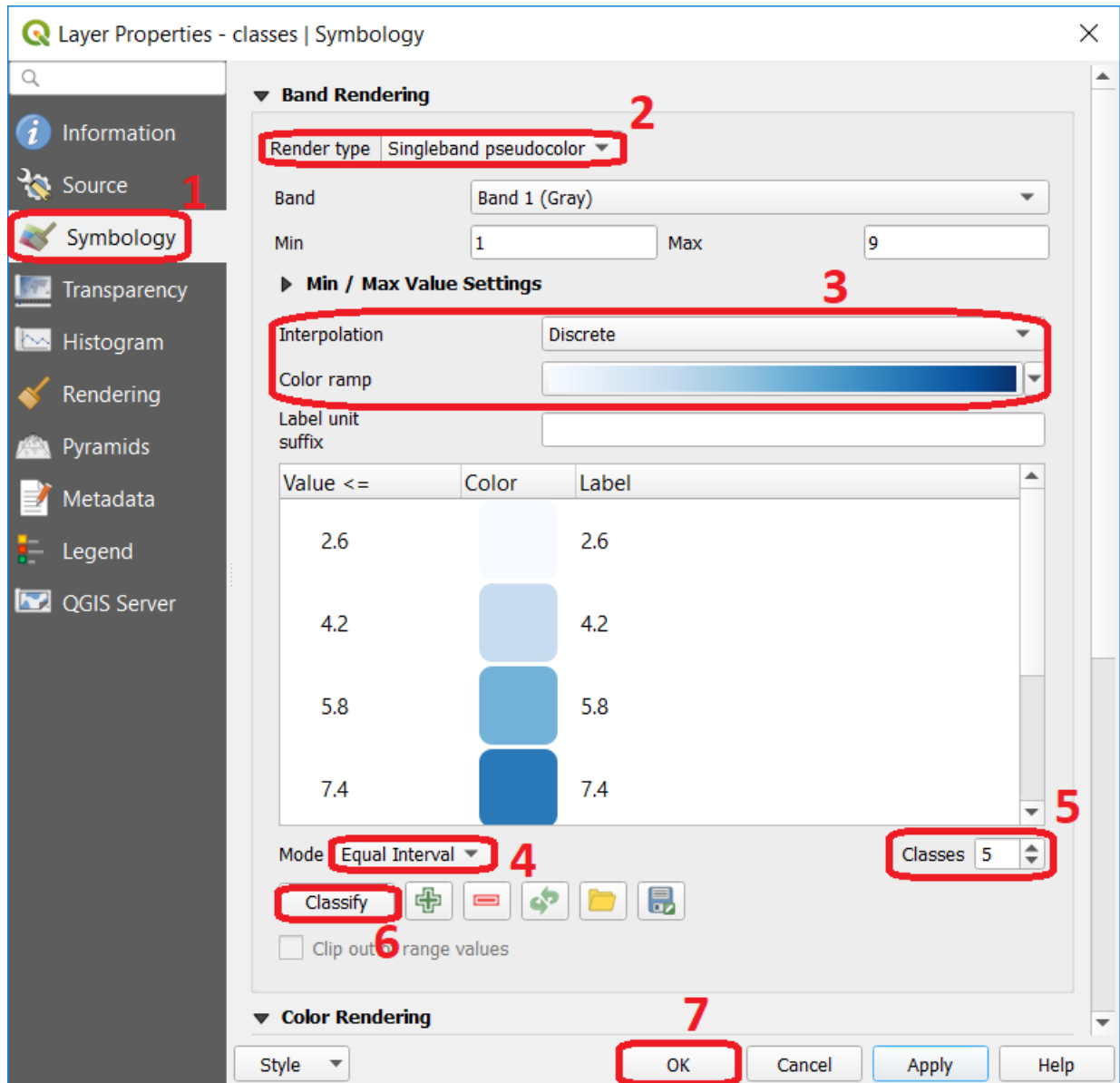


Fig. 8.1: Step to perform automatic classification via the symbology window.

## 8.3 Manual Classification

You can also manually control the number of classes and the ranges of values belonging to each class. You can do this using the controls in the same menu as before, and for example, change the **mode** or the number of **classes**. You can directly type different values for each class. Fig. 8.2. Additionally, you could define reclassification rules that will create a new raster.

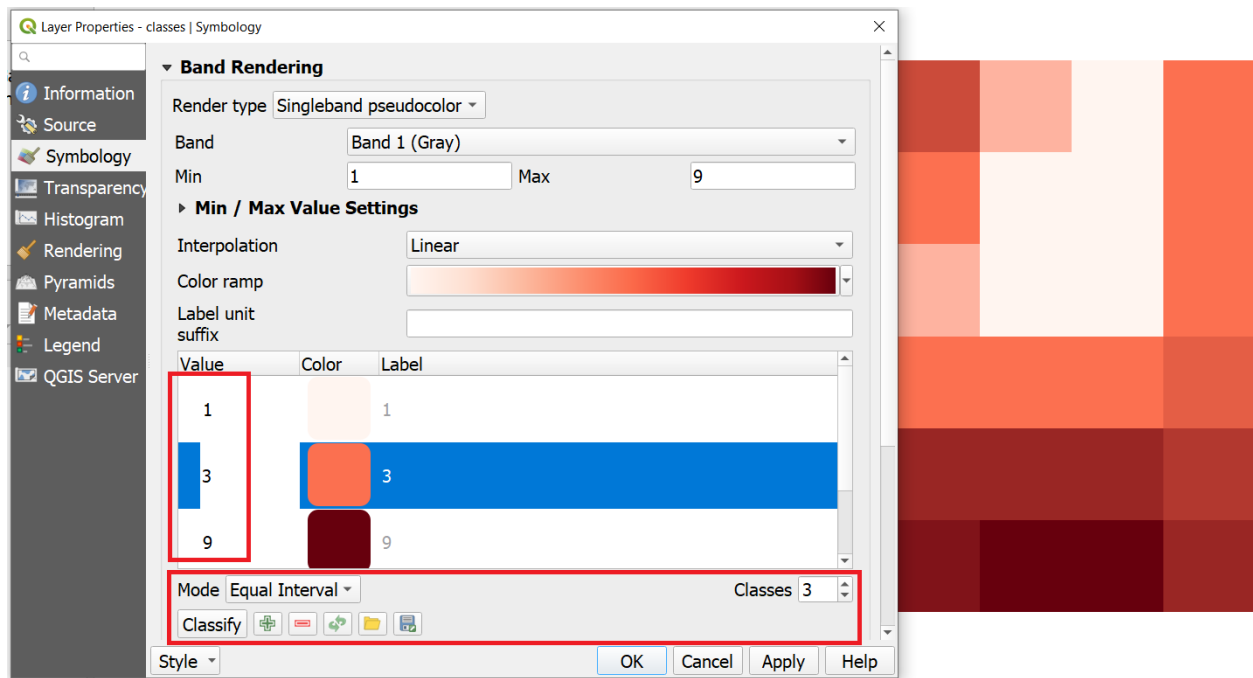


Fig. 8.2: Manual classification using the symbology window.

**Note: QGIS.** When You classify a raster using the symbology menu, you are not changing the pixel values; you are merely grouping the pixels for visualisation purposes. If you want to generate a raster whose pixel values are actually overwritten according to whatever rules you decide, then you have to create a new raster using the steps in the next task.

**Task 3** Say you want to create a new raster which group values (reclassify) according to what we shown in Fig. 8.2. That is 3 classes ( *class1* = 1, *class2* = 2 to 6, and *class3* = 7 to 9). Go to *Processing toolbox > Classify by table* and write down the reclassification rules as shown in Fig. 8.3.

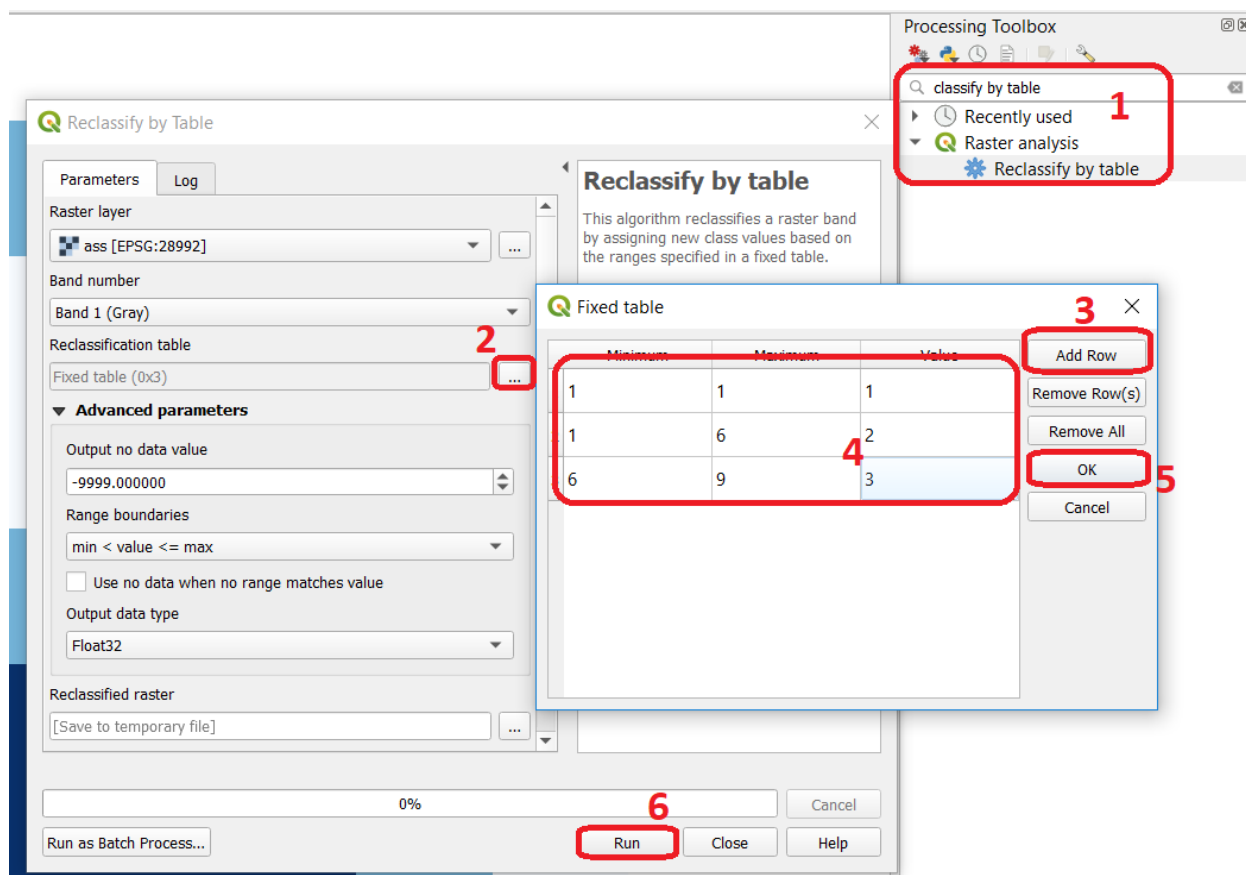


Fig. 8.3: Steps for creating a raster using Reclassify by table

**Note: QGIS.** We highly recommend that at this point you install the **Value Tool Plugin**. It allows you to see the pixel values of all the active raster at the current location of the mouse pointer. Fig. 8.4

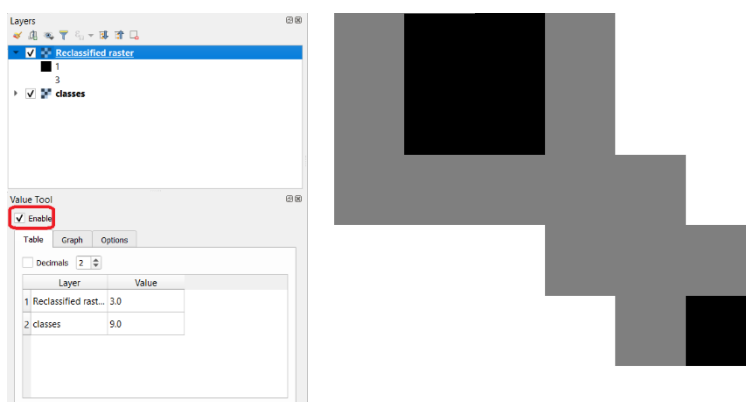
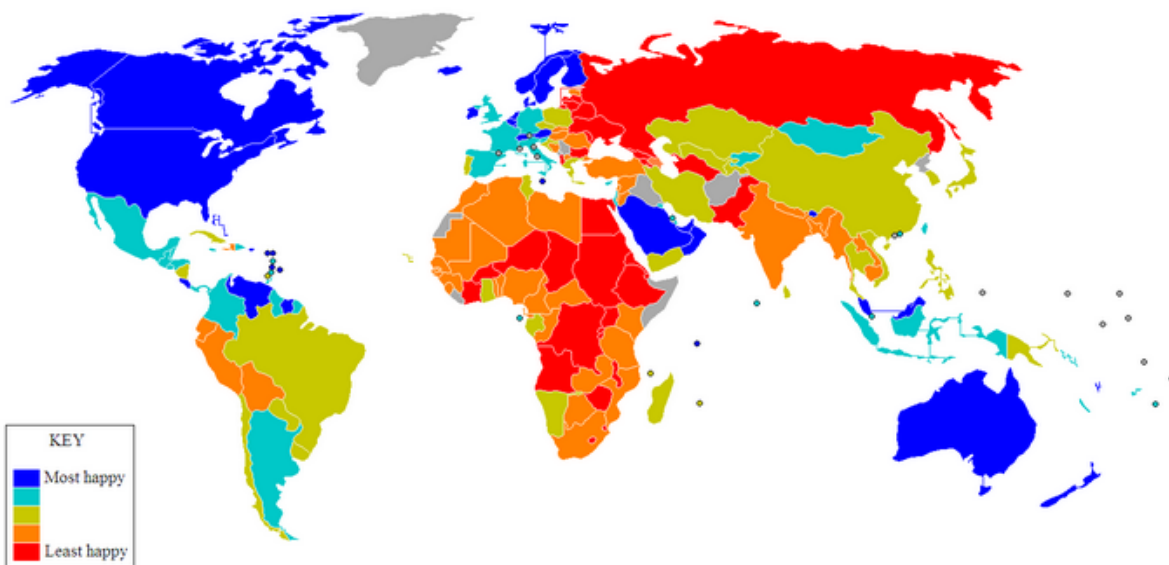


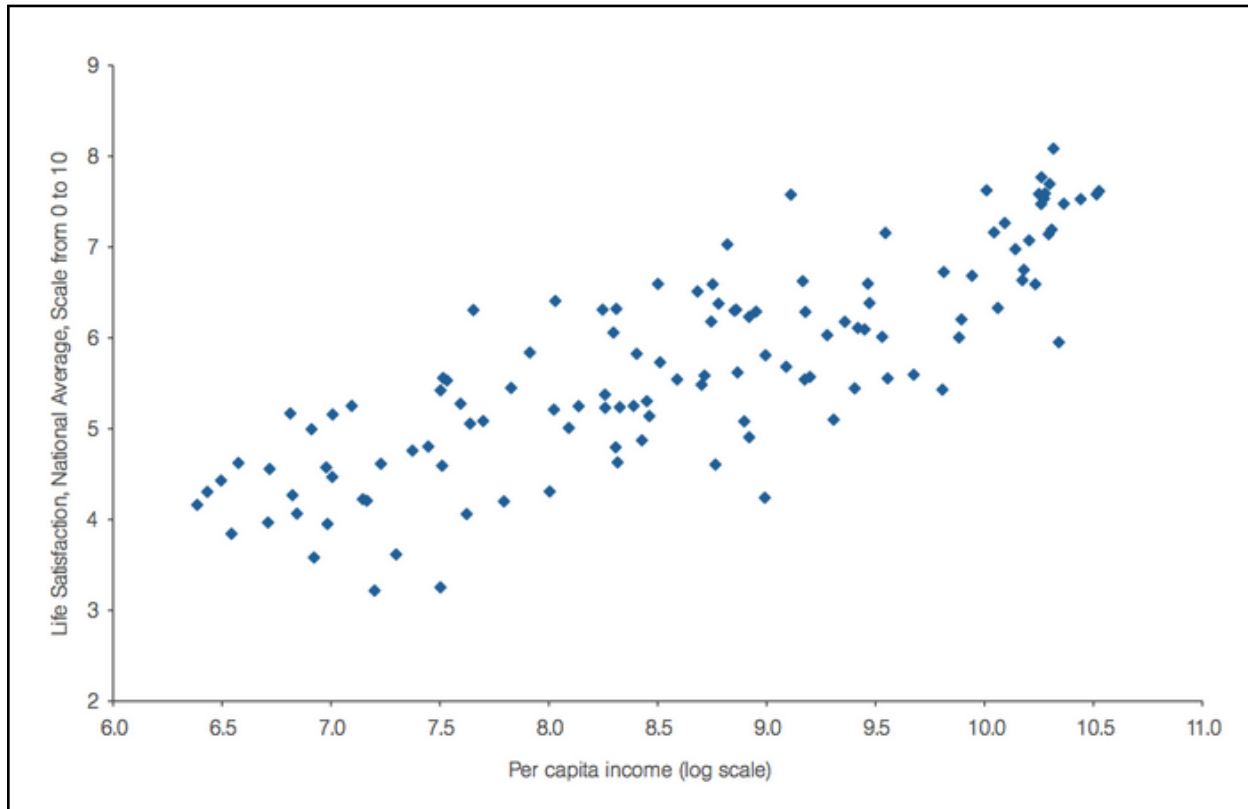
Fig. 8.4: Result of applying a manual classification using Reclassify by table

Please refer to the QGIS training manual, section [Installing Plugins](#), or watch to the video tutorial on [Installing Plugin in QGIS](#).


**Attention: Question.**


- If we have pixels whose values are not covered by any reclassification rule, what value will it be assigned to those pixels in the new raster?
- Will the new raster delete those pixels (i.e. null), assign a random value, preserve the original value, or something else?
- To test your understanding on classification techniques; imagine you are going to make a map showing the “happiness” of the countries of the world, as of 2012. You want to reclassify the data on happiness. Below you see a map and a scatter plot of the data. What reclassification technique would you apply and why?





## 8.4 Raster Overlay

During the vector exercise, you were already introduced to the concept of overlaying. In case you need to refresh your memory, check  [Overlay Analysis](#).

 [Raster Overlay](#) operations can be classified into four groups: **Arithmetic**, **Boolean and Logical**, **Conditional**, and **Decision tables**. All of these methods are cell by cell operations.

**Important: Resources.** You will require the latest LTR version of [QGIS](#), plus the dataset [raster-analysis.zip](#). For this section, you will need the following files:

- `overlay.qgs` – a QGIS project preloaded with the datasets described below;
  - `suit1.tif`
  - `suit2.tif`

### 8.4.1 Arithmetic operators

Imagine you want to find a location for a new house. You have two criteria: close to a school and in a safe neighbourhood. You prepared two input layers; 'suit1' contains information about the closeness to schools, and 'suit2' contains information about the safety of neighbourhoods.

When using arithmetic operators to combine multiple raster layers, it is imperative to evaluate the input values of the layers. In case the layers represent suitability indicated as *zero*="not suitable" and *one*="suitable", arithmetic operators can be used to combine these input layers via raster overlay; Fig. 8.5. You may wonder how you would get such a binary input layer. A suitable method is to use a *Reclassification* to convert your input layer into a suitability layer with only relevant classes.

1	1	1	1	0	0	0	0	0	0
1	1	1	1	0	0	1	1	0	0
1	1	1	1	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	1

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	1	1	1	1
0	0	0	0	0	0	1	1	1	1

Fig. 8.5: Values of two suitability layers for the location of a new house. Left: 'suit1'. Right: 'suit2'

#### Attention: Question.

- Which arithmetic operators can you use to combine these layers and obtain a suitable location for the new house?
- How many different output values will you get in your out layer?
- What would be the total suitable area in your output layer?

**Task 4** The layers shown in Fig. 8.5 are also available as raster layers in the exercise's dataset. Open the project `Overlay.qgs` and conduct an arithmetic overlay using 'Suit1' and 'Suit2' as input layers. You will need the **Raster Calculator** Fig. 8.6 Then check the results.



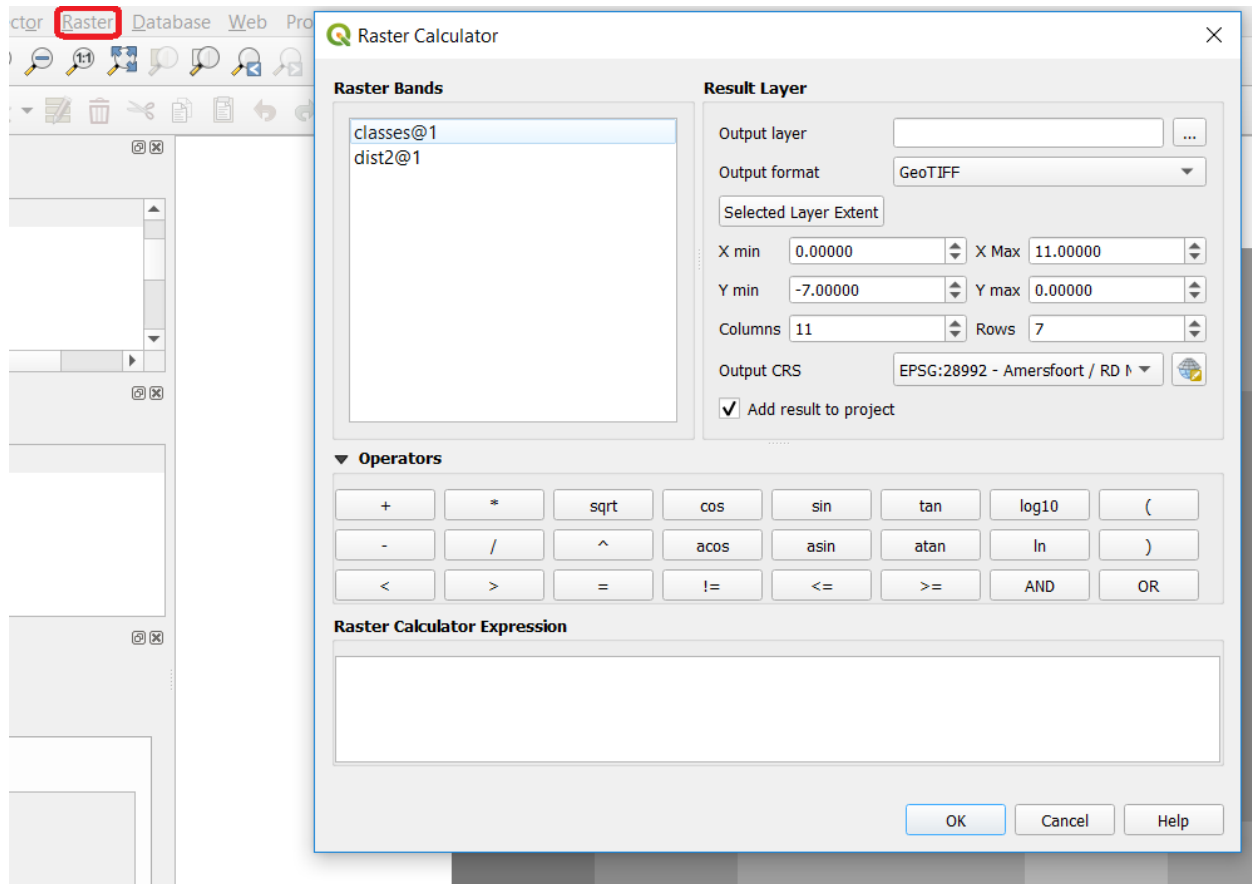


Fig. 8.6: The Raster Calculator

### 8.4.2 Comparison and Logical Operators

You may already be familiar with both Comparison (e.g. =, >, <=) and Logical operators (e.g. AND, OR, NOT). They were introduced as part of attribute selection using SQL.

Assume that the input layers for a raster overlay, have not been classified as suitable or unsuitable yet, rather they contain other values. In the example of the location of a new house. The values might represent the perception of safety in each neighbourhood, and the distance to schools. Then, values of distance would be continuous (floating point), and values of safety would be discrete (integer). See Fig. 8.7.

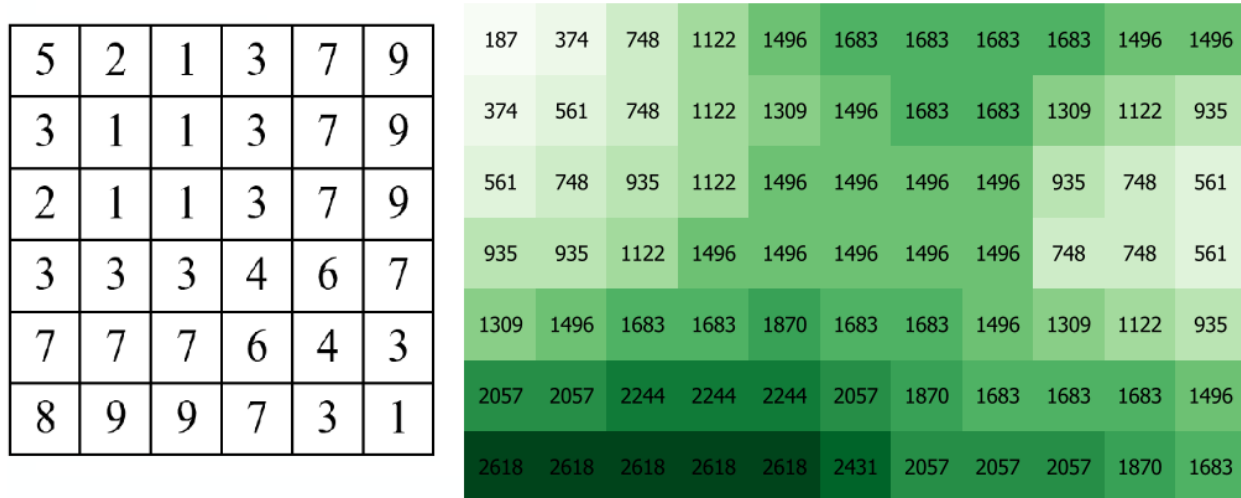


Fig. 8.7: Raster layers representing safety of neighbourhoods (left), and distance to schools (right)

**Task 5** Write down an expression to combine the raster layers in Fig. 8.7, using comparison and logical operators. Note that the size of these layers is not the same, but they do overlap. Assume that for the raster on the left (safety), pixels with values of 3 and 7 are suitable. For the raster on the right (distance to schools), pixels with values below 1000 are suitable.

**Task 6** The layers are available in the exercise's dataset: `classes.tif` (safety) and `dist2.tif` (distance to schools). Try out your expression in the **Raster Calculator**.

**Attention: Question.**

- How many output cells are suitable?
- What is the extent of the output raster layer?
- What is the range of output values?

### 8.4.3 Conditional Operators and Decision Tables

In the previous type of overlay, the output value was always either zero (un-suitable) or one (suitable). Yet sometimes, you would like to generate values expressing the degree of suitability. For example, where one is not so suitable, three is reasonably suitable, and ten is perfectly suitable. Such a thing is not possible using the type of analysis discussed in the previous section. We achieve such analysis by using **conditional statements** and **decision tables**.


**Note: QGIS.** In the **Raster Calculator**, conditional statements are implicit in the output of other operators. For example, if given two raster, `'raster1'` and `'raster2'`, we want as output the sum of pixel values of `'raster1'` and `'raster2'`, whenever the value of a pixel in `'raster1'` is more than 5. We can use the follow expression:

```
(( raster1@1 > 5) * raster1@1 + raster2@1)
```

**Explanation.** `(raster1@1 > 5)` states a condition that will return 1 (*True*) when a pixel in `'raster1'` is more than 5, and 0 (*False*) otherwise. Then, we retrieve the original pixel values in `'raster1'` by using a multiplication `( raster1@1 > 5) * raster1@1`. If the pixel in `'raster1'` is more than 5, it will return a 1, and 1 multiplied by any number will always return that number.

Finally, `+ raster2@1` adds values in `'raster2'` to the values of `'raster1'`, after applying the condition stated by the comparison operator. You can see more examples in the [QGIS documentation](#).

**Task 7** Using the case explained in [Fig. 8.7](#), write down an expression for the raster calculator that uses conditional statements to produce an output raster with different levels of suitability. Define at least three suitability levels.

An alternative to conditional statements is using decision tables. Decision tables are often used when there are many input raster or when the output raster contains classes with a value that are the result of meeting different conditions. See an example in the bottom of the explanation on  [Raster Overlay](#).

**Task 8** Rewrite the conditional statement from the previous task using a decision table.

**Attention: Question.** What is the difference between reclassification based on an input table and the Decision Tables discussed above?

## 8.5 Raster Measurements & Computations

There are several questions related to  [Raster Measurements](#) that can be answered using raster analysis. For example,

- How far are two locations?
- How long is this line?
- What is the distance to the nearest point?
- What is the area closed to this point?

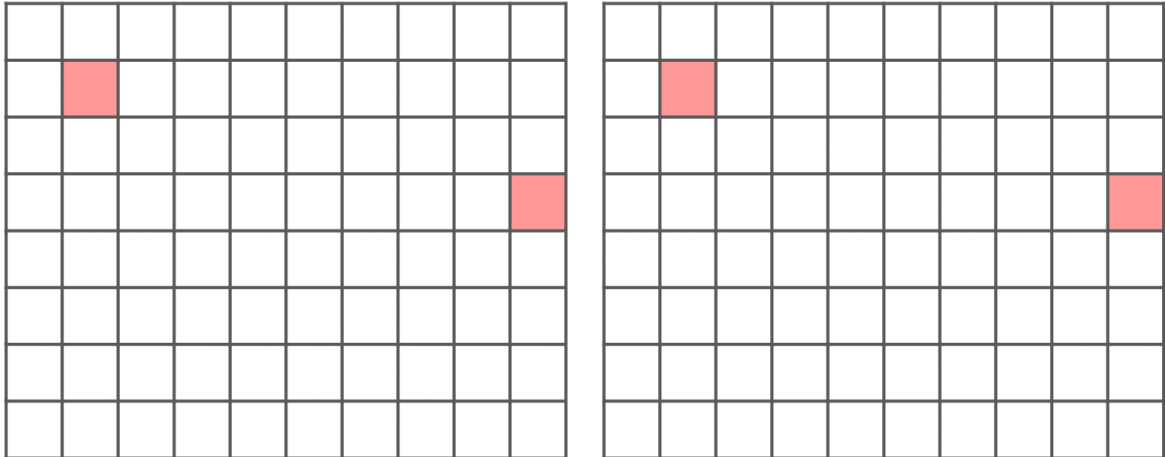
**Important: Resources.** You will require the latest LTR version of [QGIS](#), plus the dataset [raster-analysis.zip](#). For this section, you will need the following files:

- `distance.qgs` – a QGIS project preloaded with the datasets described below;
  - `raster_points.tif`
- `surface_analysis.qgs` – a QGIS project preloaded with the datasets described below;
  - `mount_etna_dem.tif` – a Digital Elevation Model

### 8.5.1 Distance

Distance, in a raster layer, can be measured as “**Euclidean**” or “**cell centre to cell centre**”. Euclidean distance is measured from the cell centre of the origin-cell to the cell centre of the destination-cell in a straight line. However, for some operations, we use a distance measured from the cell centre of the origin-cell to the cell centre of an adjacent cell until reaching the cell centre of the destination-cell.

**Task 9** Below you see two pictures. In each picture, two marked cells. Draw a line to represent the distance between the two cells. For the one *on the left*, use the concept of Euclidean distance. For the one *on the right*, use the concept of ‘cell centre to cell centre’.



**Attention: Question.** How far are the two cells, from the previous task, when the size of a cell (resolution) is  $10 \times 10 \text{ m}$ ?

**Task 10** Compute the distance over a raster layer. Open the project `distance.qgs`. You will see a layer named `'raster_points'`. Go to *Raster > Analysis > Proximity* and generate a raster distance map. Answer the following questions:

- Is it possible to select which type of distance you want to measure?
- Is the **Proximity** tool calculating the Euclidean distance or cell centre distance?

Make sure the `'raster_points'` layer is on top and use the **Value tool** to inspect the pixel values; Fig. 8.8 . It will make it easier to interpret the data.

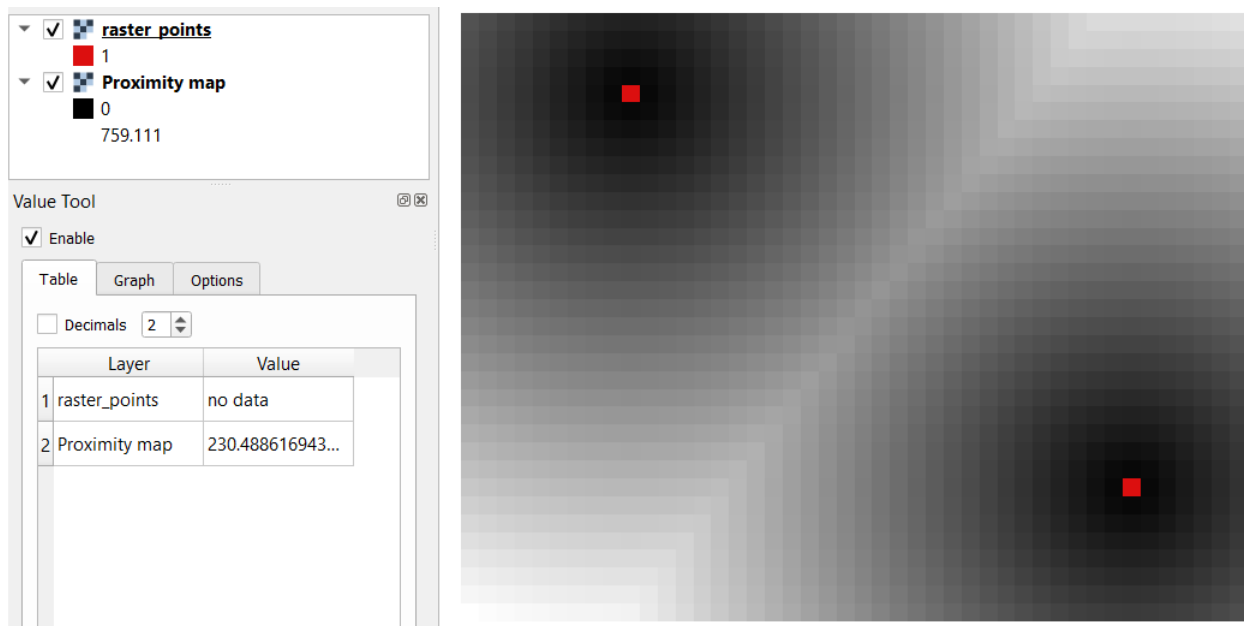



Fig. 8.8: Inspecting the distance raster

## 8.5.2 Computation of Diffusion

The computation of  **Diffusion** differs from distance computation in the sense that diffusion takes into account both distance and **resistance**. Diffusion is also referred to as the least accumulated cost distance, where cost refers to the resistance factor.

The following elements are essential to understand diffusion:

- It requires two inputs. One input contains the source(s) cells layer, and the other contains the *resistance or cost* layer.
- Distance is calculated from *cell centre to cell centre*. Because there is a difference between distances between the cell centres of the neighbouring cells, i.e. the distance of diagonally adjacent cells is longer, we have to take this into account.
- It determines the minimal cost of arriving at a particular cell. If there are multiple paths to reach a cell, diffusion calculates the value for each cell, but it will assign the lowest value.

Below you see an explanation of the computation of diffusion.

### Neighborhood functions

Resistance layer

1	1	1	2	8
4	4	5	4	9
4	3	3	2	10
4	5	6	8	8
4	2	1	1	1

■ The GIS computes the total resistance for spreading from  $c_{src}$  to  $c_n$  as  $\frac{1}{2}(\text{val}(c_{src}) + \text{val}(c_n))$


■ This is half of the resistance value of the source cell ( $c_{src}$ ), because this cell is only traveled half, plus half of the resistance value of the cell north of it

■ The value of the source cell of course is 0.

Minimal total resistance

?	
0.00	

$(4+4)/2 = 4$



Page Number: 190

## Neighborhood functions

**Resistance layer**

1	1	1	2	8
4	4	5	4	9
4	3	3	2	10
4	5	6	8	8
4	2	1	1	1

$$(4+5)/2 * \sqrt{2} = 6.36$$

**Minimal total resistance**

ITC

- The GIS computes the total minimal resistance raster for a diagonal neighbour as  $\frac{1}{2} (\text{val}(C_{src}) + \text{val}(C_{ne})) * \sqrt{2}$
- This is half of the resistance value of the source cell plus half of the resistance value of the cell to the north-east multiplied by the square root of 2

**Task 11** The best way to learn how the computation of diffusion works is by doing it manually. Use pencil and paper to compute the diffusion on the raster layers represented below.

5	7	2
1	6	3
0	8	11

**Resistance layer**  
(green cell is the origin)


**Outcome layer**

**Task 12** You can also experiment with the computation of diffusion in QGIS. In a previous task, you created a distance layer using the 'raster\_points' layer. Here, you will use that distance layer as a *resistance layer* (In QGIS this is called 'cost layer'). Go to *Processing Toolbox > r.cost*, and provide the inputs as depicted in the screenshot below; Fig. 8.9 The tool will generate more than one output, ignore all of them except for the 'cumulative cost' layer.

With the help of the **Value tool**, inspect the values of the pixels of the proximity map and of the '*cumulative cost*' layers. Make sure you understand what those values represent.

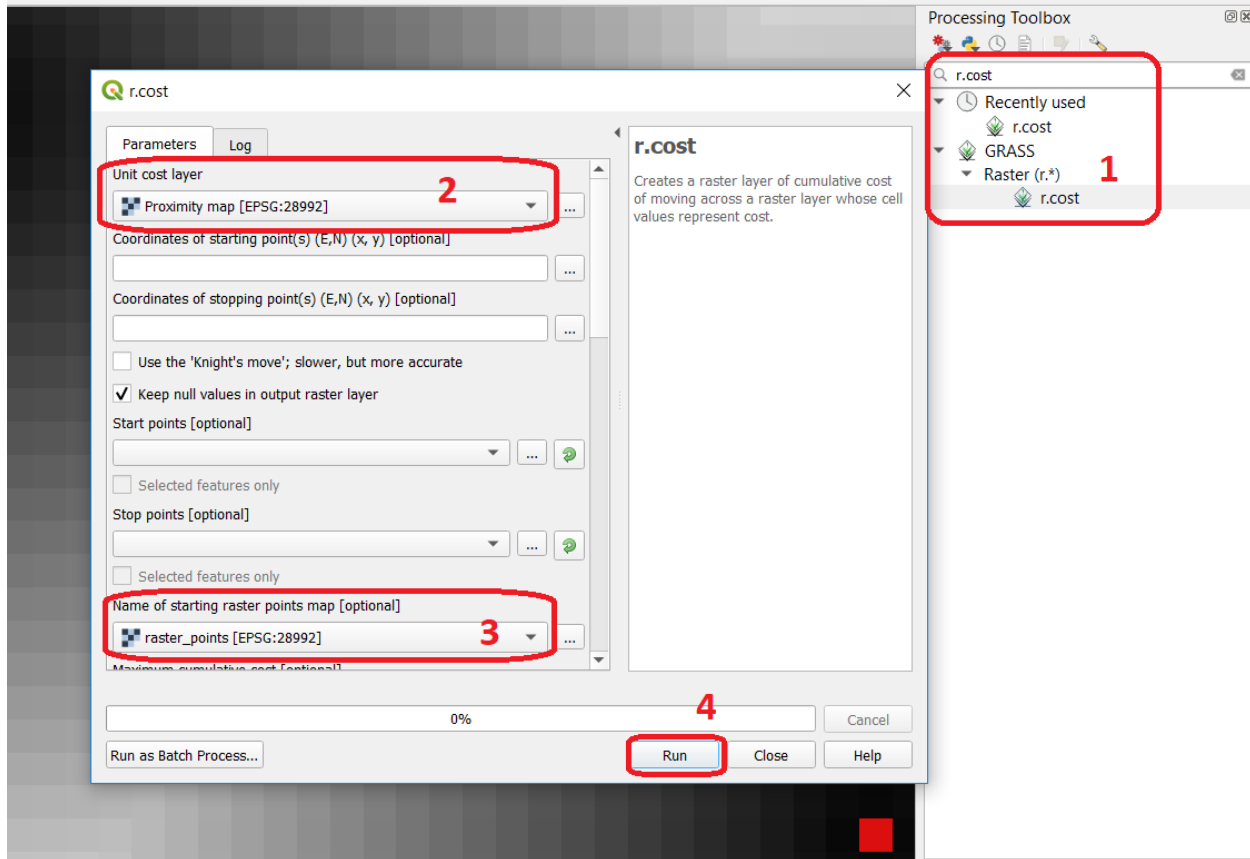


Fig. 8.9: Calculation of diffusion using the 'r.cost' tool

**Attention: Question.** Can you give some examples of applications that might use the computation of diffusion?

### 8.5.3 Flow Computation


Flow computation calculates the flow along the least-cost path for each cell. Contrary to diffusion, which computes the spread of some material in all directions, flow computation is suitable to calculate the path that water will take when flowing downhill. The procedure consists of two steps:

1. Calculation of the flow direction raster
2. Calculation of the accumulated flow

The input for flow computation is a continuous field (raster), e.g. a DEM. The computation of the **flow direction** goes as follows. See Fig. 8.10

For each cell in the input raster layer (e.g., cell 88), we determine the smallest direct neighbour (cell 74) and the smallest diagonal neighbour (cell 44). Then, we calculate the difference between the target cell and the neighbours, such as  $88 - 74 = 14\ m$  and  $88 - 44 = 44\ m$ . Then, we calculate the steepness of the neighbours. For this, we take into account the distance between the cell centres. If the resolution is

$10 \times 10 m$ , we can calculate the steepness as  $14/10 = 1.4$  for the direct neighbour, and as  $44/(10 * \sqrt{2}) = 3.11$  for the diagonal neighbour. Now, we know to which cell some material in the target cell (cell 88) will flow. *This is to cell 44 because it is the cell where the steepness is the highest.*

To compute the **flow accumulation**, we count for any given cell, how many other cells flow into it for the whole extent of the flow direction raster. For the target cell in Fig. 8.10, the flow accumulation is 7. Read a more detailed explanation on  [flow computation](#).

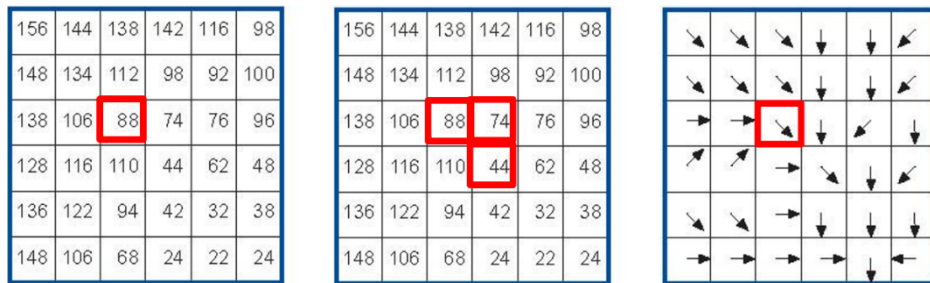
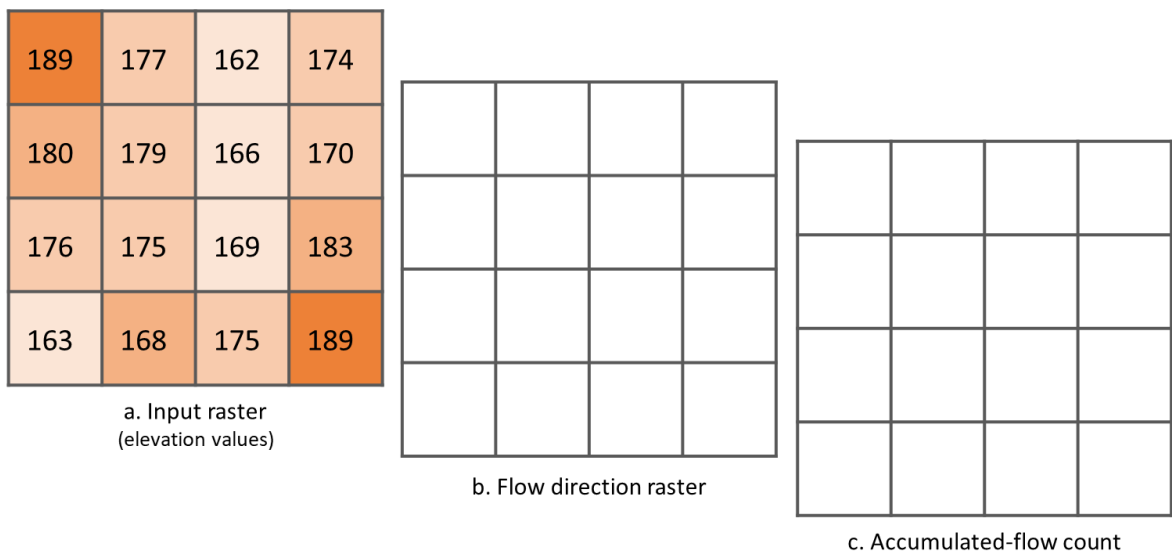



Fig. 8.10: An illustration of the flow computation in a DEM

**Task 13** Compute the flow direction and flow accumulation for the elevation raster below. Use a pencil and paper.



## 8.6 Surface Analysis

 **Surface Analysis** consists of computations such as Slope angle, Slope aspect, Hillshading, etc. A common factor among these computations is they require continuous input surfaces (e.g., elevation), and they can tell the user something about the change or shape of this surface.

**Task 14** Open the project `surface_analysis.qgis` and use your software to compute the *slope angle*, *slope aspect* and *hillshade* of the elevation raster 'mount\_etna\_dem'. We recently updated the dataset, *if you cannot find this file download the dataset raster-analysis.zip again*. Use the tools under **Raster terrain analysis** in the Processing toolbox, Fig. 8.11. Once you have the outputs, use the **Value Tool** to analyse the results.



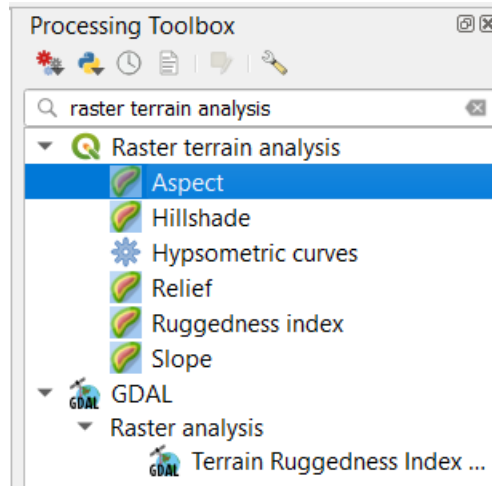


Fig. 8.11: The raster terrain analysis tools

*Section author: Ellen-Wien Augustijn, André da Silva Mano*



## DATA QUALITY




## 9.1 Accuracy

The main goal of this exercise is to provide you with the first-hand experience on how to identify and quantify errors in spatial data. We will also emphasise the common causes of the deterioration of the quality.

**Important: Resources.** You will require the latest LTR version of LTR version of QGIS, plus the dataset `data_quality.zip`. For this section, you will need the following files:

- `accuracy.qgs` – a QGIS project preloaded with the datasets described below;
  - `geodesic_marks.gpkg` – peaks signaled by geodetic marks;
  - `peaks.gpkg` – geomorphological features (peaks);
- `rmse.xls` – a spreadsheet to calculate the rmse.

### 9.1.1 Positional Accuracy


For the first part of this exercise, we will explore the concepts of  Accuracy,  Precision and bias. One of the ways of assessing the quality of a dataset is by determining how accurate the data is. Accuracy might be measured with regards to  Positional accuracy or attribute correctness, but in any case, we can only measure accuracy if we have a second data source of higher accuracy that we can regard as true or more close to the truth than the one we are assessing.

**Attention: Question.** How is accuracy different from precision?


In this exercise, the ‘*geodesic\_marks*’ dataset has higher accuracy than the ‘*peaks*’ dataset. These datasets are from Portugal. ‘**Geodesic\_marks**’ is a very high accuracy dataset from a topographic survey. ‘**Peaks**’ represents the altitude of mountain peaks extracted from a digital elevation model obtained using remote sensing – ASTER Global Digital Elevation Model.

**Note:** The ASTER Global Digital Elevation Model (ASTER GDEM) is a *product* developed and made available to the public by the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA). It is produced using data from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), a space-borne remote sensor.

The ASTER GDEM covers the Earth's surfaces between 83N and 83S. The ASTER GDEM surveys the Earth's surface and collects 22,600 tiles; each tile covers 1 per side. Only tiles that contain at least 0.01

A way of assessing the positional accuracy of a dataset is calculating the Root Means Square Error ( RMSE).

**Task 1** Open the project `accuracy.qgs`. The project has two layers: 'geodesic\_marks' and 'peaks'. Use the

**Identify Features**  to find the coordinates of two points that fall close to each other, Fig. 9.1. You can draw a small rectangle over the features you want, or you can click. Make sure that in the Identify window that will pop up, the option Mode is set to Top-down. You may need to resize or rearrange your panels to make room for the **Identify Results** window.

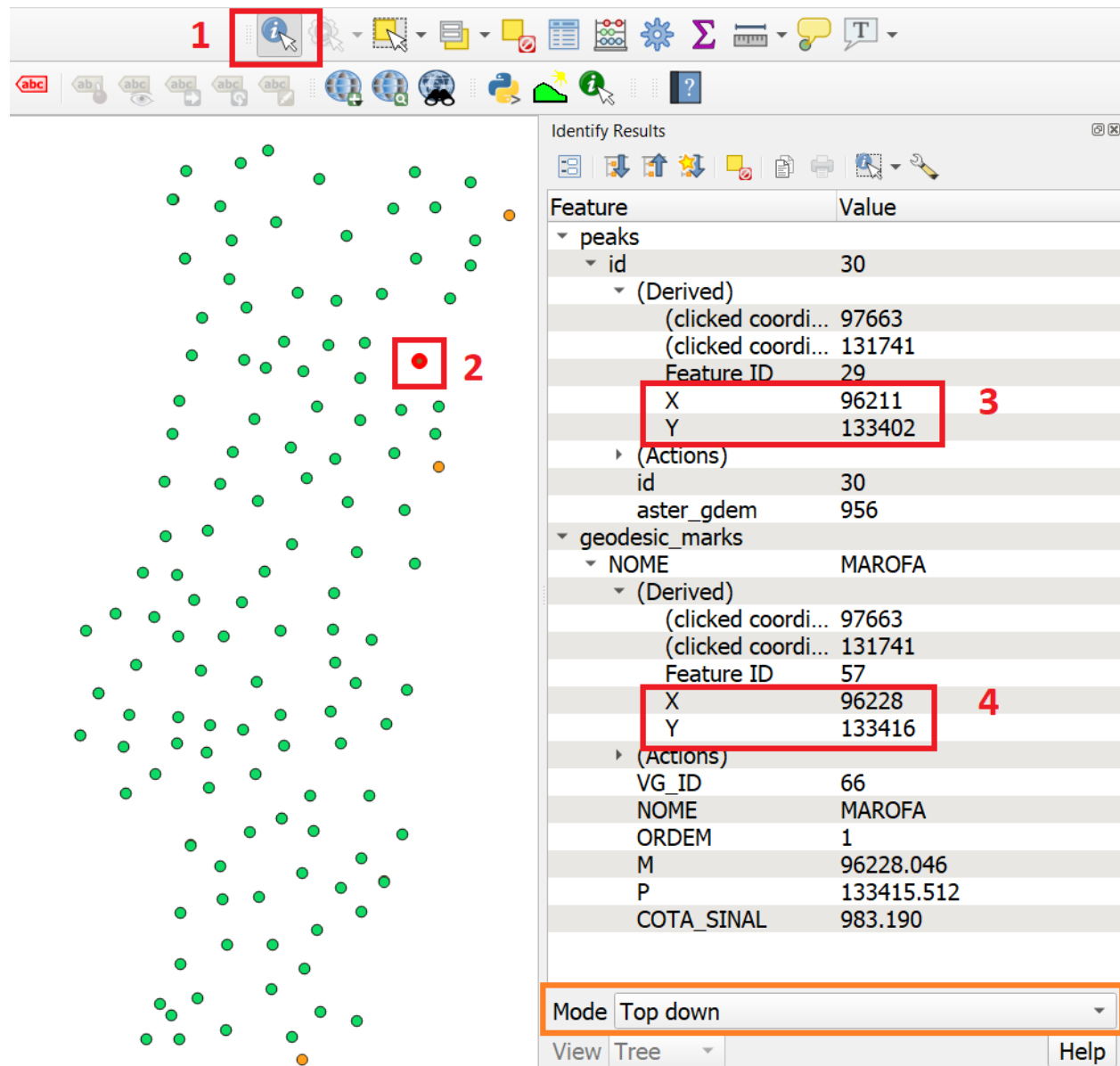


Fig. 9.1: Use of the identity tool to find coordinates in a point layer

**Task 2** Copy the coordinates from the Identify Results dialog and paste them in the spreadsheet `rmse.xls` Fig. 9.2. You will need Microsoft Excel or Libre Office to open the file. Remember that true values are the ones associated

with the *'geodesic\_marks'* layer and that the measured values are the ones associated with the *'peaks'* layer. Repeat this process for more points (minimum 4, maximum 8). The RMSE will be computed automatically at the bottom of the table.


Vertex	X (true)	X (measured)	dX true-meas	dX <sup>2</sup>	Y (true)	Y (measured)	dY true-meas	dY <sup>2</sup>
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
			0	0			0	0
		mx=	0			my=	0	
		RMSE=	0					

Fig. 9.2: Spreadsheet for calculating the RMSE

**Attention: Question.**


- What does the RMSE that you obtained represents?
- Is the error systematic?
- Is the error acceptable?

### 9.1.2 Attribute Accuracy

Accuracy doesn't refer solely to positioning; it can also refer to attributes, in which case we call it  **Attribute accuracy**. The two layers in your project contain altitudes of geomorphological features, i. e. *'peaks'*. The true values when it comes to attributes are also stored in the *'geodesic\_marks'* layer.

**Task 3** Using the Identify Tool like in the previous task. Make a quick comparison of the difference between the altitudes stored in the attribute **COTA\_SINAL** of the *'geodesic\_marks'* layer, and the altitudes in the attribute **aster\_gdem** of the *'peaks'* layer.

**Attention: Question.** Could you use RMSE to assess attribute accuracy? If yes, what would the error you obtain will represent?

Note that you can also measure the attribute accuracy of qualitative attributes, such as land cover. In that case, we use an  **Error matrix**. We will be discussed in more detail in the Course on Earth Observation.

### 9.1.3 Biased Data

Another possible problem with your data is **bias**. Biased data is systematically inaccurate data, usually for technical reasons. A classic example of biased data is the case of features that were digitised over a badly georeferenced base map. *Because the digitising process comes after georeferencing; any displacements introduced during the georeferencing propagate to all the features that are digitised.*

**Important: Resources.** You will require the latest LTR version of LTR version of **QGIS**, plus the dataset **data\_quality.zip**. For this section, you will need the following files:

- **bias.qgs** – a QGIS project preloaded with the datasets described below;
  - **road.gpkg** – a vector dataset representing a road;
  - **aerial\_photo.tif**;

**Task 4** Open the project **bias.qgs** and use the satellite image to digitise the highway that crosses the image. Refer to the section **Digitising** if need to. When you are done with digitising, turn on the ‘road’ layer and compare the results.


**Attention: Question.** You will see a significant shift between the road you just digitised and the in the road layer. Can you explain why? Use the **Quick Map services plugin** to add additional satellite imagery to help you find an answer.

Bias can also have political reasons. Maps can be statements of power and there are countless examples of using maps to project power or to lobby for a hidden agenda. Although that is not a typical case a GIS professional is exposed to, it is still worth mentioning. A classic example of political bias are the maps from the Age of Discoveries (XV to XVII centuries A.D.). In this XVI century map, **Fig. 9.3**, Asia is represented as being a possession of Portugal, but that was never the case. The Portuguese Empire in Asia consisted of commercial outposts protected by fortresses and a war navy, but at no moment did the Portuguese had control over the vast inland areas.



Fig. 9.3: Anonymous map c.1550 of Eastern Africa, Asia and Western Oceania

## 9.2 Data Acquisition Errors


A common source of errors occurs during the spatial acquisition phase. Errors originating in this phase  propagate if not corrected immediately.

---

**Important: Resources.** You will require the latest LTR version of QGIS, plus the dataset `data_quality.zip`. For this section, you will need the following files:

- **acquisition\_errors.qgs** – a QGIS project preloaded with the datasets described below;
    - `porto_buildings.gpkg` – the footprint of urban buildings from Porto (Portugal);
    - `porto.tif` – a raster base layer covering the streets of Porto.
- 

### 9.2.1 Conversion Operations Leading to Quality Loss

Conversion can have different meanings in the GIS world. In the scope of this exercise **conversion** refers to format conversion (raster to vector and vice-versa). Vector and raster data are usually associated with the representation of discrete and continuous phenomena, respectively. Having to convert between formats is a common operation in the GIS world, however, these operations always imply some type of compromise regarding what we get and what we lose, especially in the case  **Rasterisation**.

**Task 5** Open the project `acquisition_errors.qgs`. From the Processing Toolbox, use the **Rasterise** tool to convert the *'porto\_buildings'* layer to a raster using a resolution of 15 m. [Fig. 9.4](#)



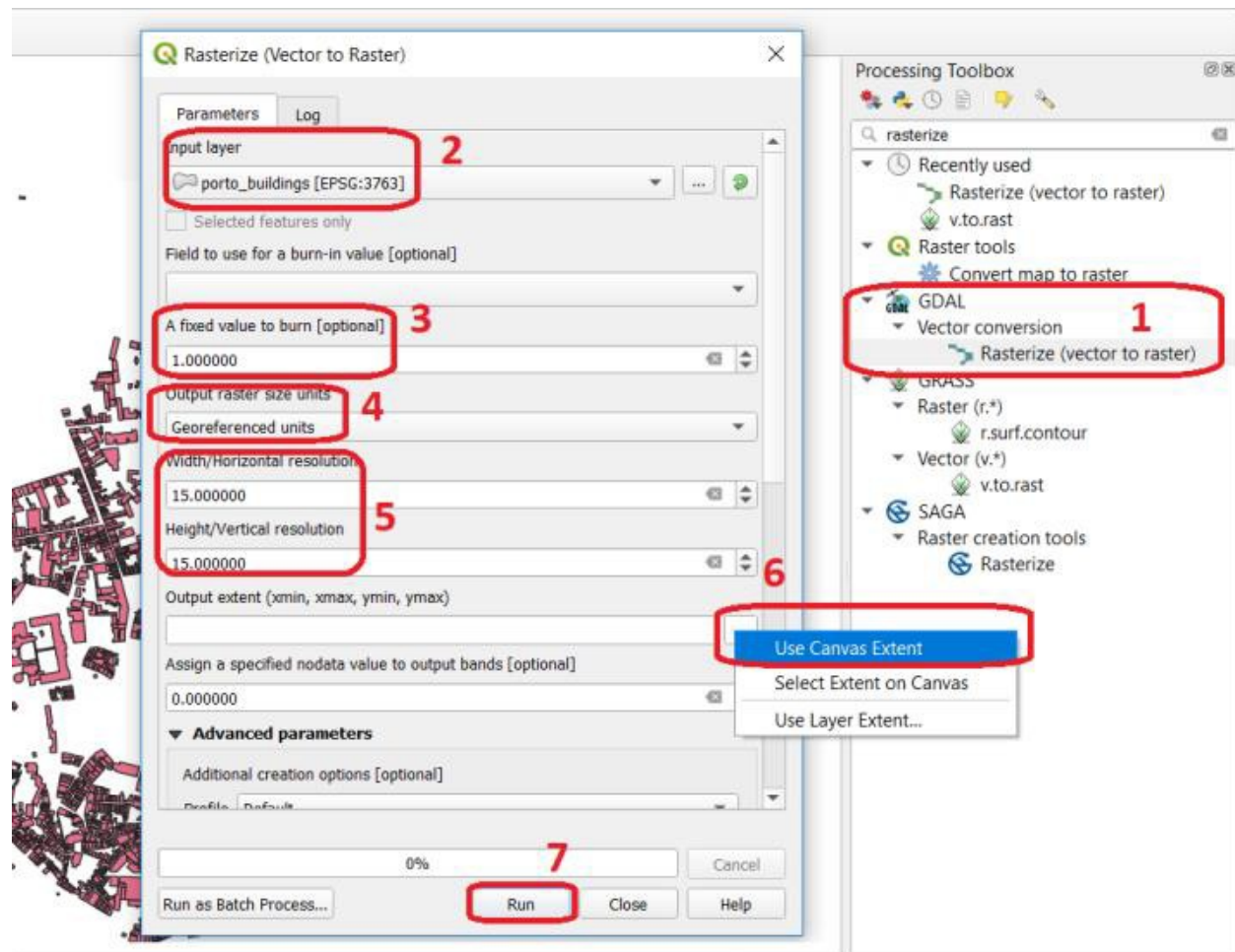


Fig. 9.4: Using the rasterisation tool

The result is not very impressive. Besides, we lost the attribute table and with a 15 m resolution, we also lost information on the shape of the buildings.

**Attention: Question.** Convert the '*porto\_buildings*' layer to a raster once more; this time use a resolution of 2 m.

- Does the 2 m resolution raster has enough quality?
- What do you lose and what do you gain when you rasterise?
- Does the 2 m resolution raster preserves all the original vector features, i.e. at least one pixel per building? How can you tell?

Converting from vector to raster usually means to give up of your semantic information, the attribute table. Nevertheless, having an attribute table associated with a raster is possible with certain software products. Still, in such cases, the attribute table can only be used from within that specific software product. In other words, it is software dependent.

A compromised solution to preserve attribute information is to create a raster layer using the numeric values of a single attribute in the vector layer. In such a case, attribute values are converted to pixel or cell values in the resulting raster layer. For example, in a vector layer representing land cover, usually, every type of land cover has a unique numeric code (e.g. *forest* = 100, *road* = 400, *water* = 700). If you set an attribute when converting to a raster layer, cells inherit the values from that attribute. Then, the cell values have an implicit semantic meaning; for this example, that



is the type of land cover.

**Task 6** Repeat the vector to raster conversion for the ‘porto\_buildings’ layer. This time feed the attribute fid into the option **Field to use for a burn-in value**. [Fig. 9.5](#)

Field to use for a burn-in value [optional]

123 admin\_code

Fig. 9.5: Rasterisation with burn-in values

**Note: Reflection.** As a result of the previous task, you will get a map like the one shown in [Fig. 9.6](#). Make sure you understand the map and how it relates to the ‘porto\_builfings’ vector dataset.




Fig. 9.6: The ‘porto\_buildings’ layer rasterised using burn-in values

## 9.2.2 Digitising Operations Leading to Errors

A common cause of errors has its origin in data acquisition through digitisation. When digitising features from a base map the resulting accuracy is related to the scale that is being used at the moment of digitising, this is related to the concept of [Perkal band](#). The basic principle is that the smaller the scale, the less accurate the digitalisation will be.




**Task 7** Digitise some streets at scale 1:15000 using the “porto.tif” layer as the base map. If you zoom in to your streets, you would realise that not only they tend to shift from the underlying reference map roads (i.e. meaning a large perkal band). They are not topologically connected.

**Attention: Question.** Read about the  [Data checks and repairs](#). What are the clean-up operations that you would need to perform on the streets layer you created in the previous task?.

You would avoid many problems with these errors, if you properly set your **snapping** and **topological edition** options, as discussed in Data Entry. But before start working with a new dataset, you should always check its quality.

---

## 9.3 Error Propagation

When  combining spatial data from multiple sources, and especially when  merging two datasets or more layers, errors like sliver polygons may occur due to  differences in accuracy. Sliver polygons are very narrow polygons that usually appear along the border of other polygons. They do not necessarily represent a problem. They may be legitimate features resulting from a geoprocessing routine, or they can be errors exposing accuracy issues.

---

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset [data\\_quality.zip](#). For this section, you will need the following files:

- `error_propagation.qgs` – a QGIS project preloaded with the datasets described below;
    - `porto_buildings.gpkg` – footprint of urban buildings from Porto (Portugal);
    - `porto_landuse.gpkg` – landuse (meadows) of Porto;
    - `porto_roads.gpkg` – road network of Porto;
- 

### 9.3.1 Errors Generated by Geoprocessing Routines

The project `error_propagation.qgs` has a layer named ‘*porto\_landuse*’ that contains areas classified as ‘**meadows**’; and another layer named ‘*porto\_buildings*’ We want to merge these two layers so that:

1. We obtain a layer whose polygons represent either buildings or meadows, and
2. Look for inconsistencies – areas that are both building and a meadow.

**Task 8** Open the project `error_propagation.qgis`. From the *Processing Toolbox*, use the *Union tool* to create a new layer which merges the ‘*porto\_buildings*’ and ‘*porto\_landuse*’ layers. [Fig. 9.7](#).

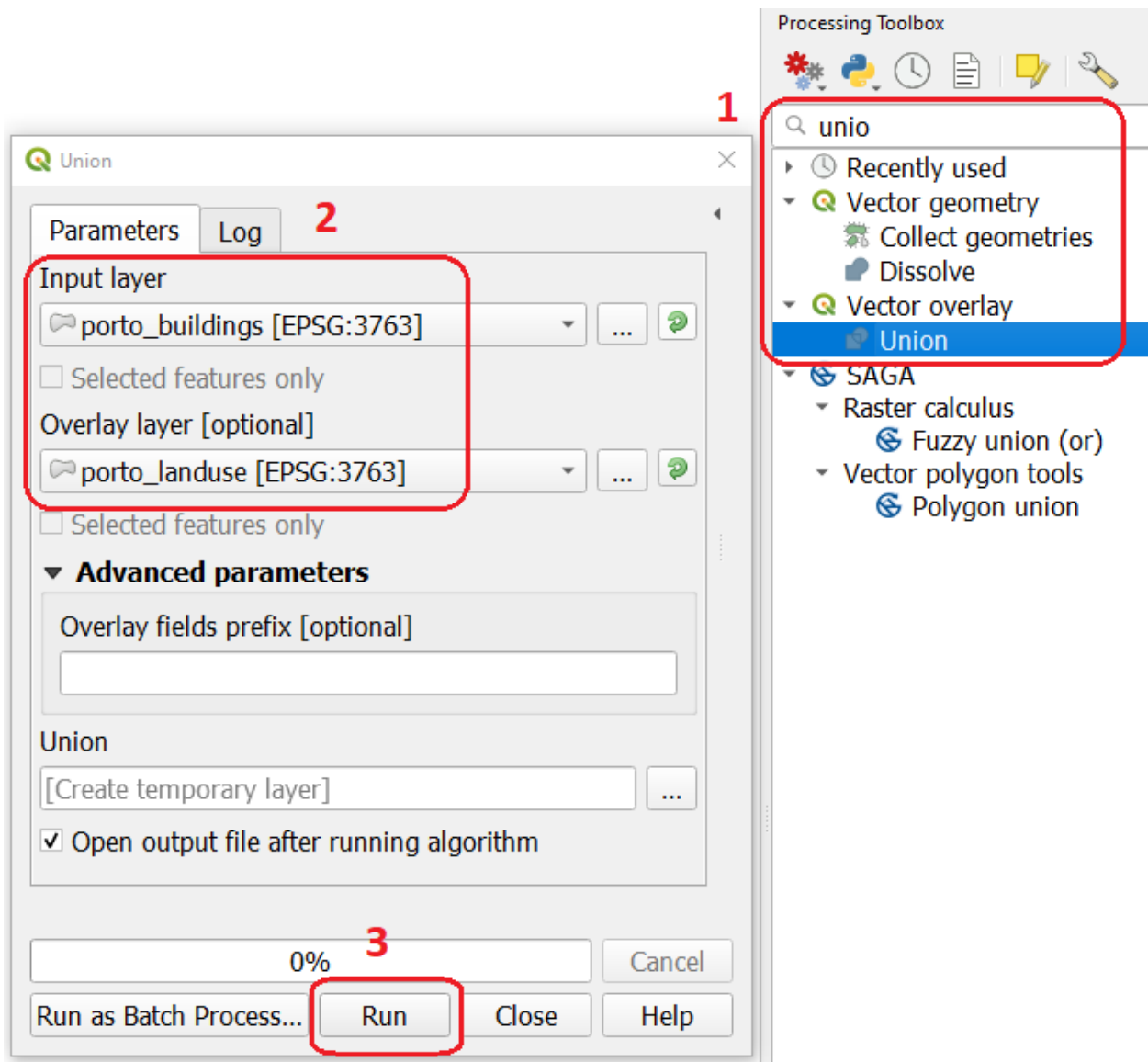


Fig. 9.7: The Union Tool

Now, we could filter the union-layer to find out if any of the resulting polygons are classified as both 'meadow' and 'building'. But, before that we will focus on finding **sliver polygons**.

**Attention: Question.**

- Can you think of a filter that identifies silver polygons in the union-layer?
- Do you think you could also find silver polygons before the union operation? How?

**Task 9** Sliver polygons are usually very small. In the layer that resulted from the union operation; go the layer *Properties.. > Fields > Field Calculator* . Add a new field to compute the area of the polygons using the **Field Calculator** Fig. 9.8.

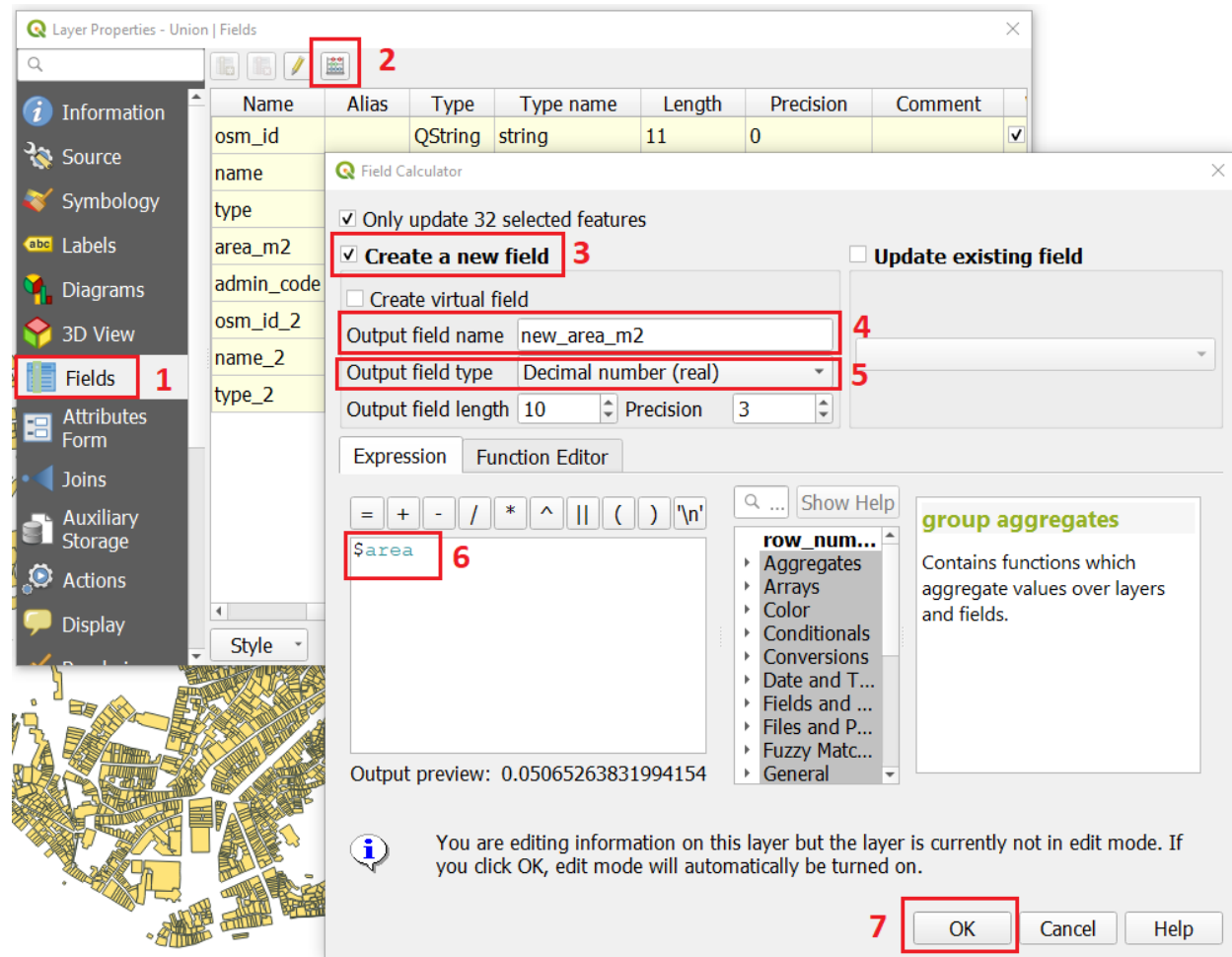



Fig. 9.8: Adding a field to the attribute table using the field calculator

**Note: QGIS.** When you press *OK* (step 7), QGIS will automatically enter into edit mode to commit the changes into the attribute table of the vector layer. You can toggle off editing mode using the button . You will be asked to save the edits - make sure you do.

**Task 10** Using the field you just created. Filter the polygons whose area is less than  $x$ ; where  $x$  can be any value as long it is suitable for a particular case. For this case use  $2\text{ m}^2$ , polygons smaller than this are potentially sliver polygons. Refer to section [Selection by Attributes](#) if you need to. Once the filter is applied, browse the attribute table and zoom to specific features to take a close look at the sliver polygons. [Fig. 9.9](#)

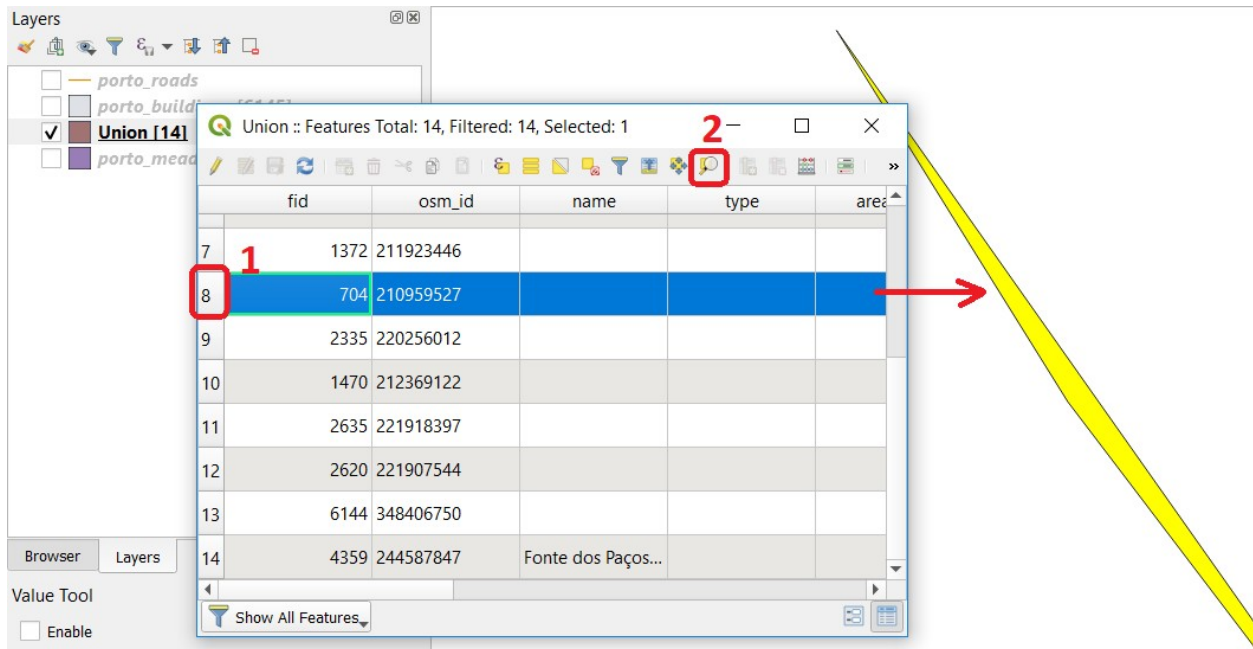


Fig. 9.9: Browsing the filtered attribute table to find sliver polygons

In this case, you would probably proceed to delete those polygons and clean the dataset. But keep in mind that knowing if a sliver polygon is an error or not is, it is a tricky question because it depends on what your data is supposed to represent. Having very small polygons can simply represent a very small feature! Therefore depending on the nature and scale of the data, a particular sliver might be an error or not. In cadastral datasets, for example, sliver polygons are most likely errors.

**Attention: Question.**

- What problems might occur if our data has sliver polygons?
- Can you think of other methods to detect sliver polygons?

### 9.3.2 Conceptual Errors

Conceptual errors are errors that arise not from acquisition or processing but from logically erroneous thinking. These type of errors are associated with the way reality is being modelled, and they have a direct relation with Logical consistency and Completeness of datasets and analytical approaches.

**Example.** You are making a map to support biologists who want to study a particular species of amphibious that thrives on any small pond. With that in mind, you produced a raster map of water ponds of the study area using a cell size of 20 m; i.e. that will mean that the minimum size for an object to be properly represented in your dataset is 400 m<sup>2</sup>.

However, when you show the map to the biologists, they remind you that this species has been found in ponds much smaller than 400 m<sup>2</sup>. *If you insist on using such raster dataset for further analysis, the final result may not be valid for the biologists.*

To explore the idea of conceptual errors, we will use a simple example that applies network analysis. In this example, we want to estimate the fastest route from point A to point B.

**Task 11** From the Processing Toolbox, use the **Shortest path (Point to Point)** tool to calculate the fastest route between any origin-destination over the *porto\_roads* layer. Fig. 9.10.

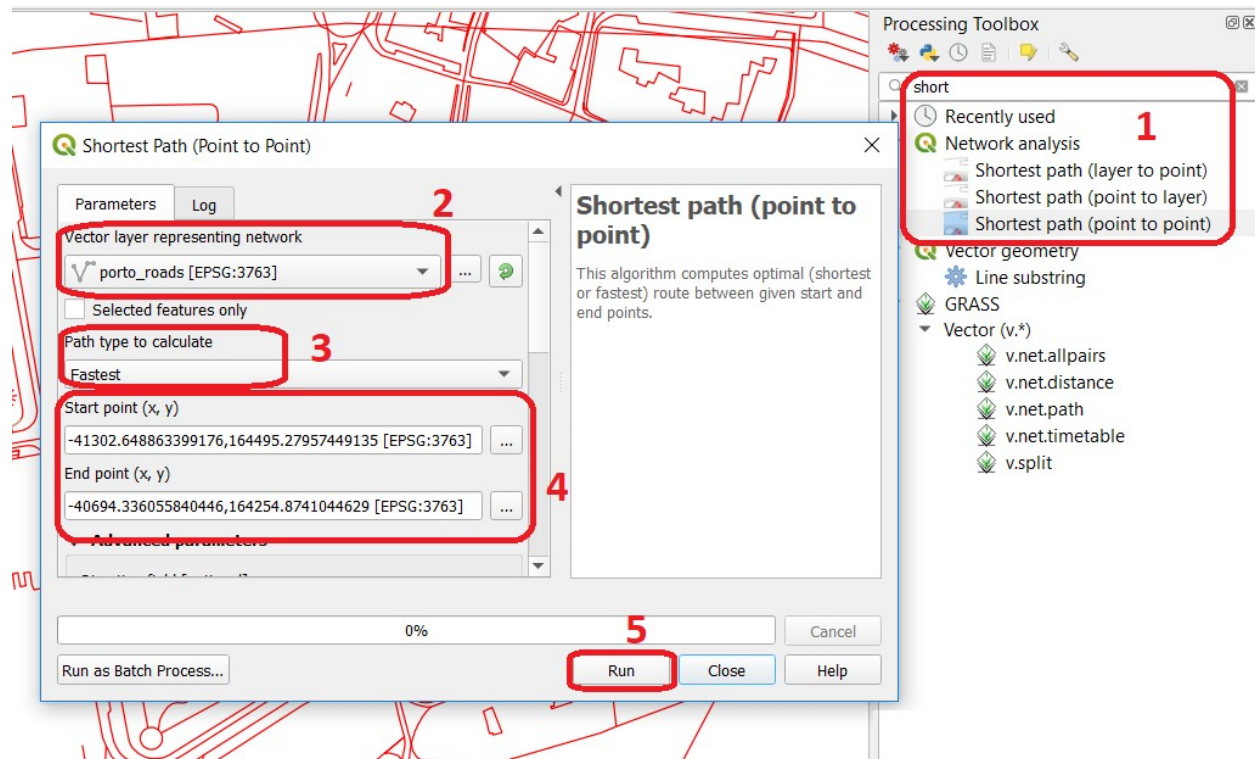


Fig. 9.10: Using the shortest path tool for network analysis

#### Attention: Question.

- This analysis of the previous task has some conceptual problems that make it very prone to errors. Can you identify some of them?
- **Uncertainty** is related to the **Sources of errors**. Can you describe how the sources of errors influence uncertainty?

## 9.4 Data Management and Metadata

GIS projects can require massive amounts of data. As the project complexity increases the effort to manage the necessary data starts to need more time.

We will not cover this topic entirely here. Still, we will state some simple guidelines to improve data and project management. In particular, we will focus on:

1. Workspace organisation
2. Naming conventions
3. Metadata

It is essential to organise your input and output data in such a way that you can easily continue your work later. This will save you a lot of time, and it will prevent confusions when working with tens of datasets.

---

**Important: Resources.** You will require the dataset [data\\_quality.zip](#).


---

**Task 12** Take a look at the contents of the `data_quality.zip`. Put attention to the file structure and naming conventions.

- Are the files named consistently?
- Is the folder structure adequately organised?
- Is it easy to distinguish between different types of data?
- Are the datasets stored using well-known data formats or standards?
- Would you have organised the data for this exercise differently? How?

**Task 13** As you moved along the exercises you might have produced some output files, some of them have experimental results, others have definitive results.

- Have you saved them in a proper and accessible folder?
- Can you trace back the changes you made during every processing attempt from the file names?

A relevant topic for data quality is **Metadata**, often defined as “*data about the data*”. Metadata provides additional or auxiliary information about a geospatial dataset that may not be easily obtainable from the dataset itself. For example, sometimes you can read the coordinates of a geometry in a vector data set. Still, you may not be able to determine the spatial reference system from the coordinate values alone. Ideally, metadata includes details related to the [Lineage](#) and  [Completeness](#) of the data. There are standards and directives for metadata of spatial data. In Europe, an important directive is the [INSPIRE](#) initiative.

From a practical point of view, it is not easy to maintain all necessary metadata, mainly if you aim at being 100% compliant with a particular standard or recommendation. Additionally, each software package has its metadata management system, which makes the interoperability of metadata difficult.

This is not the same as saying that we should not care about metadata. However, we can be pragmatic and consider at least the following principles:

1. Save metadata in a format that is not software-dependent (a simple *Readme* file is often enough).
2. Include the lineage of the data:
  - how was the data aggregated?
  - Where to find the original data?
  - What geoprocessing operations were performed?
3. Include contact information of the author or owner of the dataset, if you expect others to use the dataset.
4. Historical information. Is the dataset representing a geographic phenomenon in 2013 or in 2020?
5. Licensing: copyrights and restriction on using the data.

**Attention: Question.**



- How can metadata be useful?
- Would you add something else to the list of principles above?

*Section author: André da Silva Mano, Ellen-Wien Augustijn & Manuel Garcia Alvarez*



## VISUALISATION

### 10.1 Cartographic Grammar

The cartographic grammar defines a set of ‘rules of thumb’ for the  Visualization of spatial data and the correct representation of spatial and temporal information on a  Map. Cartographic grammar ensures that maps effectively communicate information to the map user.

For a more detailed explanation on Cartographic Grammar, visit <https://kartoweb.itc.nl/TMT> The **Thematic Map Tutor** will help you to understand which visual variable to use depending on the type of data you want to visualise. This will help you in completing the tasks below.

**Attention: Question.** Assume you were asked to make a map using following data:

*No. of physiotherapists per 100.000 inhabitants per province (2010):*

<b>Groningen</b> 3,5	<b>Friesland</b> 3,2	<b>Drenthe</b> 3,0	<b>Overijssel</b> 2,9	<b>Flevoland</b> 4,9	<b>Gelderland</b> 2,8	<b>Utrecht</b> 2,5
<b>Noord-Hol- land</b> 2,6	<b>Zuid-Hol- land</b> 2,9	<b>Zeeland</b> 2,8	<b>Noord-Bra- bant</b> 2,7	<b>Limburg</b> 2,9		

1. What is the measurement level of the data?
2. Which visual variable(s) is (are) suitable to depict this data?
3. Which type of thematic map would you use to depict this data?

**Task 1** Suppose you have to design symbols for a small scale (1 : 100.000) geological map. The following rock types have to be represented on the map:

- igneous rock
- metamorphic rock
- consolidated sedimentary rock
- unconsolidated sedimentary rock

You also have to map the terrain morphology, according to the following classification:

- relatively flat terrain (slopes 0-7%)
- rolling to hilly terrain (slopes 7-30%)
- steep, rugged terrain (slopes > 30%)

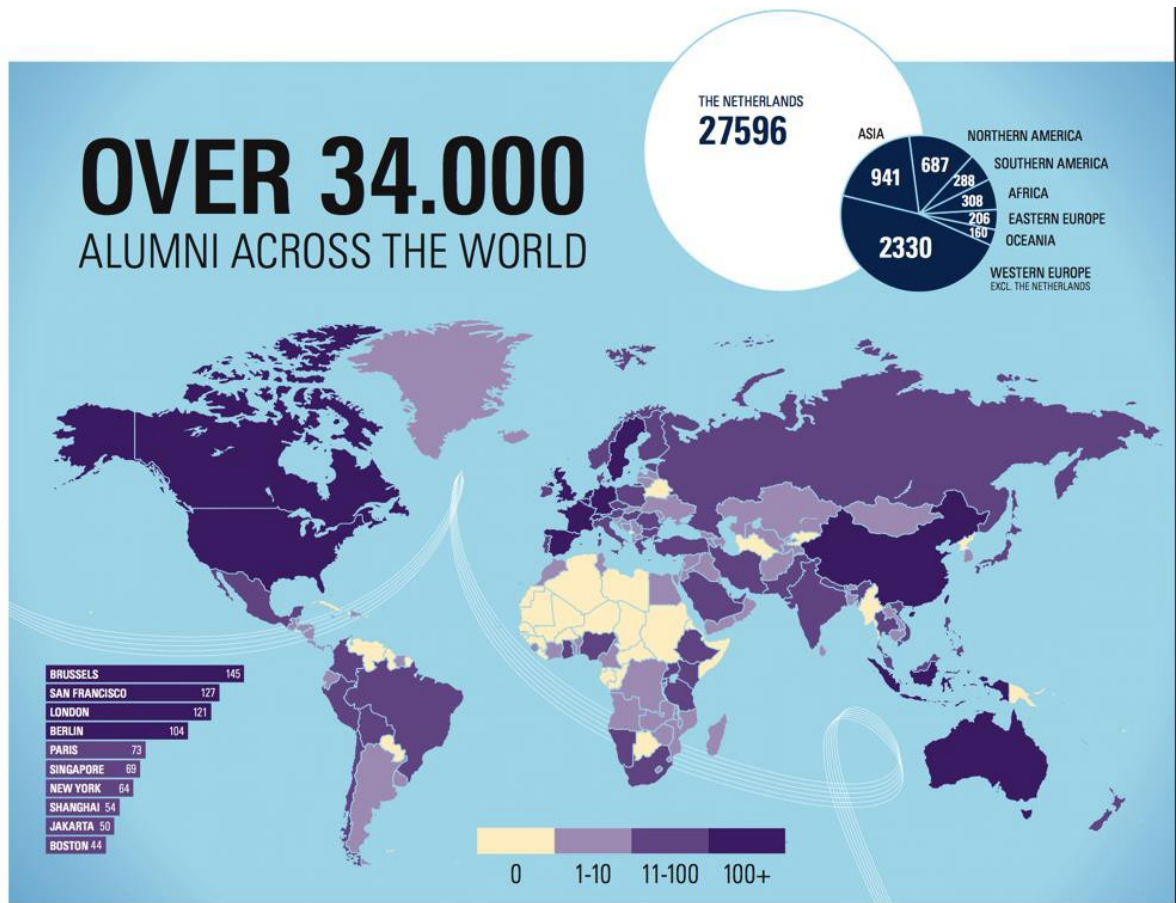
All possible combinations of rock type and terrain classes may occur. Describe how you would set up this map: *you do not have to draw the map or legend symbols, instead, describe:*

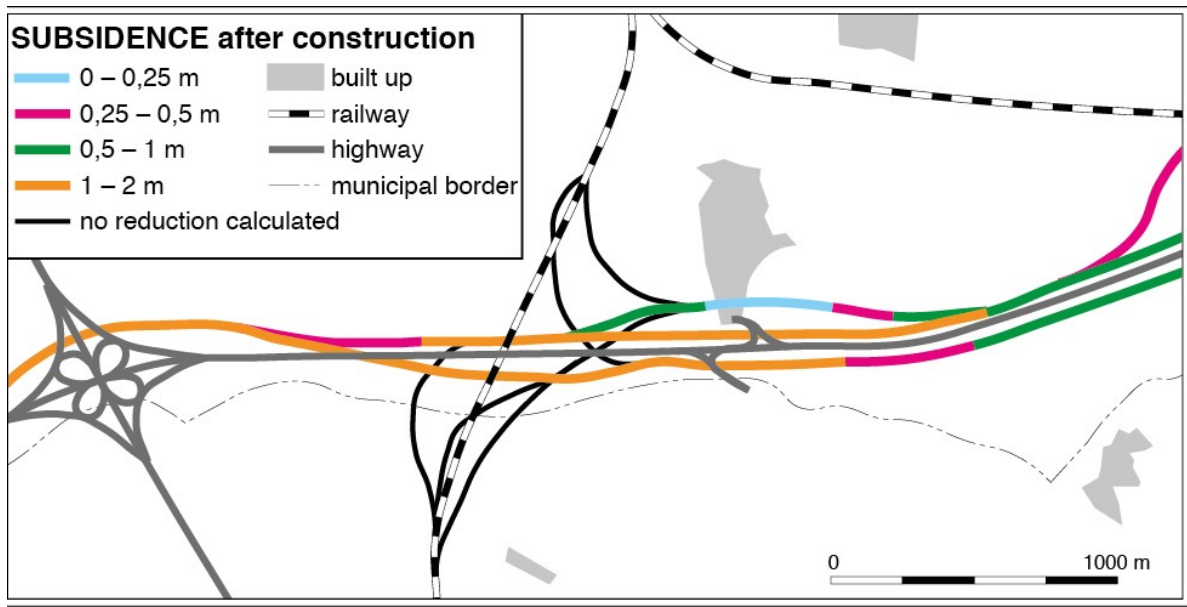
- which layer(s) of information you would put in the map and what visual variable(s) should be used to combine those layer(s) in a single map,
- which legend titles and categories you would use, and
- which type of map(s) to represent the data.

## 10.2 Critical Map Viewing

With the knowledge, you gathered in the previous section on Cartographic Grammar; you can critically look at existing maps and identify their shortcomings.

**Task 2** Look carefully at the two maps below and use what you have learned about depicting quantitative data to analyse them critically.



**Attention: Question.**

- What are some of the shortcomings these two maps above?
- How will you solve such shortcomings? Give some examples.

**Task 3** Before the next session in the Virtual Classroom, find a map (in the newspapers, on the web, in a research paper or a book) that you find interesting. You may select a map because you think something is wrong with it, but also because you particularly like it or think it is a well-made map.

**Attention: Question.**

- What is interesting about the map you find in the previous task?
- Can you give a summary of its flaws and strengths?

## 10.3 Creating a Topographic Map

In this section, you will create a small scale topographic map of a part of Switzerland, including the depiction of the 3D terrain using hill shading and layer tints.

Here, you will follow the “search, try & learn” principle. That means that we do not show you how to do things. Instead, we ask you to accomplish specific tasks and provide you with relevant information that will help you to achieve the tasks. The software skills necessary for this exercise were covered in previous sections.

In this exercise, you have to focus on the production of a cartographically well-designed map. It will be your responsibility to study the task carefully and search through the reference information for ideas and answers. With that knowledge, you have to decide which strategy to use to complete the task.

**Important: Resources.** You will require the latest LTR version of [QGIS](#), plus the dataset [visualization.zip](#).

### 10.3.1 Reference Material

- Earlier exercises, using QGIS;
- The [QGIS training manual](#) and the [QGIS user guide](#).
- The texts on Cartographic Visualisation in the Living Textbook;
- The appendix on *Shaded Terrain Models*, which contains quick introductions to creating a shaded terrain model and other useful tools.

### 10.3.2 Understanding the Data

This is a detailed description of the datasets for creating a topographic map in this section.

**Raster data** `dem_90m.tif` This is part of a *Digital Elevation Model*, produced by NASA from the *Shuttle Radar Topography Mission* (SRTM). During an 11-day mission in February of 2000, NASA obtained elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of Earth. SRTM consisted of a specially modified radar system that flew onboard the *Space Shuttle Endeavour*. The data is freely available at <http://www.jpl.nasa.gov/srtm>.

Each cell has a value that represents the height in meters. The SRTM data was originally stored using geographical coordinates. However, this version is in meter coordinates in **UTM zone 32N** on the **WGS84 datum** (EPSG: 32632).

**Vector data** All vector data was derived from the “EuroGlobalMap”, a Pan-European Database at Small Scale. It was an initiative from *EuroGeographics*, a cooperation of all European topographic services. The vector layers provided here are from version 7.0 (September 2013). The EGM Database is intended to be used in map scales of about 1 : 1.000.000. The EGM data was initially stored using geographical coordinates. However, this version is in meter coordinates in **UTM zone 32N** on the **WGS84 datum** (EPSG code 32632). EGM is open, and it can be downloaded from <http://www.eurogeographics.org/products-and-services/euroglobalmap>.

For this exercise, We have removed many of the attributes in the original datasets, and kept only small selection. A description of the attributes per dataset is given below:

- `builUpArea.gpkg`: NAMN1 = Name in German;
- `lakes.gpkg`: NAMN1 = Name in German;
- `watercourse.gpkg`: NAMN1 = Name in German;
- `railways.gpkg`: TYPE: 31 = secondary, 33 = primary; TUNNEL: 0 = not in tunnel; 1 = in tunnel
- `ElevP.gpkg`: Most important mountain tops and passes. Attributes NAMN1 = Name in German; ZV2 = height in meter above sea level;
- `towns.gpkg`: NAMN1 = Name in German; PPL = population (in 2013);

The road data from EGM is notoriously incomplete and too general for the scale of the map that you will make. Therefore, we included data from the *OpenStreetMap* database. We extracted the road data for the categories that depict main roads. The OSM data was originally stored using geographical coordinates. However, this version is in meter coordinates in **UTM zone 32N** on the **WGS84 datum** (EPSG: 32632). Only a small selection of the original attributes of OSM was kept in this version. Those are:

`osm_roads.gpkg`: **osm\_id**: unique id of each segment; **type**: motorway, primary, secondary, or trunk; **tunnel**: 0 = not in tunnel; 1 = in tunnel.

*Note that the OSM data is very detailed. It is up to you to decide if you need all categories, or if it is better to delete or not show some of them. This will depend on the requirement of your user, and the choices of symbology that you make of this and the other data layers.*

### 10.3.3 Map Making

**Task 4** Open the QGIS project `topographic_map.qgs`. It contains all the layers you will need. Make a topographic map of this part of Switzerland (the “Berner Oberland”, highlands of Kanton Bern), that adheres to the following requirements:

1. The map shall present data in the information categories mentioned below. The visualisation shall be correct for the type of data, and it shall also be tailored for the specific combination. The required information categories are:
  - a. **Terrain.** The shape of the terrain shall be visualised using hill-shading in conjunction with layer tints. Consult the appendix *Shaded Terrain Models* to know how to create such a model. Refer to the theory in the Living Textbook and lecture slides to find examples of how to achieve such depiction. Give priority to the design of a sensible and readable visualisation. Something that gives the user a good impression of the shape of the mountains, in the country.
  - b. **Infrastructure: roads and railways.** The map shall show the number for the important roads.
  - c. **Cities and towns.** Place names shall be included for at least the larger cities.
  - d. **Hydrography: lakes and rivers.** The most important rivers and lakes shall show their names.
  - e. Optionally, the map shall include additional data that you gather from other sources (e.g., the Internet, atlases, and other). Useful additional information might be the *names of mountain tops*, the *famous tourist sites*, and others.
2. The map shall fit on an *A4 landscape paper* sheet. The outer bounds of the map shall be rectangular and match the extent of the DEM. The projection of the data shall be Universal Transverse Mercator (**UTM**) **Zone 32N** on the **WGS84** datum. The data is already provided in that projection.
3. The map shall contain all necessary marginal information, such as title, legend, scale bar, etc.
4. The map shall be created for colour printing. The resulting map shall be exported as a PDF file. Use the **Print Composer** of QGIS to achieve this.

---

**Important:** This is a complex task! Please do not be satisfied too easily. Make tests of your results so far and study them critically. Ask family and friends, supervisors and your fellow students to give their opinion.






---

**Attention: Question.** Examining the topographic map you created in the previous task. What problems did you encounter during the map-making process?

*Section author: Barend Köbben, André da Silva Mano & Manuel Garcia Alvarez*



## RADIOMETRIC OPERATIONS

A  **Radiometric operation** uses or changes the values the pixels in an image. In this section, you will apply some of the most common radiometric operations in image processing. But first, you need to get familiar with how sensors capture  **Electromagnetic radiation** and how we can  **Display an image** based on the  **Tri-stimuli theory**. And how we represent colour on paper using the principles of the  **Subtractive colour system**.

### 11.1 Image Display

---

**Important: Resources.** You will require the latest LTR version of LTR version of **QGIS**, plus the dataset **data\_histogram\_operations.zip**. When you unzip the dataset, you will find the following files inside:

- **SPOT270611.img** – SPOT 5 image with 4 spectral bands.
  - **tm24aug99.img** – TM image of the 24th of august 1999.
  - **tm24aug99\_sub.tif** – a subset of **tm24aug99.img**
  - **topo34f.img** – scan of a topographic map in RGB.
- 

We can **enhance** the visualisation of images through well-known plugins and providers available in QGIS. In this exercise, we will make use of some of the QGIS tools. *Note that none of these tools changes the actual values stored in the raster datasets.* They simply change the way the image is displayed to highlight features that are not obvious when using the default visualisation settings.

**Task 1** Disable the default contrast stretch. Go to *Settings > Options > Rendering* tab and scroll down to Contrast enhancement.

Set the algorithms *Single band grey*, *Multiband colour (byte/band)* and *Multiband colour (> byte/band)* to **No Stretch**. Set the *Cumulative pixel count cut* to 2.0 and 98.0%. Click *OK* to confirm. See Figure [Fig. 11.1](#)

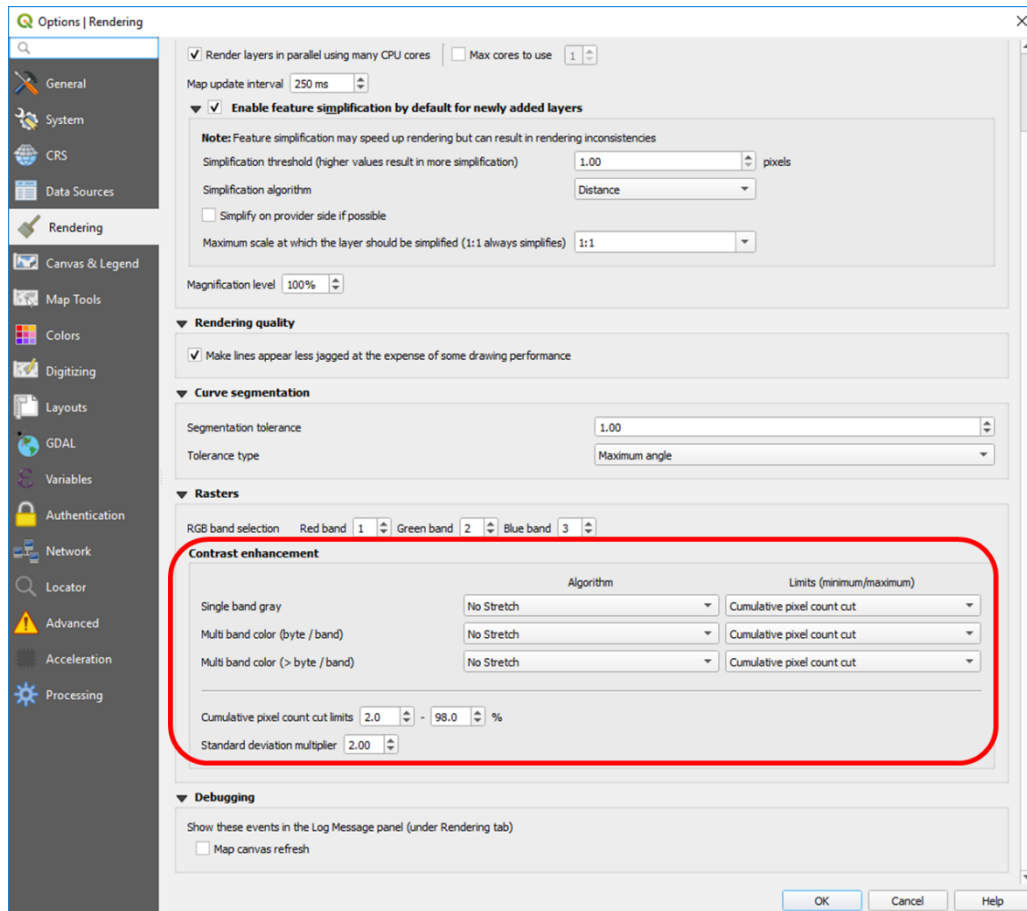


Fig. 11.1: Changing the default settings of contrast enhancement to 'No Stretch'

### 11.1.1 Single Band Display & Relative Brightness

In general, remote sensing images can be displayed using **Pseudo colour** and **Colour composite**. A common colour composite is the so-called **True Colour**. Single-band display uses **pseudo colour**.

**Task 2** Create a new QGIS project and open the `topo34f.img`. If required, change the colour composite for this layer such that Hydrographic elements like water bodies display in cyan colours (use the legend in the right down corner as reference). *Right-click* on the layer > *Properties* > *Symbology*; change the band selection for Red, Green and Blue to *Layer\_1*, *Layer\_2*, and *Layer\_3*.

**Task 3** Add the `Spot270611.img` to the project. Change the display from the default *Multiband colour* to *Single-band Gray* and select 'Band 1' as the *gray band*. *Right-click on the layer* > *Properties* > *Symbology* > *Render type* > *Singleband gray* > *Apply*. The band will be displayed in a greyscale with poor contrast. See Figure Fig. 11.2



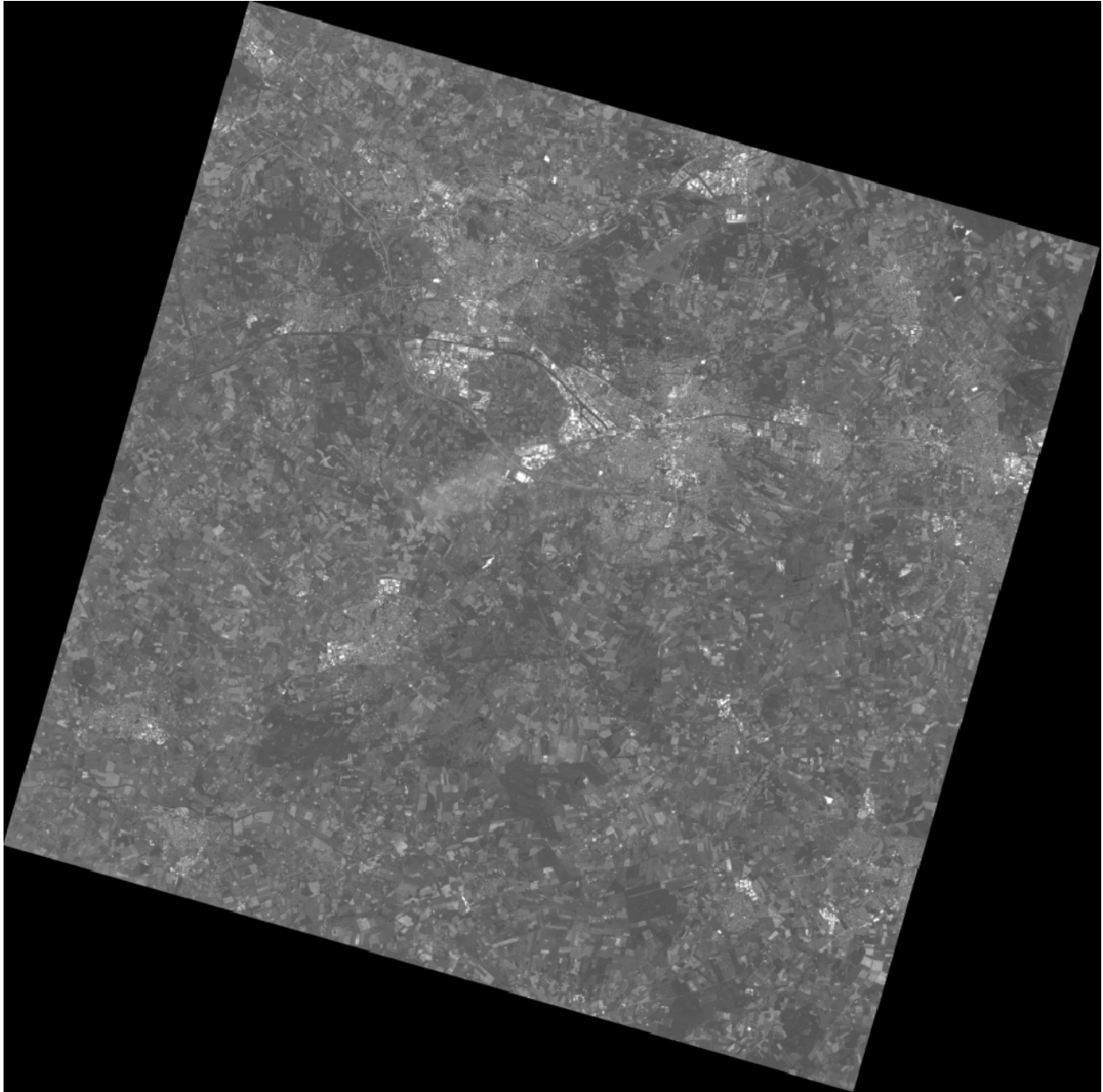


Fig. 11.2: Displaying band 1 of 'Spot270611.img' as greyscale

Then, set min and max values for the contrast stretch. Set **contrast enhancement** to *Stretch to MinMax*. Select *Cumulative pixel count cut* and set the limits to 35% and 98%. Set **Accuracy** to *Actual (slow)*. See Figure Fig. 11.3. Click *Apply*. This will copy the DN values associated with 35 and 98 cumulative percentages the Min and Max of the contrast, respectively.

Contrast enhancement Stretch to MinMax

▼ **Min / Max Value Settings**

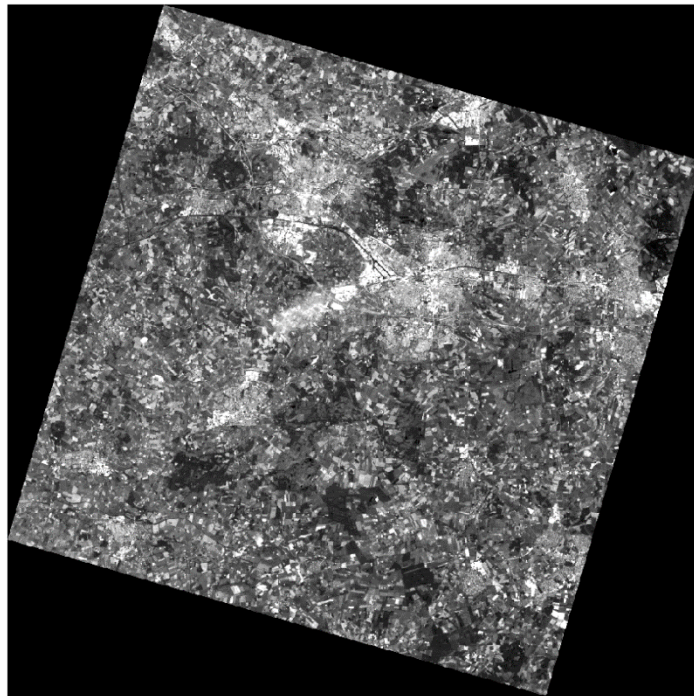
☐ User defined  
☒ Cumulative count cut 35.0 - 98.0 %  
☐ Min / max  
☐ Mean +/- standard deviation × 2.00

Statistics extent Whole raster

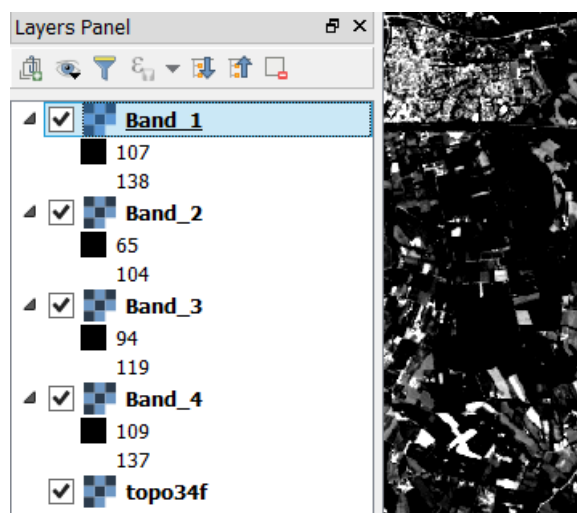
Accuracy Actual (slower)

Fig. 11.3: Contrast enhancement with 'Stretch to MinMax'

We chose a 35% for the minimum because the raster file does not contain image data values for the whole scene. Approximately 35



Repeat the previous task. This time apply a *MinMax Stretch* to all bands of `Spot270611.img`. You can copy a layer by doing right-click on the '`Spot270611.img`' and then choose *Duplicate*. Rename each layer name such that it includes the band number, see the example below.



**Task 4** Compare the results of each band by toggling the visibility of the layers off and on. Give special attention to the comparison of band 3 and band 2. These two bands are displayed with similar composition, and yet they look quite different from the others. This proves that those bands captured different spectral properties.

**Task 5** Use the topographic map `topo34f.img` to find areas in the `Spot270611.img` with Water (Cyan), Buildings (Purple) and Evergreen Forest (Green with overprinted symbols). Then, identify the relative brightness in each of the four bands associated with the areas listed above. Fill in the table below.

Do not spend too much time in identifying representative objects and filling the table. Remember that the decision, whether something is grey or light grey is subjective; thus use the same subjectivity when you fill in the table. If you think a class is represented with more than one brightness, you may select more boxes.

Band	Water	Build-up	Evergreen Forest
1	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>
2	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>
3	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>
4	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>	<input type="checkbox"/> <i>Black</i> <input type="checkbox"/> <i>Dark grey</i> <input type="checkbox"/> <i>Grey</i> <input type="checkbox"/> <i>Light grey</i> <input type="checkbox"/> <i>White</i>

**Note: Reflection.** While working on the previous task, you should have noticed that different classes of land cover may have similar or different brightness within a specific spectral band. Moreover, the same class of land cover may have different brightness in other bands.

### 11.1.2 Multiband Display: Colour Composites

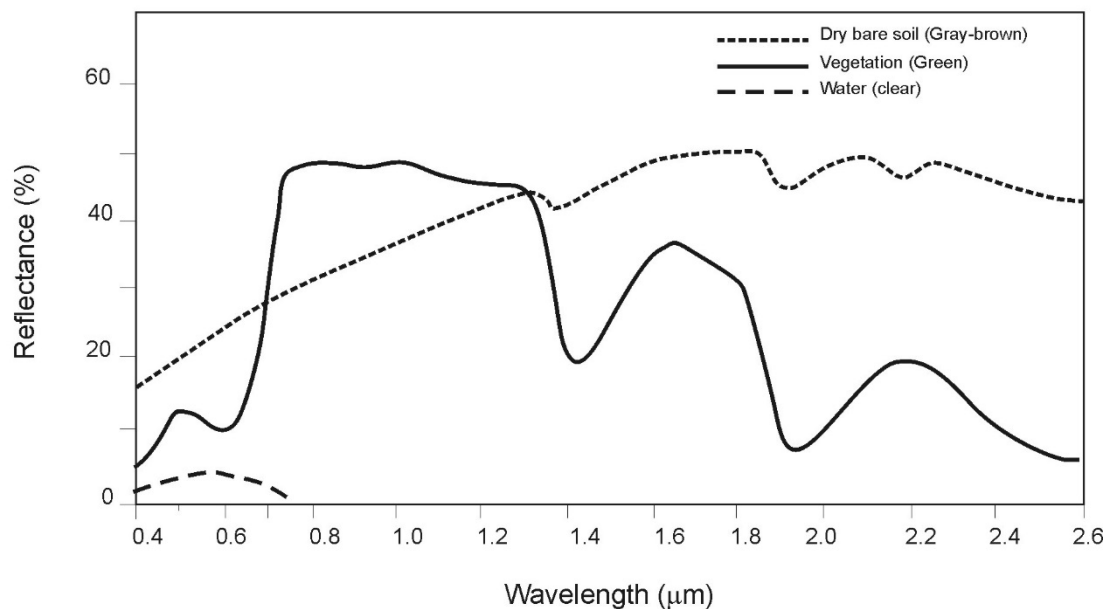
This section will help you to understand the relationship between the spectral property of a class, the selection of spectral bands for visualisation, and the choice of spectral bands in a colour composite. Suppose you have a **SPOT XS** image which includes land cover the types: soil, vegetation and water. Such an image will be displayed with a *contrast stretch* with the band combination of 3, 4 and 2 for RGB.

**Attention: Question.** Which colours will the land cover types above will have in the colour composite mentioned above? You were introduced in the lecture on how this works for one and two spectral bands. Now it is your turn to predict the case for a 3-bands colour composite.

Use the reflectance curves below to estimate how much each of the three land cover types will reflect relative to each other, for each of the 'SPOT XS' bands.

For this, assume that there are no other land cover types in the image; i.e. the land cover with the highest reflectance will have the highest Digital Number in a band. Thus, it will be displayed with the highest brightness (i.e.,  $DN=255$ ). Likewise, the land cover with the lowest reflectance in a band will be displayed with the lowest brightness (i.e.,  $DN=0$ ). For the land cover types that fall between reflectance values, use linear interpolation to estimate its brightness value.

Spectral curves and bands



Terra Aster	<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="5"/>	<input type="text" value="6"/>	<input type="text" value="7"/>	<input type="text" value="8"/>	<input type="text" value="9"/>
Landsat 8 OLI	<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="5"/>	<input type="text" value="6"/>	<input type="text" value="7"/>		
Landsat TM 5	<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="5"/>	<input type="text" value="6"/>	<input type="text" value="7"/>		
Landsat MSS	<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>					
Spot XS	<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>					

**Task 6** Copy RGB brightness values that you estimated into the table. Then use an [RGB calculator](#) to determine the

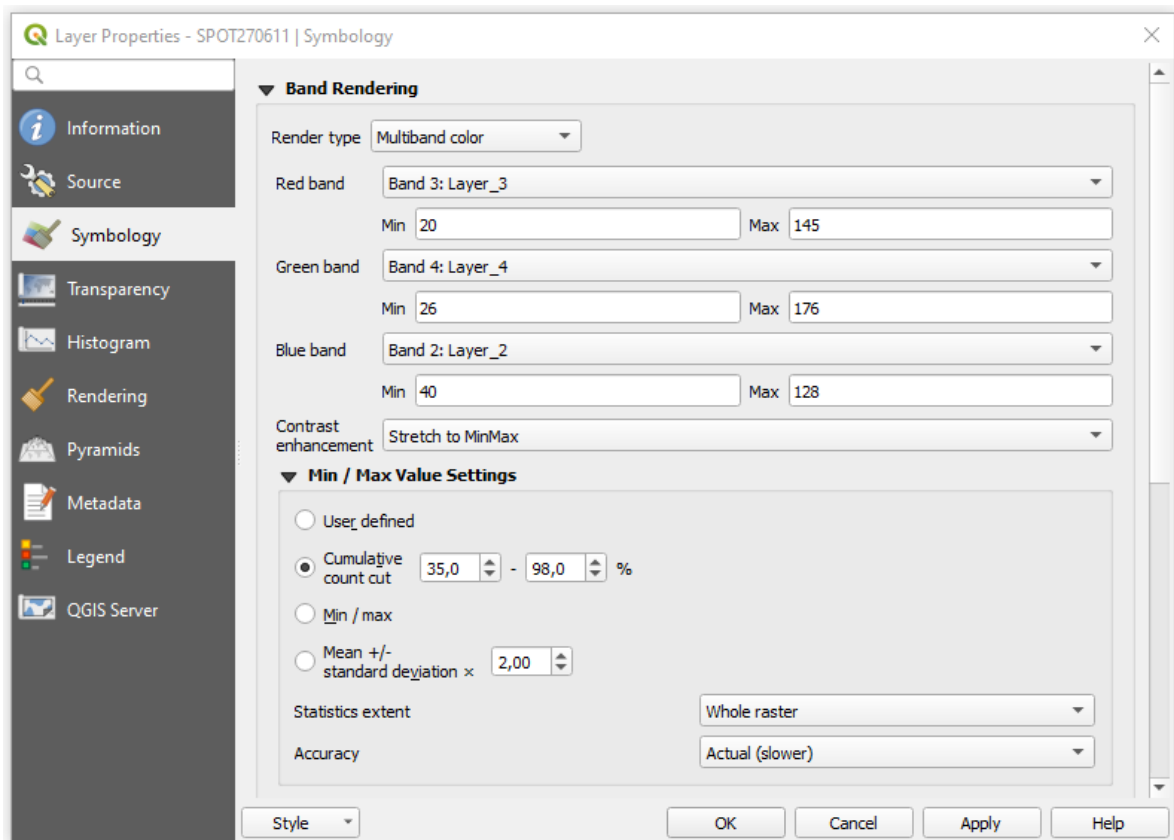
approximated colour of each land cover type in this colour composite.

Major land cover	Band 3 (Red)	Band 4 (Green)	Band 2 (Blue)	Colour in composite
Dry-bare soil				
Vegetation				
Water				

**Note: Reflection.** Based on only theoretical knowledge, you can predict the colour of a land cover type in a colour composite. This is assuming there are no other land cover types present in an image, which may have a lower or higher reflectance. Take the time to understand this statement and study the topic once more if necessary. You could also discuss this topic further in the virtual classroom.

In the previous task, you determined the relative brightness for three types of land cover. Then, you determined the theoretical colour in a specific colour composite. In the following task, you will compare the theoretical colour for water and vegetation with the actual colour in the image when using a colour composite.

**Task 7** Open the `Spot270611.img` in QGIS and use a band combination of 3, 4 and 2 for Red, Green and Blue. Set the contrast enhancement to 35% and 98% for all bands. Use **Actual (slower)** for *Accuracy*, as shown below.



Zoom in to the image so that you can easily confirm what colour was assigned to water. Also verify whether the theoretical colours, determined in the previous section, match the colours that you see in this colour composite.

**Attention: Question.** When comparing the theoretical colours and the result of the colour composite, you will find out that there are discrepancies. **What could be the causes behind such discrepancies?**

---




**Note: Reflection.**

You should realise that by knowing the relative **spectral reflectance** of a class in every spectral band helps to understand and interpret images. Such spectral reflectance is visualised as relative brightness. Moreover, you could predict, for a given sensor, the ranges of DN values for a specific object, if you know and understand the spectral properties of such objects.

Some type of land cover might have similar or the same spectral reflectance property in specific ranges of the Electro-magnetic (EM) spectrum. The integration of more bands in the analysis, and the assessment of the spectral properties in other ranges of the EM spectrum, can result in the successful discrimination of more types of land cover. This is true not only in the case of land cover but also for many other objects in a remote sensing image.

---

## 11.2 Image Enhancement by Histogram Operations

 **Image enhancement** describes a set of operations that aim to improve the way certain features in an image are displayed. One method to enhance images, for interpretation or analysis, uses  **Histogram operations**. These operations use the  **Histogram** of an image to control how it is displayed, and they are usually known as **global contrast enhancement operations**.

---

**Important: Resources.** You will require the same dataset as in the previous section: [data\\_histogram\\_operations.zip](#).

- To experiment with contrast enhancement, we will use a TM image of the 24th of august 1999, `tm24aug99.img`. This image covers a large water body and different types of land cover, and it has areas covered by clouds.
-



## 11.2.1 Contrast Stretching

**Task 8** Display the `tm24aug99.img` using the band combination 4, 5 and 3. Set the *Stretch to MinMax* to a *Mean +/- standard deviation* of 2.0, and the *Accuracy* to **Actual (slower)**; as shown below.

**Band Rendering**

Render type: Multiband color

Red band: Band 4: Layer\_4  
Min: -51.6949 Max: 154.111

Green band: Band 5: Layer\_5  
Min: -41.6123 Max: 114.607

Blue band: Band 3: Layer\_3  
Min: -28.2911 Max: 137.2

Contrast enhancement: Stretch to MinMax

**Min / Max Value Settings**

☐ User defined

☐ Cumulative count cut: 2.0 - 98.0 %


☐ Min / max

☒ Mean +/- standard deviation × 2.00

Statistics extent: Whole raster

Accuracy: Actual (slower)

**Task 9** Zoom in to an area covered by clouds over the mainland (centre right). Go to *Layer Properties > Symbology*. Change the *Statistics extent* to *Current canvas* and click *OK*.

Click on **Zoom Full**  to zoom out to see the whole image. Instead of seeing clouds in whites and pinks, you now can see clouds in several colours.



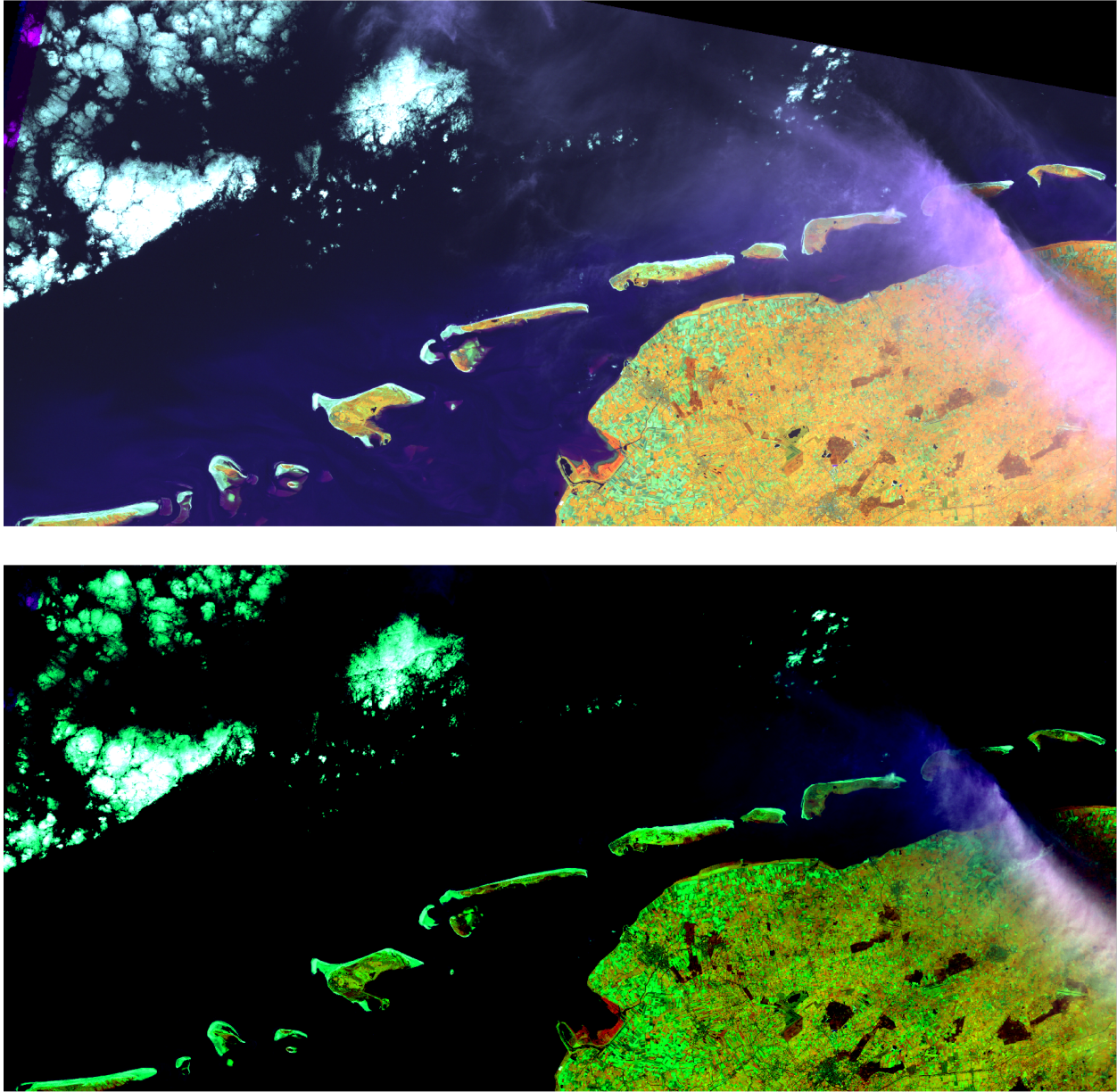
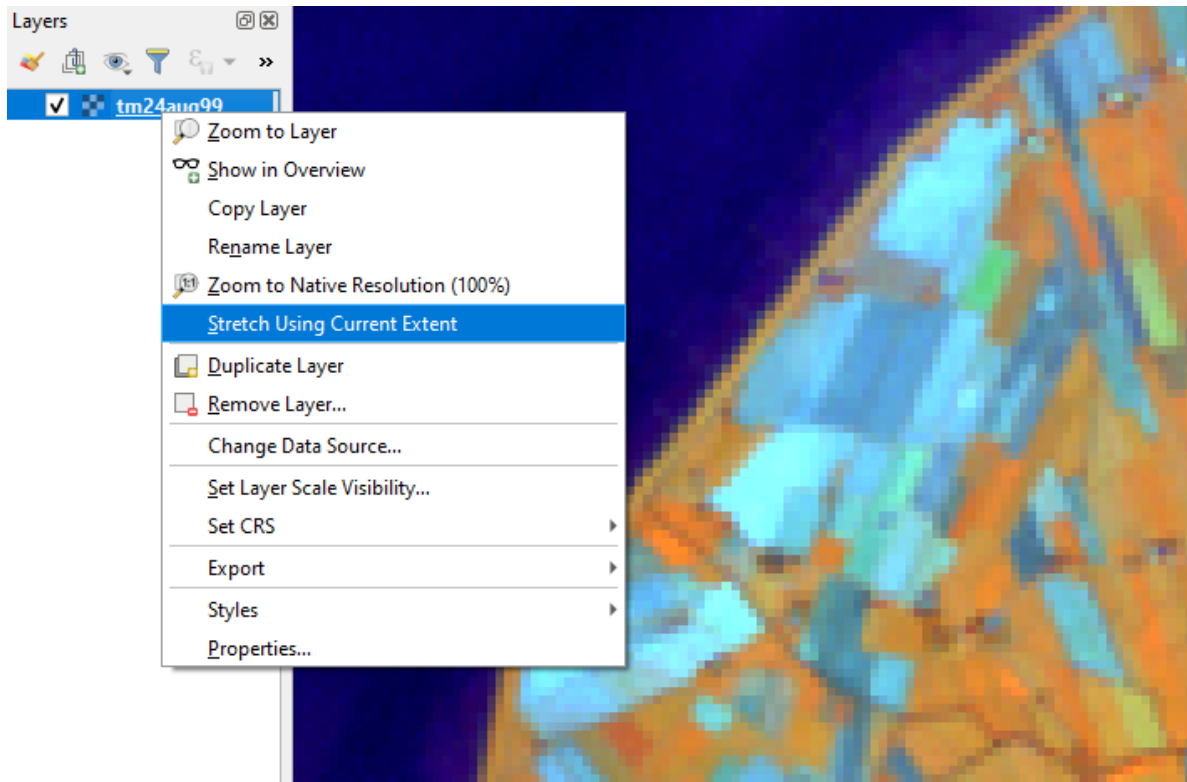


Fig. 11.4: Top: Image 'tm24aug99' with global contrast enhancement. Bottom: Image 'tm24aug99' with local contrast enhancement for clouds

**Attention: Question.** Do you also think that the mainland is displayed in a not-so-good way?

**Task 10** Zoom into an area with mostly land and some water, and re-apply the same *Contrast enhancement* method as before. You can use the context-sensitive option **Stretch using current extent** as shown below, but verify that it does what you intend.

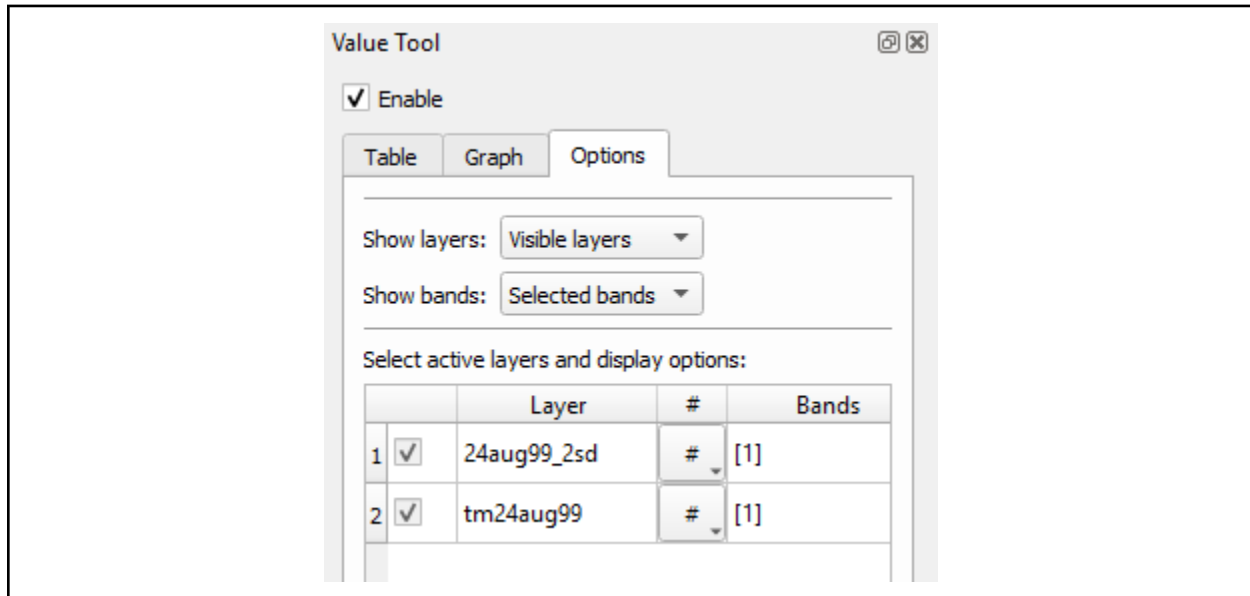


You will notice a change in contrast in the image. This is because the part of the image that is currently visible include different types of land cover; thus, different local statistics. The results of a contrast stretch based on local statistics changes when the range of values used in the computation changes.

**Task 11** Set the **Contrast enhancement** back to *Mean +/- standard deviations* and the **Statistics extent** to *Whole raster*. Then, right-click on the 'tm24aug99' layer and do *Export > Save as...* For **Output mode** select *Rendered image* and enter a self-explanatory name for the output file. Save the file to an existing folder. The file will automatically be added to the Map View.

**Attention: Question.** Is the exported image different from the original `tm24aug99.img`?

Compare the properties of the two files, especially the histograms. Determine whether they only look-alike, or if they are the same. You can use the **Value tool** for the comparison by setting it up in such a way that the tool shows the values of the same band for both images, as shown below.



### 11.2.2 Choosing Min and Max values

To choose the *min* and *max* values for a contrast stretch, the user has to consider which areas of an image are of interest, or which types of land cover are relevant for specific purposes.

**Task 12** Remove the exported image from the project; keep only the original image. Zoom into an area on the mainland which is primarily dark orange/brown; they represent areas with forest. Then, on Layers Panel, *right-click* the original image > *Properties* > *Symbology* > *Min / Max Value Settings*. Set the *Statistic extent* to **Canvas extent**. See Figure Fig. 11.5. Click *Apply*. This will compute the histogram statistics only of the part of the image that is visible in the Map View. Notice that there are other options to compute statistics than just *Min/Max*. For example '*Mean +/- standard deviation*'.

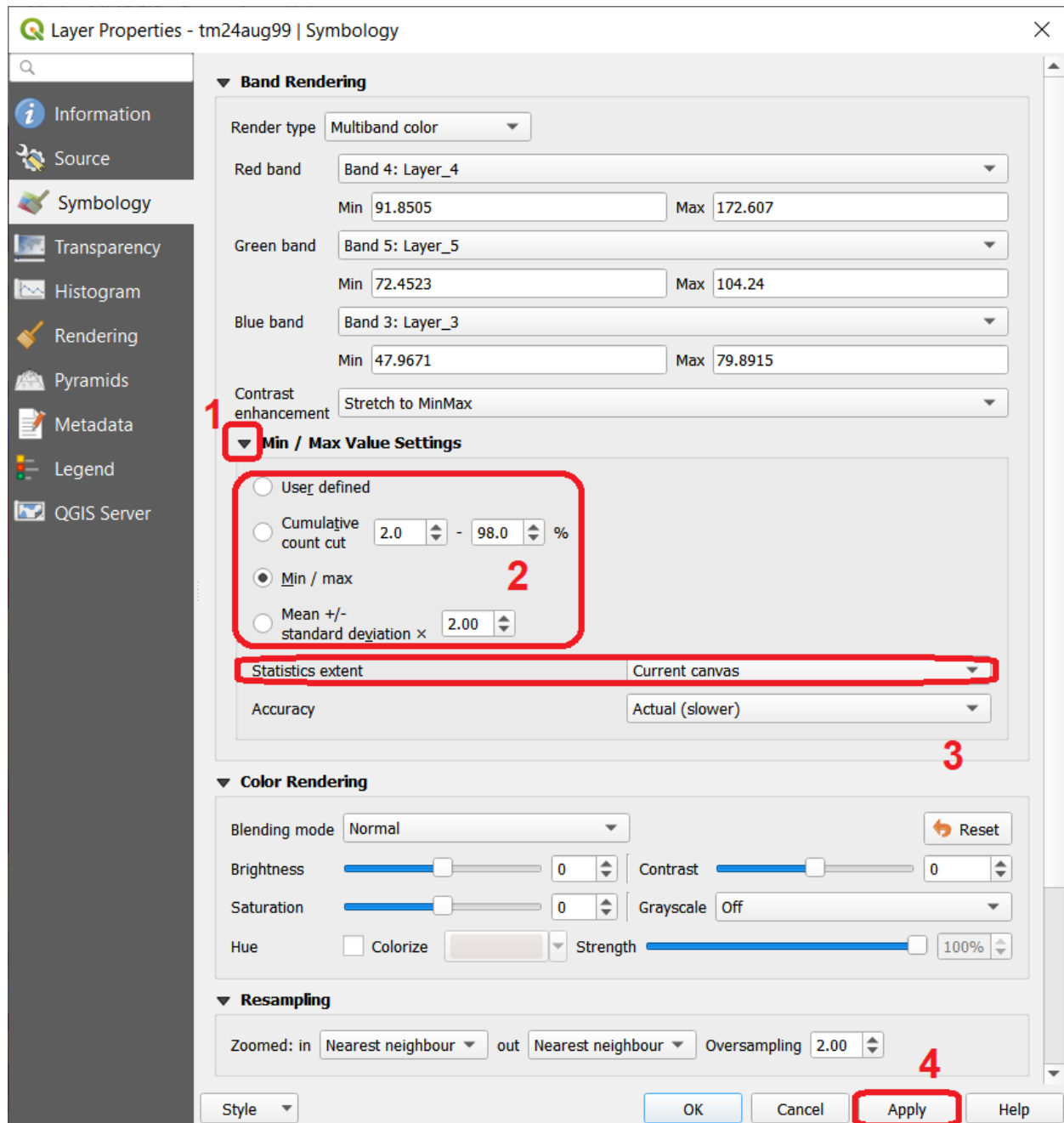


Fig. 11.5: Applying local statistics for histogram stretching based on the 'Canvas extent'.

**Note: Reflection.** To correctly apply contrast enhancement for specific types of land covers, you need to know which are the types of interest. What their spectral signatures are; the specifications of the spectral bands of the sensor which you have chosen; and you need knowledge of additive colour mixture.

**Task 13** Add the `tm25aug99_sub.img` to the project, and display it using a band combination 4, 5 and 3. this image covers shallow water and land with various types of land cover. Analyse the histograms of the three bands for this image; *right-click* on the layer > *Properties* > *Histogram*. Compute the histogram if necessary. Then,

*Prefs/Actions > Show selected band* and choose the band you want to inspect.

**Attention: Question.** In which band on display do you expect a significant difference in DN Values between water and land? Use your knowledge on EM radiation. A bi or tri-modal trend in the histogram is an important clue.

**Task 14** Use the histograms to identify approximate values for a *contrast stretch* which will enhance the contrasts between types of coverage on the part of `tm25aug99_sub.img` with land. Save the result using *Export As.. > Rendered image*. Remove the resulting layer from the project.

**Task 15** Repeat the previous task. This time use the histograms to set a contrast that will enhance the image for distinguishing shallow water.

**Note: Reflection.** It should be clear to you that for some contrast enhancement methods, statistics of the data play an important role, e.g. mean and standard deviation, minimum and maximum. However, when you know the spectral properties of the objects of interest, the characteristic of the scene, and the sensor; you can interpret the histograms directly and make improvements to make effective use of the brightness values in an image.

## 11.3 Image Enhancement by Filter Operations

 **Filtering** describes a set of radiometric operations used to enhance images. Filters are applied to images for the sake of  **Noise reduction**,  **Edge Detection**, and  **Edge enhancement**.

### Important: Resources.

You will require the latest LTR version of QGIS, plus the dataset [Data\\_Filter\\_Operations](#). When you unzip the dataset, you will find the following files inside:

- `tm1999_b4.tif` – A scene from band 4 of the Landsat TM. Enschede in 1999.
- `tm1999_xs_ml_classification.tif` – Classification of the types of land cover in Enschede. From Landsat TM 1999.
- `SW-NE_3x3.txt` – Definition of a custom filter.

QGIS offers the possibility to apply all kind of filter kernels on images which are displayed in a viewer. In this exercise, we will use tools that apply filters and store the output as temporary files. In such a way, we can easily compare different results.

**Task 16** Install the **Profile tool** plugin. Go to *Plugins > Manage Install Plugins*, and install the plugin.

**Task 17** Set the default contrast stretch to use the 2% and 98% of the cumulative pixel count for grayscale images. In the *Settings* menu, go to *Options > Rendering* tab. Scroll down to *Contrast enhancement settings*, and set the default value for **Single band gray** to *Stretch To MinMax*. Then, set **Limits (minimum/maximum)** to *Cumulative pixel count cut*. Make sure that the cut limits are set to 2.0% and 98.0%. Click *OK*.

**Note: QGIS.** QGIS is not specifically tailored for Remote Sensing and does not provide filter tools directly. Such filtering tools are available through the processing toolbox using external providers like *SAGA* and *GRASS*.

*For an overview on how to use the Processing Tools in QGIS, watch the [introduction to processing](#) video tutorial.*

**Task 18** In the *Settings* menu, go to *Options > Processing* and check that you have the SAGA and GRASS providers enabled.

### 11.3.1 Linear Filters

#### Smoothing Filter

**Task 19** Apply a linear filter to the 'tm1999\_b4' image. In QGIS, open the tm1999\_b4.tif. Your project should assume the same Spatial Reference System as the image (*EPSG:32632 WGS84/UTM zone 32N*). In the **Processing Toolbox**, open the SAGA tool called **User defined filter**.

Confirm that 'tm1999\_b4' is the input and click **Default filter matrix (3x3)** to open an empty filter kernel. Enter the weights of an *average filter kernel*. Ensure that the sum of weights is equal to 1. Confirm with OK. In the **User defined filter dialogue** execute the kernel by clicking OK. The output is added to the Map View as a temporary file.

In the **Layers panel** right-click the 'Filtered Grid' layer and rename to 'Average'.

**Attention: Question.** Which kernel weights did you use in the previous task? Write them down.


**Task 20** Explore the filter results around the Twente Airport. Reset the zoom to fit the image to the **Map View**. Next, change the scale, in the box at the bottom of the Map View, to 1 : 75,000.

Zoom in to the major runway of the *Twente Airport*. See Figure Fig. 11.6.

We will use the **Profile tool** to compare the results of the average filter and original image. If you do not know how to install the *Profile Tool* plugin, watch the video tutorial on [installing plugins in QGIS](#).

Start the Profile tool. Select one of the layers in the **Layer Panel**. Zoom into the centre of the image and draw a profile (line) across the major runway. Click on **Add Layer**. Select the other layer in the **Layers panel** and click on *Add Layer* again.



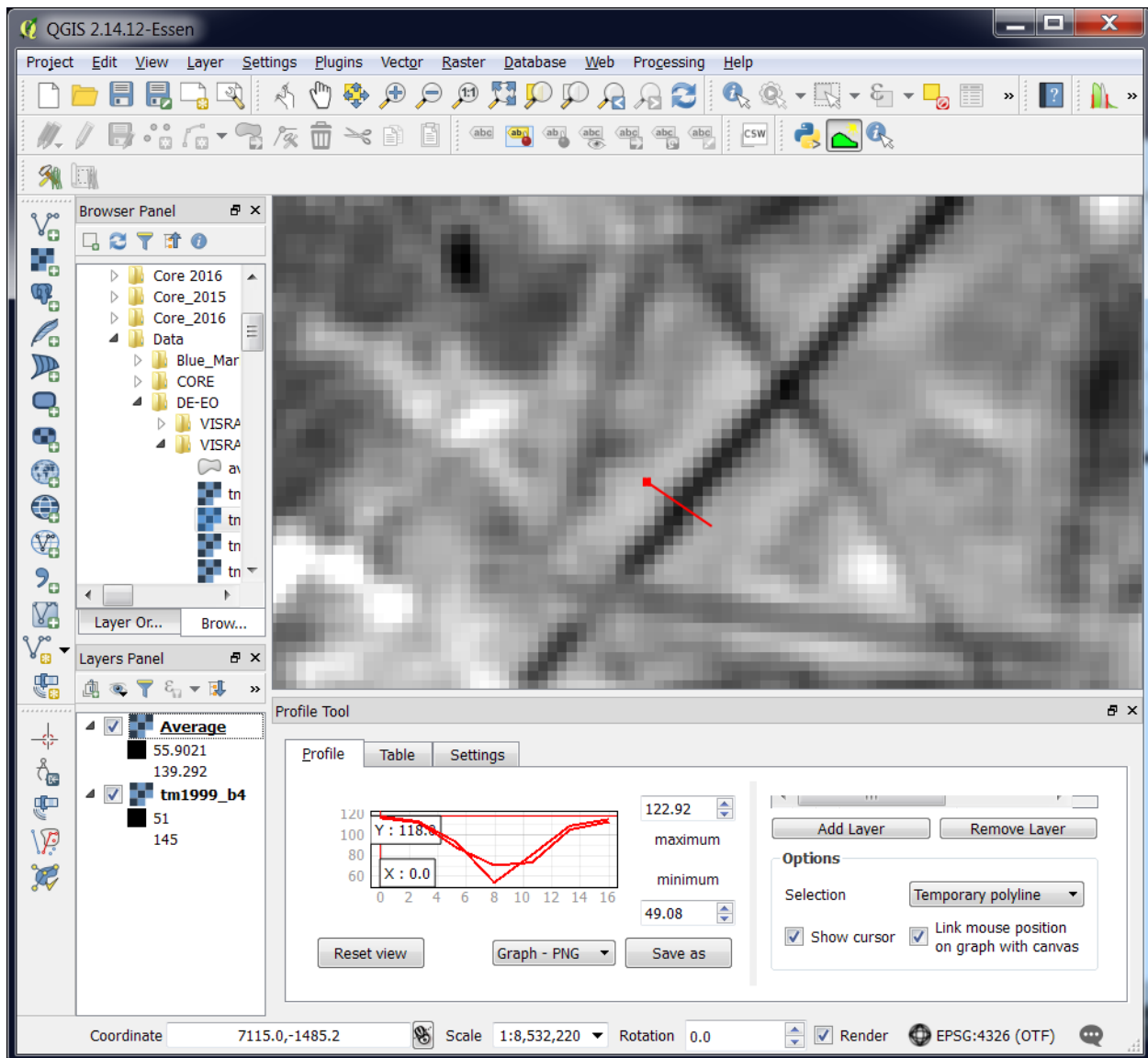


Fig. 11.6: Comparison of the result of a smoothing filter to 'tm1999\_b4.img' using the 'Profile tool'

**Attention: Question.** Just by looking at the graph of the Profile tool, can you tell which profile belongs to the layer with the average filter?

**Task 21** Draw profiles at different locations, and confirm your knowledge of the effects of applying an average filter (smoothing) to the image.

## Gradient Filter

**Task 22** Use the **User defined filter** tool to apply a filter using the weights in the figure below, to the original 'tm1999\_b4' layer. Rename the resulting layer to 'Laplace'.

-1	-1	-1
-1	8	-1
-1	-1	-1

**Error:** You will notice that the output of the Laplace filter contains many pixels with *No Data* (hollow pixels). Possibly, you also noticed an error message when executing the filter. This is an issue with the tool itself, possibly a bug. Please, ignore the *No Data* pixels this time and concentrate on the outcome in the rest of the image.

**Attention: Question.**

- Is the kernel above a detection kernel? If yes, what does it detect?
- Does the layer resulting from the previous task contain the same brightness as the original image for area objects?

**Task 23** Examine the result of the Laplace filter. Toggle on and off the visibility of the 'Laplace' layer to check what has happened. Zoom in to the edge of the image until you see individual pixels. Toggle the visibility of the 'Laplace' layer again.

**Attention: Question.** What phenomena do you observe? Can you explain it?

**Task 24** Open the histogram of the 'Laplace' layer. Go to *Properties > Histogram*; check the values in the image.

**Attention: Question.** Around which value does the histogram has its centre?

**Task 25** In the **Profile** tool add the 'Laplace' layer and toggle the visibility for the other layers. Confirm that the filter kernel detected two edges, i.e. both sides of the runway.

The 'Laplace' layer looks rather artificial. The brightness of the original image is gone; the lighter and darker areas in the original have now a common grey tone and high contrasting pixels at the edges. This filter has detected the changes (edges) between local lighter and darker pixels. The circular build of the kernel that you applied, i.e. all negative weights around the centre with positive weight, detected changes in all directions.

**Note: Reflection.** We can use the output values of an edge detection kernel to discriminate **edges with high contrast** and **edges with low contrast**. For example, a field with relatively low values and a neighbouring field with relatively high values will result in edges with high contrast. In comparison, edges with low contrast may be the result of adjacent pixels which have a slightly different value. You could use a threshold to select only edges with high contrast and delimit the edges between areas or fields.



Directional filter kernels have **positive** and **negative** weights, and their output values could be positive or negative. Positive values represent edges which correlate positively with the kernel, negative values represent edges which correlate negatively – ‘they are opposite to’ –. We can also use the sign of the values to discriminate different classes or types of edges.

If you do not understand these statements, review the lecture material once more.

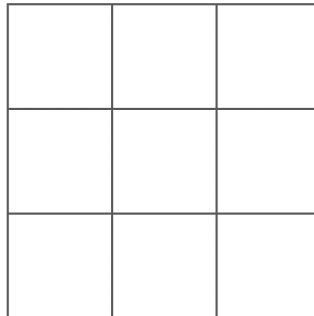
## Edge Enhancement

**Task 26** Repeat the steps of the previous task, but this time use kernels with the following values for the centre: 12, 16 and 200. Increasing the centre value will increase the weight of the centre pixel in the original image. When using a value of about 16 for the centre pixel; the kernel will calculate the Laplace enhancement of the image. Then, the resulting layer will look like the original image.

## Custom Filters

The Laplace kernel detects edges in all directions. We can also define kernels which detect edges in specific directions. In the next task, you will use a custom filter defined in a text file. Check the content of `SW-NE_3x3.txt` by opening in a text editor.

**Attention: Question.** Which are the weight of the filter define in `SW-NE_3x3.txt`? Write them down.



**Task 27** In the Processing Toolbox, use the **r.mfilter** tool of GRASS to detect edges in a specific direction. Use the file `SW-NE_3x3.txt` as *Filter file*.

Check the results and confirm that one of the runways of the *Twente Airport* was not detected at all! Also, ensure that the edges of the main runway were detected. The result should show positive values on one side of the runway and a negative on the other. This is because of the correlation of the results with the positive and negative weights in the kernel.

Notice also, that the edges in the resulting layer have a slope and a direction. In this case, the kernel has detected edges in the SW-NE direction.

## 11.3.2 Enhancement using Non-linear Filters

### Rank-Order Filter

**Task 28** In the Processing Tools, look for the SAGA **Rank filter**. Select the 'tm1999\_b4' layer as the input grid. Set the *Search Mode* to *Square* and a Radius of 1. Use a *Rank (Percent)* of 50. This settings essentially define a **median filter**. Execute the filter.

Use the **Profile tool** to inspect the results and confirm the difference between the original image and the result of the *Average and Median* filters. Pay special attention to locations where you expect variations, for instance, around the edges between areas.

### Majority Filter

**Task 29** Add the tm1999\_tm\_xs\_ml\_classification.tif into the Map View; this layer contains a set of classes representing land cover. You will notice that there are many isolated pixels inside some homogeneous areas. For example, the yellow pixels identify maize, which usually does not grow in such small parcels.

In the **Processing Toolbox**, open the **r.neighbors** tool. Select 'tm1999\_xs\_ml\_classification' as input raster and set the neighbourhood operation as 'mode' (also known as **majority filter**).

The results will be shown as greyscale or black and white. You can apply a pseudocolour to visualise the results properly. Copy the style of the 'tm1999\_xs\_ml\_classification' into the result of the majority filter layer. See Fig. 11.7

In the **Layer panel**, right-click on the 'tm1999\_xs\_ml\_classification' layer and copy the style. Then, right-click on the majority filter layer and paste the style.

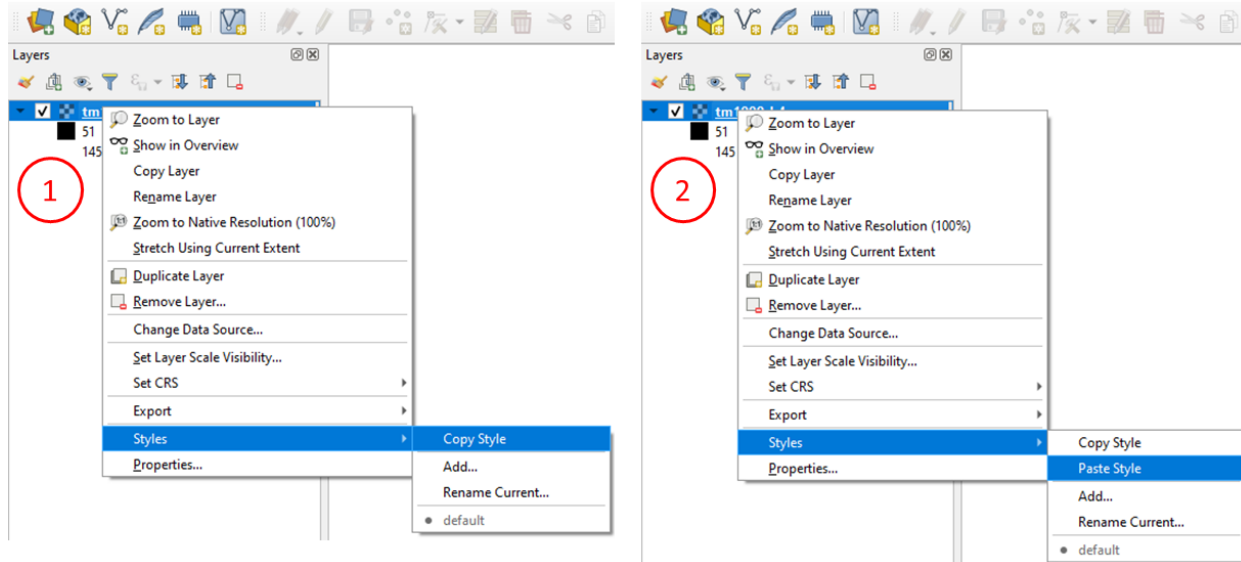


Fig. 11.7: Copying the style between two raster layers in QGIS

**Task 30** Toggle the visibility of the filtered result and compare it with the original landcover layer. Confirm that most isolated pixels have disappeared, and that thin lines of pixels surrounded by homogeneous areas also disappeared (e.g. a runway in the Twente Airport). The main runway should still be distinguishable.

**Attention: Question.** Can you explain why one of the small runways of the Twente Airport disappeared after applying a majority filter?

**Task 31** Use the `r.neighbor` to compute another mode filter on the `'tm1999_xs_ml_classification'` layer. This time use a neighbourhood of size 5; which is to apply a  $5 \times 5$  kernel. Compare the result of this filter with the  $3 \times 3$  majority filter.

**Attention: Question.** What do you observe when comparing results of  $3 \times 3$  and  $5 \times 5$  majority filters around the main runway of the Twente Airport?

**Task 32** Experiment with the application of consecutive filters. Apply a  $3 \times 3$  majority filter to the results of the already existing `'3x3 majority'` filter layer. Compare these results with the result of applying a single  $5 \times 5$  majority filter. You should see that the results are not the same.

---

**Note: Reflection.** In summary. You should acknowledge that in the case of neighbour operations, such as the majority filter. The results change depending on the size of the kernel, and the number of times a filter is applied to a dataset.





---

*Section author: Wan Bakx, André Mano & Manuel Garcia Alvarez*



## RADIOMETRIC CORRECTIONS

The procedures described in this section fall within the group of “pre-processing” techniques in image processing. They focus on correcting artefacts originated from atmospheric disturbances.

As shown in Fig. 12.1, *radiometric correction* techniques are applied to Remote Sensing data before any *geometric correction*, and before performing the main *data analysis*. Radiometric corrections fall within three groups:  *Imperfect sensor correction*,  *Sun angle correction* and  *Atmospheric correction*. Atmospheric corrections are needed to “clean” the images from disturbances produced by water vapor, aerosols and gases in the atmosphere, and retrieve pure radiances from a target on the ground. A common technique for atmospheric correction is  *Haze correction*.

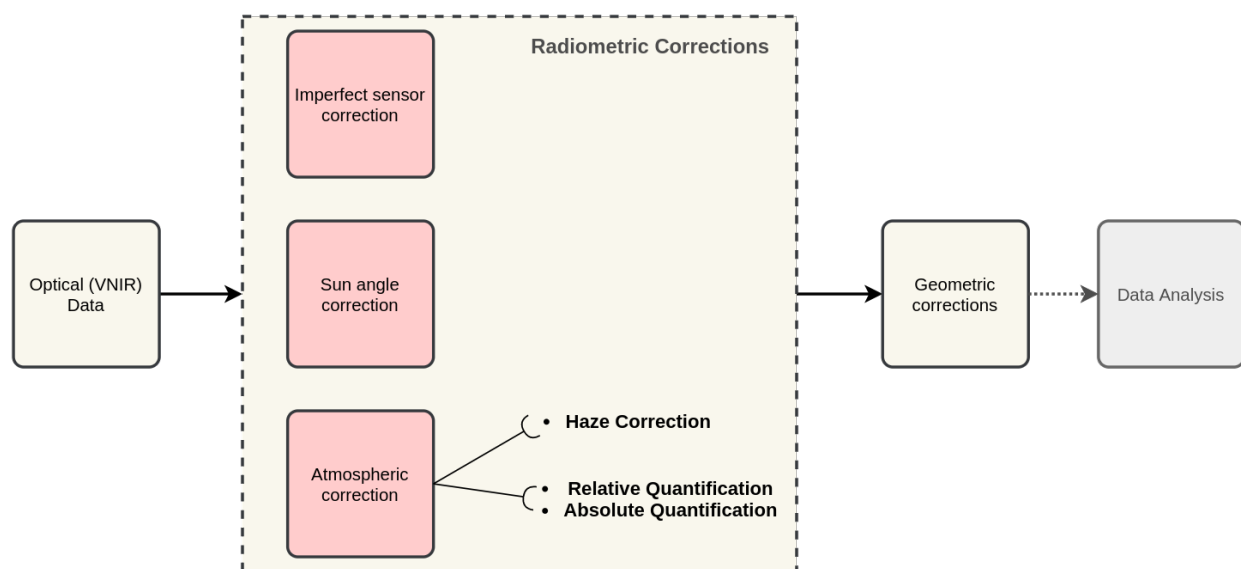


Fig. 12.1: A sequence of pre-processing steps to apply radiometric corrections on optical imagery

---

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset [Data\\_Atmospheric\\_Corrections.zip](#). When you unzip the dataset, you will find the following files inside:

- `etm99.img` – Scene from Landsat Enhanced Thematic Mapper, bands 1 to 4. Wadden islands, 1999.
- `pan.img` – Scene from SPOT, panchromatic band.
- `tm89.img` – Scene from Landsat Thematic Mapper.

These three images are provided with almost cloud-free skies, and the coordinate system is **EPSG: 32631**.

## 12.1 Haze correction

**Task 1** Use the [Satellite and sensor database](#) and the file-properties to find information about the spectral specifications of *SPOT PAN*, *Landsat TM B1 to B4* and *Landsat ETM B1 to B4*. Then, complete the table below.

Satellite/sensor	Filename(s)	Date	Resolution GSD
SPOT/HRG panchromatic	PAN	03-Jul-1989	
Landsat-5/TM	TM89	21-Dec-1989	
Landsat-7/ETM+	ETM99	03-Nov-1999	

**Attention: Question.** Are the spectral characteristics of the TM and ETM sensors different?

**Haze has an additive effect to the overall image, which results in higher DN values.** Therefore, haze reduces the contrast of the image. Because this effect depends on  $\lambda$ , its influence differs for each band.

In the following task, we will assume that there is a real *blackbody* present in the images; and do a subtraction to reduce pixel values to a minimum of zero.

**Task 2** Open `etm99.img`, `pan.img`, and `tm89.img` in QGIS. and take a look at the images. Display all images using **bands 4, 3, 2 for RGB** and *No Stretch*. Open the **Value tool** to explore the pixel values.

Assume that the blackbody in these images is *deep, clear and unpolluted water*. The island in the images is surrounded by a tidal flat, which causes tidal currents with a tidal frequency. It is not easy to locate clear water around the island in such situations. We recommend using a body with following water as assumed-blackbody, for example the small lake on the island.

Follow the steps below to complete the following table. You have to determine the **minimum values** for the assumed-blackbody (lake) on the images.

Image	Band 1 DN	Band 2 DN	Band 3 DN	Band 4 DN	Pan DN
PAN	<b>n.a.</b> <sup>1</sup>	<b>n.a.</b>	<b>n.a.</b>	<b>n.a.</b>	
TM89					<b>n.n.</b> <sup>2</sup>
ETM99					<b>n.n.</b>

1. Use the **Value tool** to read DN values. Note that the tool shows pixel values for all bands of the multispectral image. *Read the value for the proper band!*
2. Select the '*ETM99*' image, go to Properties and change the symbology to single band grayscale for Band 4. Set minimum to 0 and maximum to around 40 to increase the contrast for the lowest DN values (these are objects with the lowest reflectance properties, which are more likely to be water).
3. Zoom into the area around  $[lat, long] = [53.46145, 5.652432]$  or  $[x, y] = [676191, 5927074]$ , for comparison. This location is a small lake on the island and shows much lower values than most of the sea in the image. Search the lake for the lowest DN value. Keep a list of lowest values on scratch paper and copy these to the table.
4. Repeat steps 2 and 3 for images '*PAN*' and '*TM89*'.

<sup>1</sup> Not available

<sup>2</sup> Not necessary

**Attention: Question.** Compare the minimal value found in the lake or tidal flat with the minimum value in that band.

- Why is there a discrepancy between the values? Think about the spectral and spatial characteristics of the band and the spatial dimensions and spectral property of the object.
- Why are there decreasing values for haze for 'ETM99' from Band 1 to Band 4?
- How can you explain the difference between the Haze values of the different images?
- What about the 'PAN' image? What range on the spectrum does it cover? Does this relate to the observed difference between 'ETM99' channels and the panchromatic channel?

**Task 3** Correct the image for haze by subtraction values using the Raster calculator. In QGIS go to *Raster > Raster Calculator*. Specify the formula to subtract the Haze value from Band 4 of the 'ETM99' image, and specify the name of the output file.

Select the newly created *Band 4 with haze correction* on the **Layers Panel**. Go to Properties and modify the contrast stretch as follows: **minimum: 0, maximum: 255**. Do the same for the *Band 4 without haze correction*.

**Attention: Question.** Compare the two versions of band 4 of the 'ETM99' image. Do they look different?

Verify that the histograms of the haze-corrected bands have shifted towards the origin. Both histograms have the same shape before and after haze correction, but a different location. It also shows that there are some pixels with DN values lower than the small lake on the island. These negative values should not exist in EO images, but they are caused by the fact that we use a GIS to do the calculation. Before continuing, we have to correct this artefact.

**Task 4** From the **Processing Toolbox**, use the SAGA module **Reclassify values (simple)** to set all negative values to 0. For *Grid* choose the 'haze-corrected' layer, and for *Replace Condition* choose [0] Low value <= grid value < high value.

Edit the Lookup table; add a row and enter -255 for *Low Value*, 0 for *High Value* and 0 for *Replace with*. Just like in the table below. Click **OK > Run**. This will replace all values in the range [-255, 0] with 0, and produce a new raster layer.

	Low Value	High Value	Replace with
1	-255	0	0

**Task 5** Calculate the Haze correction for all bands of 'ETM99', for band 4 of 'TM89' and the 'Spot PAN', including the reclassification.

**Attention: Question.** Suppose you had a SPOT image with the values for the channels that correspond to R, G, and B. Would the Haze values be different than those in the TM scene? Why or why not?

## 12.2 Sun Angle Correction

So far we have corrected for haze by simplifying the illumination formula, such as:

$$DN = (T\sigma + T\delta_i)R_i$$

The sun angle determines the amount of irradiance per unit area on the ground. We can correct for sun angle attenuation by normalising the sun angle to a zenith situation. We do this by dividing the image values by the sine of its sun elevation angle (in degrees). This correction will result in slightly higher values:

$$DN' = \frac{DN}{\sin(\sigma)}$$

Where  $DN$  is the input pixel,  $DN'$  is the output pixel value, and  $\sigma$  is the sun angle. Note that because the angle is smaller than 90 the sine will be smaller than 1 and as a result  $DN' > DN$ .

In case you have many images with various sun angles, you can calibrate them to the zenith. Or you can select one of the images, which serves as a reference for sun angle correction and correct the other to match the reference sun angle.

### 12.2.1 Absolute Sun Angle Correction

Usually, you will find the sun elevation angle  $\sigma$  in the header file of the images. However, this time, we provide you with the value for the 'SPOT PAN' image, in the table below.

**Attention: Question.** What is the sine for the  $\sigma$  of the 'SPOT PAN' image you used in the previous section Write the answer in the table below.

Scene	$\sigma$	$\sin(\sigma)$
PAN	58.9°	

**Task 6** Use the **Raster Calculator** to correct for the sun angle on the 'haze-corrected SPOT PAN' image.

**Task 7** Locate some particularly dark and bright pixels with the **Value tool** in the 'SPOT PAN', and examine the difference between *haze-corrected* and *haze-and-sun-angle-corrected* values.

**Attention: Question.**

- In which way have the values of the *haze-corrected* and *haze-and-sun-angle-corrected* images changed? Are the changes constant, linear, or exponential? Were the changes predictable?
- Why did we apply haze correction before sun angle correction? Does the other of these corrections make a difference?
- In case of multiband images, we do not need to calculate the sun-angle separately for each of the bands. Why not?

Section author: Wan Bakx



## DIGITAL IMAGE CLASSIFICATION

In this exercise, we use a **Sentinel-2** satellite image covering Enschede and its surroundings from September, 2016. Before going forward with the exercise read about  [digital image classification](#).

---

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset [digital\\_image\\_classification.zip](#). When you unzip the dataset, you will find the following files inside:

- `s2_25_sept_2016.tif`, a Sentinel-2 satellite image of Enschede.
  - `Training_areas.shp`, containing the sites visited during a fieldwork.
  - `digital_image_classification.qgs`, a project containing the files above.
  - `Field_Photos_201`, a folder containing photos of the surroundings of training areas.
- 

The methodology for mapping land cover using satellite images consist of the following general steps:

1. Interpretation of satellite image
2. Field data collection
3. Field data analysis
4. Supervised Classification
5. Accuracy assessment

In this exercise, you will conduct the steps mentioned above except for the **Field data collection**. However, we briefly explain that step in the sections *Field Survey* and *Organising Field data*. The workflow in [Fig. 13.1](#) summarized the process of applying *digital image classification* for land cover mapping.

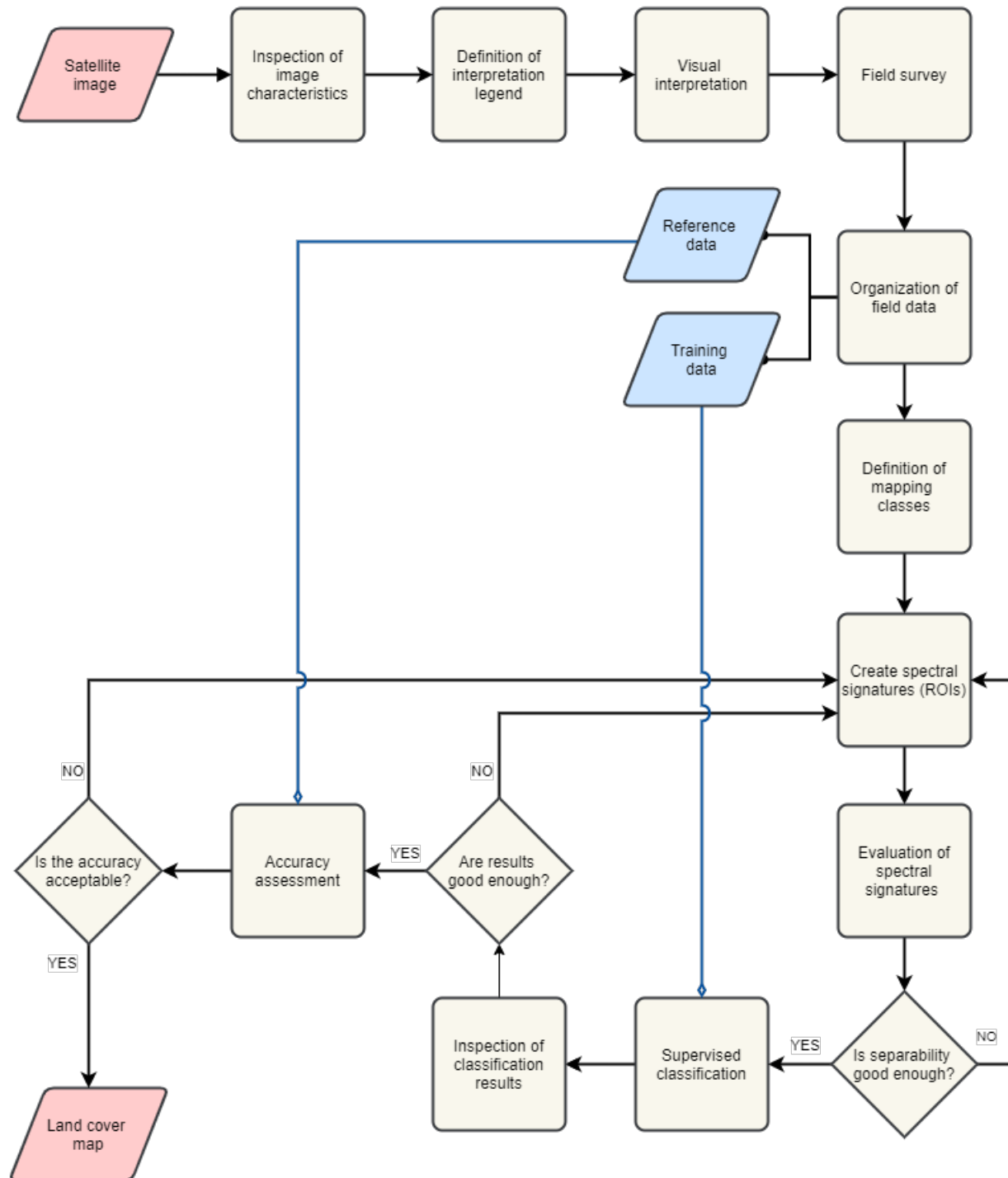


Fig. 13.1: A workflow of the process of digital image classification for land cover mapping

## 13.1 Inspection of Image characteristics

The first step is to get familiar with the dataset by reviewing the characteristics of the image dataset. We will use only some of the spectral bands of the Sentinel image because of their differences in spatial resolution.

**Task 1** Inpect Table 3.1., open `s2_25_sept_2016.tif` in QGIS and get familiar with this dataset. Notice that the bands in the image refer to the numbers under the column **Exercise Band**.

Table 13.1: Spectral bands of Sentinel-2. The bands used in this exercise appear in bold.

Band	Wavelength ( $\mu m$ )	Bandwith ( $\mu m$ )	Resolution ( $m$ )	Exercise Band
Band 1 (VIS)	0,443	0,02	60	—
<b>Band 2 (VIS)</b>	<b>0,49</b>	<b>0,065</b>	<b>10</b>	<b>1</b>
<b>Band 3 (VIS)</b>	<b>0,56</b>	<b>0,035</b>	<b>10</b>	<b>2</b>
<b>Band 4 (VIS)</b>	<b>0,665</b>	<b>0,03</b>	<b>10</b>	<b>3</b>
<b>Band 5 (RE)</b>	<b>0,705</b>	<b>0,015</b>	<b>20</b>	<b>4</b>
<b>Band 6 (RE)</b>	<b>0,74</b>	<b>0,015</b>	<b>20</b>	<b>5</b>
<b>Band 7 (RE)</b>	<b>0,775</b>	<b>0,02</b>	<b>20</b>	<b>6</b>
<b>Band 8 (NIR)</b>	<b>0,842</b>	<b>0,115</b>	<b>10</b>	<b>7</b>
Band 8A (NIR)	0,865	0,02	20	—
Band 9 (NIR)	0,94	0,02	60	—
Band 10 (SWIR)	1,375	0,02	60	—
<b>Band 11 (SWIR)</b>	<b>1,61</b>	<b>0,09</b>	<b>20</b>	<b>8</b>
<b>Band 12 (SWIR)</b>	<b>2,19</b>	<b>0,18</b>	<b>20</b>	<b>9</b>

## 13.2 Legend & Visual Interpretation

The interpretation process involves to visually examine the image dataset that will be use in the classification, and define a set of classes of interest. In this case we will focus on the definition of classes of land cover. For example, *forest, buil-up areas, water bodies, grass fields, etc.*


### 13.2.1 Interpretation Legend

The interpretation legend defines a list of *legend units* in terms of the characteristics of the image dataset. For example, one unit can be define as *all black and dark-blue pixels* in the image.

**Task 2** Start QGIS and open the `s2_25_sept_2016.tif`, display it with the band combination **7,8,3**.

Define between 5 to 10 legend units for this dataset. After we established a legend, we can start with the visual interpretation.

## 13.2.2 Visual Interpretation

One the purposes of  **Visual image interpretation** is to identify objects in the image dataset. For this we use the legend units defined in above and the ability of the human eye to recognize certain objects. For example, a legend unit defined as *black and dark-blue pixels* can be recognized, after considering the size, shape and texture of the objects , as **water**.

**Attention: Question.** Base on your knowledge on the image characteristics and EM radiation, **which classes of land cover can you name and identify?**. If your now sure use the **QuickMapService** to load the standard map from Open Street Maps.

---

## 13.3 Field Survey

The purpose of the fiel survey or *fieldwork* is to conduct a survey to determine what the *legend units* based on the image characteristics are in reality. In this case, what type of land cover and land use they represent.

During the survey we have to visit and check different locations for all the legend units. Then, we record our observations at each location. *The locations for the survey should be chosen in such a way that all legend units are covered.*

In this course, it is not possible to conduct a fieldwork. Instead, we provide you with a dataset with observations made during a fieldwork for interpreting the Sentinel-II image in Enschede, NL. The datasets contains locations and photos for each visited site. If you were to conduct a field surves, you should collect similar data.

**Task 3** Add the `training_areas.shp` to QGIS. Use the **Identify tool** to inspect the location of visited sites during the fieldwork over the 's2\_25\_sept\_2016' image. Put attention to the attributes **TA\_ID**, it tells the file name of the photo(s) that belong to that location; for example. *AV\_1*. If more than one photo is available for a location, files names contain also a literal; for example *AV\_1a* and *AV\_1b*. See [Fig. 13.2](#)

Fig. 13.2: Field data: training data and photos.

---

## 13.4 Organising Field data

After fieldwork, we have to structure the fieldwork data, define classes and relate such classes to the image. Thus we have to define **field classes** using the fieldwork data. Filed classes are classes recognizable in the field based on certain criteria. In our case the creteria is related to land cover and land use. Field classes should define at same level of detail.

**Task 4** Compare the field classes, as defined in the attribute **Landcivr** of the 'training\_areas' layer, with the colours on the satellite image when displaying bands 7,8,3. Make a list of up to 10 map classes you would like to classify, and write down which colour(s) corresponds them. See the example in the table below.

Class	Colour (7-8-3)
Water	Black/dark blue

## 13.5 Defining Mapping Classes

In the previous task, you practiced how to define field classes and associate colours. However, for the remainder of this exercise, we will use the following list of macroclasses (MC\_info) and ID codes (MC\_ID).

MC_ID	MC_Info
1	water
2	maize
3	evergreen
4	grass
5	bare
6	buildings
7	tarmac
8	heath
9	deciduous
10	rapeseed

It is *fundamental* that you follow the nomenclature above, because the accuracy assessment assumes that the classification results contain such classes.

---

## 13.6 Defining Spectral Signatures (ROIs)

**Note: QGIS.** The QGIS **Semi-automatic Classification** plugin uses the term ‘Region of Interest’ (ROI) instead of ‘training sample’ which is the term you will see in the reading materials. *These terms are equivalent and have exactly the same meaning.*

---

**Task 5** Install the plugin **Semi-automatic classification**. If necessary, open the `s2_25_sept_2016.tif`.

**Task 6** Use the **Semi-automatic classification** plugin to create the first training sample. While you are at it, follow these recommendations:

- Make sure that you take samples from homogenous areas, and that the *standard deviation* is kept low (check the statistics of the ROI).
- It is advisable to make several subclasses for the same macroclass. For example, you will notice that not all pixels with water have exactly the same colour. Some areas with water have black pixels and some others have dark-blue pixels. Instead of collecting samples under the macroclass *water*, it is better to split water in two subclasses, for examples *water\_black* and *water\_blue*, and take samples for each subclass separately. The set of all pixels associated to an specific class is what we call a **spectral signature**.
- Give subclasses distinct class names (e.g. *grass\_yellow*, *grass\_orange*).

Watch the video tutorial on [Creating training sets](#) to know how to complete this task in QGIS.

**Note: QGIS.** We will apply a classification based on macroclasses (information class), and not on subclasses (spectral classes). When we classify according to macroclasses, the classification algorithm will group subclasses and label the results according to macroclass. However, *the classification algorithm will not merge the spectral signatures*.


---

**Attention: Question.** Why is it advisable to make separated (spectral) subclasses when taking training samples?

**Task 7** Define ROIs for all macroclasses in the table above and for all relevant subclasses. Remember to group subclasses with the correct macroclass.

---

## 13.7 Evaluation of Spectral Signatures

You can visualise spectral signatures (ROI) statistics. One option to evaluate signatures (ROI) are  **feature space images**, which are two-dimensional histograms.

**Task 8** Create and analyze feature spaces. Click on *Add highlighted items to scatterplot* as shown in minute 11:15 onwards of the video tutorial on [Creating training sets](#). Inspect the feature spaces and evaluate the separability of your ROIs.

**Attention: Question.** What band combination creates a suitable feature space to evaluate the separability of your training samples?

To evaluate the quality of the training samples, we can use the standard deviation as reference.


**Task 9** Analyse the standard deviation for each of your ROI. Click on *Add highlighted signatures to spectral signature plot > Signature details*.

When you are satisfied with the signatures (ROI) of your training samples, you can move to perform a digital image classification.

---

## 13.8 Supervised Image Classification

In this exercise, we will use the **Maximum Likelihood** classifier for the supervised classification, but we encourage you to experiment with other classifiers and compare the results.

**Task 10** Use the **Semi-automatic classification** plugin to test your first training sample set. Configure the 'Classification algorithm' as shown in [Fig. 13.3](#). Then use the **Active classification preview pointer**  to generate a temporary classification. Then, visually inspect the classification results.

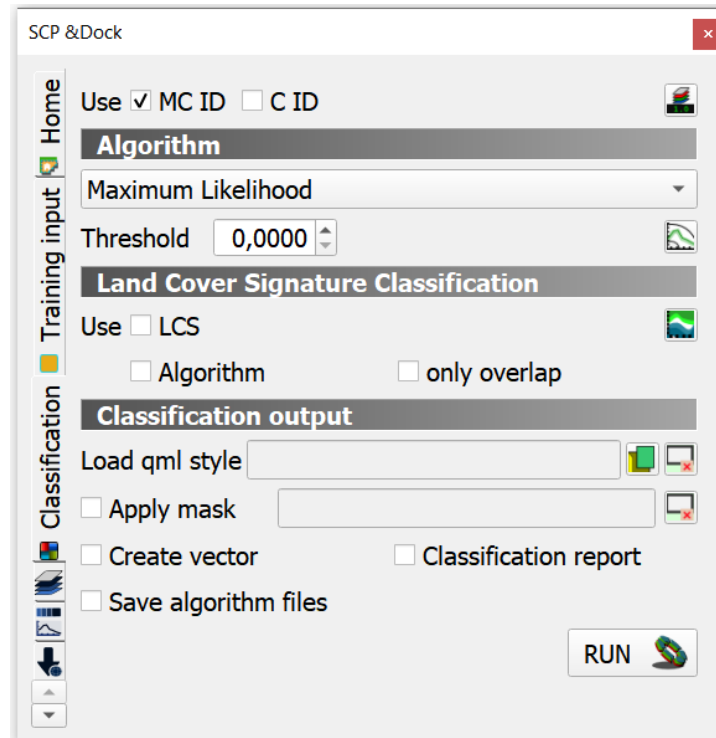



Fig. 13.3: Setting a classifier

Watch the video tutorial on [Running a supervised classification](#) to complete this task.

### 13.8.1 Inspect Results & Improve the Classification

**Task 11** Based on the visual inspection of your first classification results, add additional ROIs that capture areas where the algorithm fails to classify the image correctly.

Then, inspect the classification results again, and if necessary add more ROIs until you are satisfied with the results.

**Task 12** Run the final classification, by clicking on the button 

**Task 13** Generate a shapefile from the ROIs you used to classify the image. [HOW? GIVE SOME HINT?]

**Task 14** Make sure you save your training samples for later use. In the **Semi-automatic classification** plugin, this is done automatically.

## 13.9 Accuracy assessment

---

**Important: Resources.** In this section, you will require the dataset [reference\\_data.zip](#). This dataset contains the following file:

- `reference_data.shp`, dataset containing the locations and types of land cover for the accuracy assessment.
- 


### 13.9.1 Data for Accuracy Assessment

For this exercise, the data for the accuracy assessment have been collected for you. The `reference_data.shp` contains the **ground truth** information of 187 locations, that we will use in assessing the accuracy of the supervised classification.

The areas in the `reference_data.shp` were randomly selected. Areas which fell on the edge of two or more classes were considered.

**Task 15** Add `reference_data.shp` to QGIS and open its attribute table. Make sure the values for **MC\_ID** and **MC\_Info** attributes match the ones you used for the macroclasses during the supervised classification of the 's2\_25\_sept\_2016' image.

### 13.9.2 Accuracy Report

In the accuracy assessment, we generate an  **Error matrix** between the classification results and the reference data. The accuracy report will provide the *overall accuracy* of the classification, as well as *user and producer accuracies*.

**Task 16** Perform the accuracy assessment. Watch the video tutorial on [Accuracy assessment of supervised classification](#).

### 13.9.3 Improve Accuracy

The accuracy of the classification can be improved by identifying the sources of errors in the process of supervised classification, and then minimizing their effects on the classification results.

**Task 17** Locate and inspect areas which you suspect are sources in your classification result. Then, try to improve the *overall accuracy* by adding or excluding ROIs from your training dataset.


Once the accuracy of the supervised classification is *acceptable*, you have completed the digital image classification process and generated a **land cover map**. Save all your results.

**Attention: Question.** What determines if the accuracy of the digital image classification is acceptable or not?

*Section author: Monika Kuffer, Andre Mano & Manuel Garcia Alvarez*




## SYSTEMS AND MODELS

In this section, you will get familiar with some of the basic concept of  Systems Modelling.

---

### 14.1 Systems



A  **System** is defined as “a combination of parts that interact and produce some new quality or function in their interaction.”

A System consists of:

- Parts or elements
- Interaction between these elements
- Emergent properties

**Attention: Question.** Think about an example of a *system* and answer the following questions:

- What are the **elements** in your example of a system?
- Give some examples of **interactions** in your system. Then, identify if the interactions are related to a *flow of material* or a *flow of information*.

 **Interactions** between elements can have many different forms. Interactions may form loops. When *A* impacts *B*, while at the same time *B* impacts *A*, we say that there is a  **feedback in the system**. We distinguish **positive** and **negative** feedback loops.

An example of a *positive feedback* is population growth. The larger the population, the more people are born; the more people are born, the larger the population. In a *negative feedback* loop the system stabilises itself according to the rule: *the larger something is, the smaller something becomes*. For the population example, if there is a limited supply of food, and the population grows, there is less food for each individual.

**Task 1** The table below contains a few examples of *interactions*. Identify if such interactions result in *positive* or *negative* feedback.

Example	Type of Feedback
The more we study, the more tired we get, and the more difficult it is for us to understand the content of the course, therefore the more we have to study because we learn less.	
The more sheep we have, the more wolves we get because they have more food; the more wolves we get, the more sheep are eaten which means there is less food for each wolf and wolf population decreases.	

Systems and their elements do not always change immediately in reaction to internal or external influences. When the effects of changes in some of the elements of a system, do not affect other elements immediately, we referred to them as the *delayed effects*.

**Attention: Question.** Can you think of an example of a delayed effect in a system?

Although thinking about the different elements and interactions of a system may seem very abstract. Such activity provides the starting point for building models, and models help us to gain understanding about the world we live in.

---


## 14.2 Models

Modelling is a very broadly used term, which may cause “confusion”. It can be used at different abstraction levels, and it may refer to modelling different types of information or systems. Many things we do or products we produce can be regarded as a model. If we take a picture of a landscape, this can be regarded as a model. A printed map is also a model. In the exercise on *Data Modelling*, you created a model of reality as a point-vector layer or a continuous raster.

The models we refer to, in this section, describe processes that happen on the Earth’s surface at various scales. Such models aim to represent systems. Modelling the Earth as a system may require to capture its natural components. For example, we could portrait the system Earth as the combination of four subsystems: *atmosphere (air)*, *lithosphere (land)*, *hydrosphere (water)* and *biosphere (life)*. However, we may also model the Earth as human-induced systems, such as *urbanisation*, *human health* and *agriculture*.

**Attention: Question.**

- Can you define a system that takes place on the Earth’s surface, and it is interesting for studying during your MSc research project?.
- What makes such a system relevant or interesting?


Many modelling processes start with a conceptual  **Model**. A conceptual model is often written on a piece of paper, but it can also be a mental picture of some phenomenon.

You can create a conceptual model by following these steps:

1. Define the goal (write it on the bottom of the page)
2. Describe the input of your model (write it on the top of the page)
3. Fill in all the steps (processes) needed to go from input (top) to goal (bottom)

**Task 2** In many universities, they teach conceptual modelling by asking the students to create a conceptual model for making a cup of tea or coffee. *Create a conceptual model for making tea or coffee.*

A conceptual model serves as the first step towards creating such a computer model. A conceptual model for making tea represents a non-spatial model. However, as a geo-scientist are interested in spatial and temporal models. You are also interested in implementing your model using modelling software, using your spatial data.

When  **modelling a system**, the components of a conceptual model includes three elements: *boundaries, variables and parameters*.


**Attention: Question.** For the conceptual model, you created in Task 2.

- Can you define the boundary of the model, a variable and a parameter?
- Is the variable you defined, a state variable?

## 14.3 The Purpose of a Model

Models are used in many application, and they can be classified in various ways. Here, we recognise five ways of classifying models in GIS-based applications based on their characteristics:


- Purpose of the model
- The methodology underlying the model
- Scale
- Dimensionality
- Implementation logic

The *purpose* of the model refers to whether the model is **descriptive, prescriptive or predictive** in nature. The *methodology* underlying the model refers to its operational components. Here, we have two groups: **stochastic** models and **deterministic** models. Read about models according to *scale, dimensionality, and implementation logic* in  **System Modelling**.

**Attention: Question.**

- An example of a deterministic model could be the hydrological flow in a watershed. **Why do you think hydrological models are deterministic?**
- When a model is stochastic, we usually run such model many times with the same input parameters, and present then present the average or a range of the modelling results. **Why is this necessary in stochastic models?**

## 14.4 Modelling Processes

The process of building models is more complicated than defining their elements and relationships, capturing them in a conceptual model, and implementing such conceptual model in software, see Fig. 14.1 The  **Modelling** is an **iterative** process.

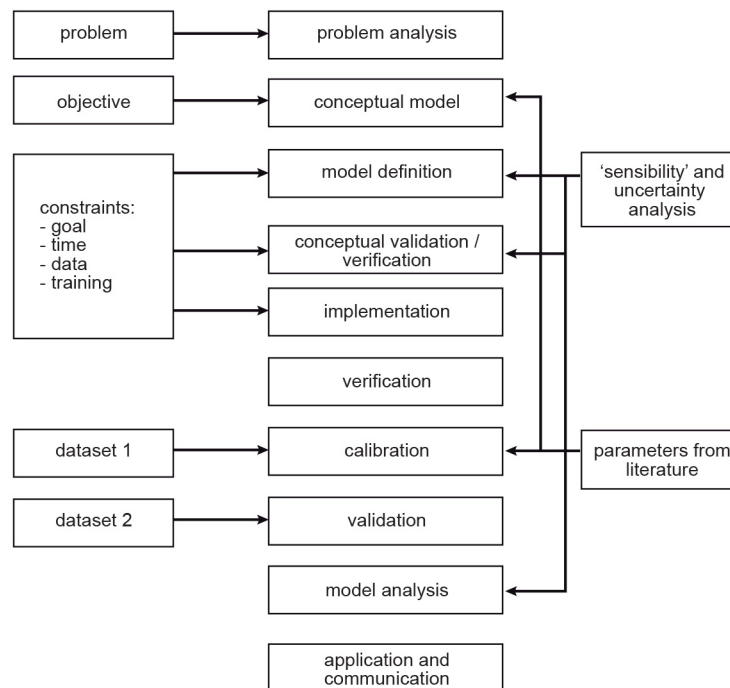


Fig. 14.1: A summary of the steps in the modelling process

**Task 3** Define in your own words, why the modelling process is iterative? and what will induce you to make another iteration? or when would you stop iterating a model)?

Four of the steps in Fig. 14.1 deserve some attention:

- Verification
- Calibration
- Validation
- Sensitivity analysis

**Verification**, in the context of computer simulation, is the process of confirming if the conceptual model was implemented correctly. It may seem obvious that when you made the conceptual model, you know how to implement it. Yet, a lot of things can go wrong; therefore, we need verification. **Calibration** is the process of adjusting the model parameters.


**Task 4** Give an example of a problem during the implementation of a model that can be detected during verification.

When you look at Fig. 14.1 you see that there is a line between calibration and model definition. **Validation** is the most challenging task; validation is the process of proving that the model is a good representation of reality. Many models are not fully validated. Note that the diagram shows that calibration and validation use two different datasets.

**Attention: Question.** Why it is a good idea to use different datasets for calibration and validation?

**Task 5** Below you see a table with examples of models. Indicate if they are *static* or *dynamic*.

Example	Static or Dynamic?
An animated map	
A film	
A simulation model	
The implementation of the tea model	
A model of the evacuation of a building	

The  **Implementation logic** of a model refers to two groups of approaches: *deductive* approaches and *inductive* approaches. To explain the difference between those groups, we will use an example from education. The concepts ‘deductive’ and ‘inductive’ can also be applied to learning and teaching. In a *deductive approach*, the teacher presents students a new concept, explains how it works, and then makes the students practice using the concept. In an *inductive approach*, the teacher presents students many examples that show how the concept is used; the students should ‘notice and realise’, by way of the examples, how the concept works [OB2009].

The table below shows some differences and examples for inductive and deductive approaches.

Deductive Modelling	Inductive modelling
For the Verification of hypothesis (top-down)	Is bottom-up – hypothesis are developed from the data
Hypothesis is defined by the researcher	E.g., machine learning
Common in models used for description	Can be used for prediction

**Attention: Question.** When would you prefer to use an inductive over a deductive approach?

## 14.5 Evaluating Systems Models

The best way to learn about systems modelling is to experiment with a number of existing models and determine which characteristics they replicate. In this section, you will use the web version of Netlogo. Netlogo is a software designed to create agent-based simulation models and Cellular Automata (CA) models. You will see some examples and then reflect on their similarities and differences.

**Task 6** Open the web-version of [NetLogo](#). Then, click on *NetLogo Web*. You can load an existing model by selecting it from *Search the Models Library* dropdown list. See [Fig. 14.2](#)

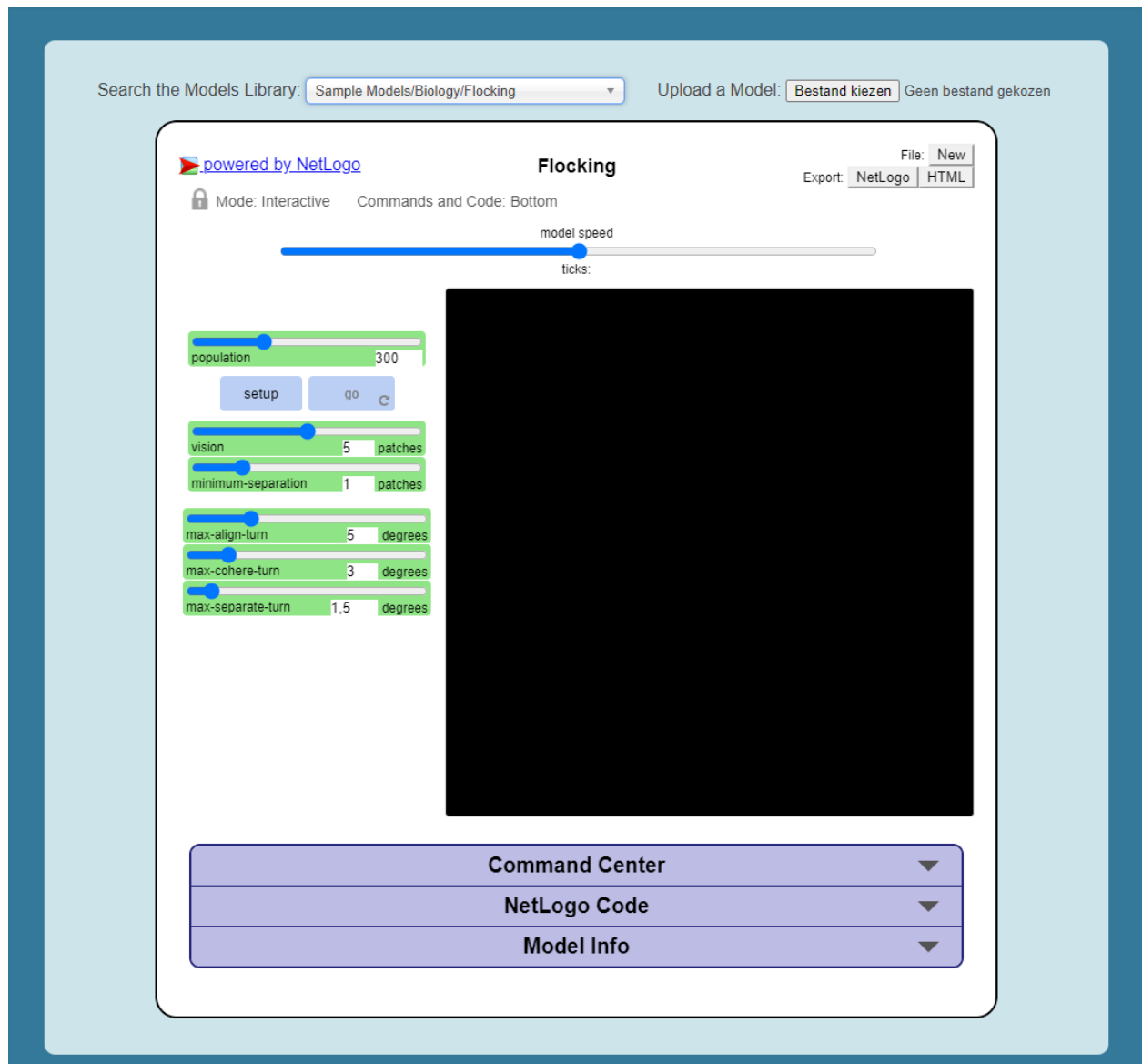



Fig. 14.2: The NetLogo web interface

Now, we will experiment with the following models:

- Flocking (Sample Models - biology),
- Wolf Sheep Predation (Sample models - biology), and
- The CA 1D Rule 30 (Sample Models- Computer Science)

**Task 7** We will start with the Flocking. This model demonstrates how spatial patterns (Flocks) emerge based on three simple behavioural rules: *alignment*, *separation* and *cohesion*. Go to *Search the Models Library* > search for *Sample Models/Biology/Flocking* > click *setup* > *go*.

---

**Note: Reflection.** The Flocking model is an  **agent-based model**. Agent-based models consist of three components: *agents* (with behaviour), the *environment* in which these agents exist and *time*.

---

**Attention: Question.** Open the *Model Info* tab underneath the model window. Answer the following questions regarding the Flocking model:

- Is this a spatial model? Explain your answer.
- Do you think this model contains any agents? If so, which agents and how are they represented in the model?
- What happens when you change the *minimum-separation* from 1 to 2 patches?
- If you were to calibrate this model, which elements would you calibrate?

The second type of model we will evaluate is a CA model. Before we open the example model in Netlogo, you will first practice how this type of model works via a pencil and paper exercise.

**Task 8** Below you see a lattice (raster) where each cell has a value. This raster represents the starting values of a CA model simulation. With each iteration of the model, the value of the cells will be determined by considering their 8 neighbouring cells, and applying the following rules:

- If the current value is 1. Change to 0 if less than four neighbours are 1
- If the current value is 0. Change to 1 if less than four neighbours are 0

For the raster on the left, *compute the values for the next iteration of the model. Follow the rules listed above.* Write down the values on the raster on the right.

1	0	1	1
1	1	1	0
0	0	1	1
1	1	1	0


**Attention: Question.** One of the items of CA models is the *state variable*.

- Do you think the model presented in the previous task has a state variable?
- Do you think the model presented in the previous task will ever reach a *steady-state*, where the output does no longer change?

Now that you have gained some understanding of how a CA model works, it is time to experiment with a CA model in NetLogo.

**Task 9** Open the model *CA 1D Rule 30*, check the *Model Info*, and run the model. Then, describe how this model differs from the pencil and paper model above.

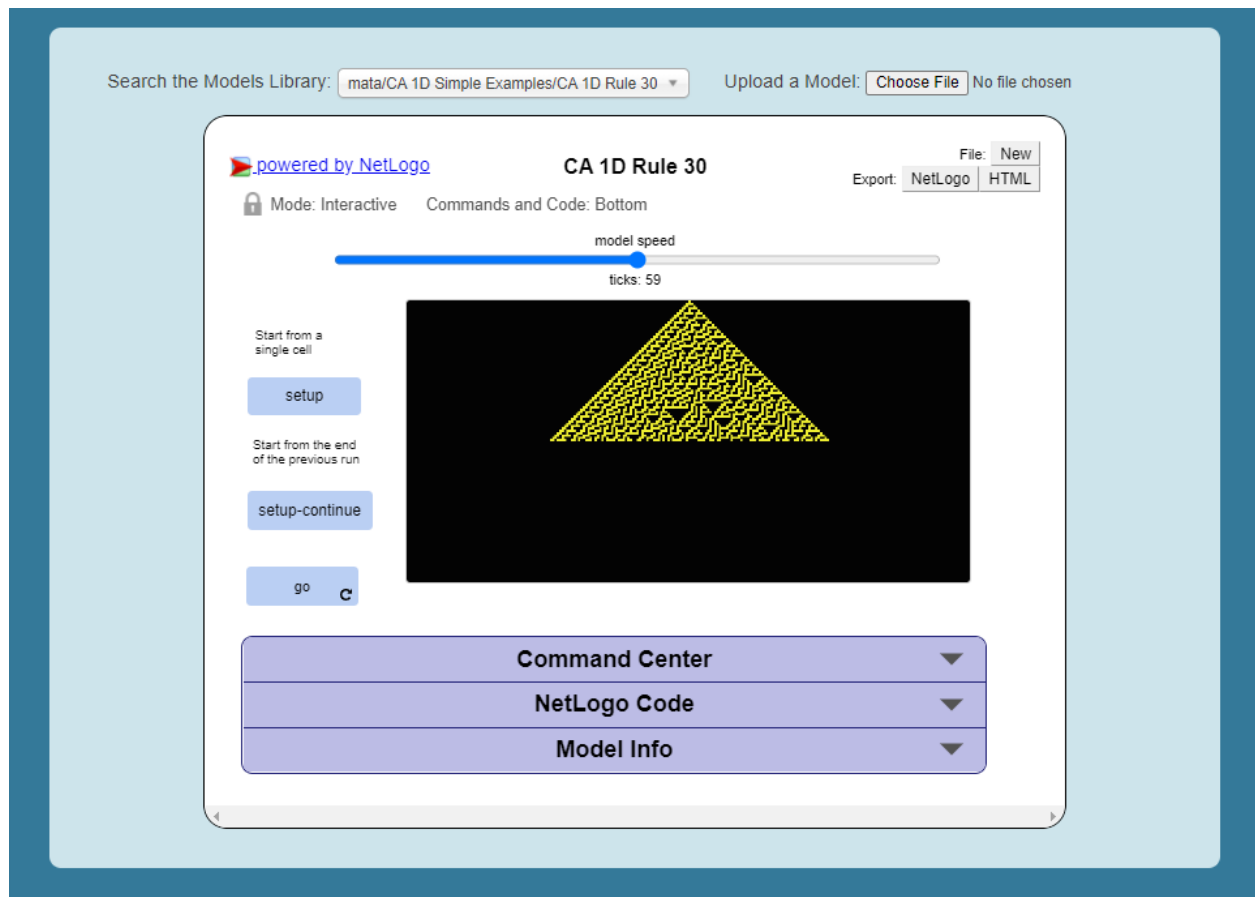


Fig. 14.3: The 'CA 1D Rule 30' in NetLogo

**Attention: Question.** This 'CA 1D Rule 30' model represents a very "synthetic" implementation of a CA model. Can you think of an example of a CA model that is more realistic and aligns with your interests?

**Task 10** Finally, we will explore 'Wolf Sheep Predation' model. Go to NetLogo and load *Wolf Sheep Predation > model-version*, choose *sheep-wolves-grass*. Run the model, and put attention to the simulation.



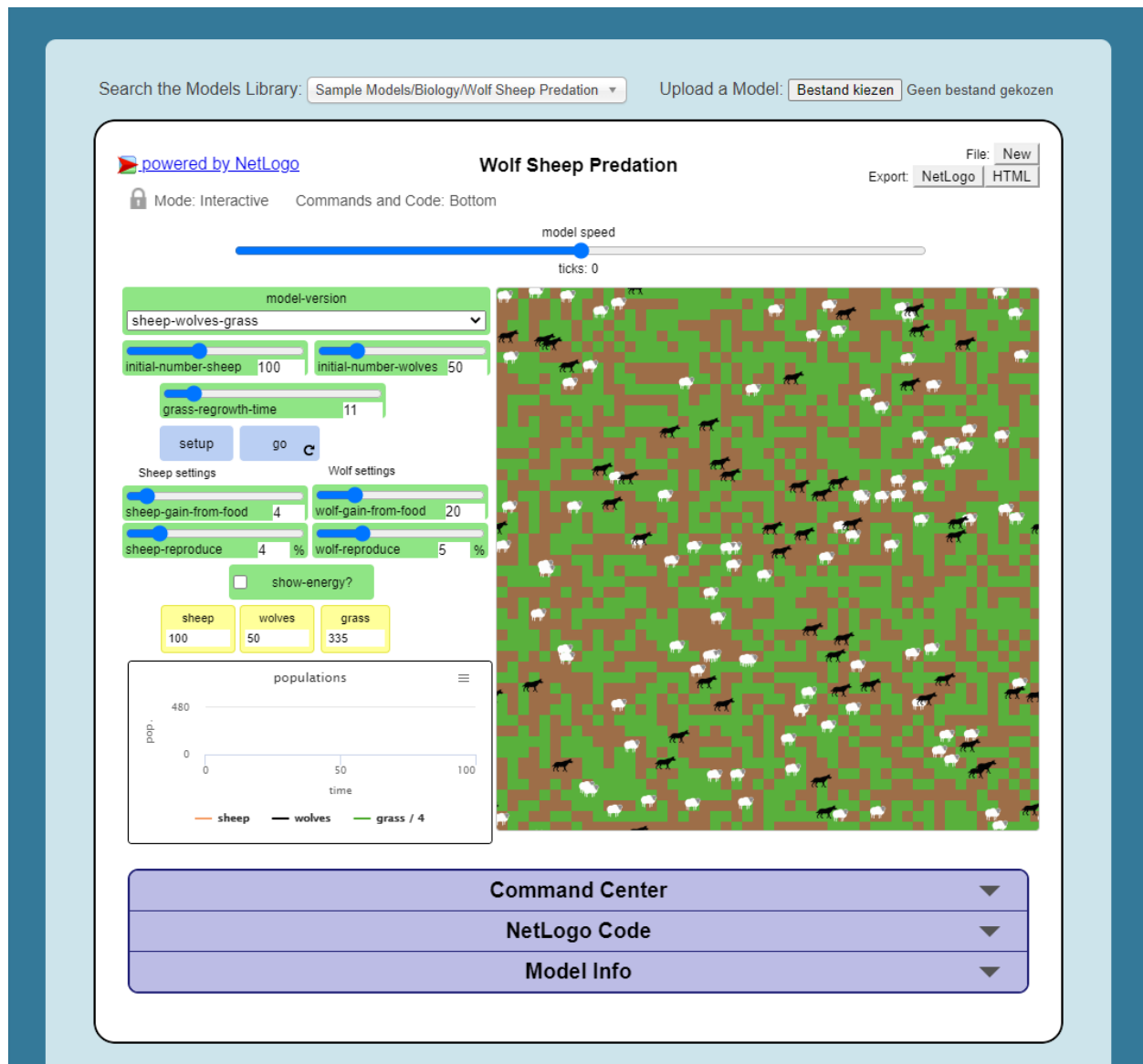


Fig. 14.4: The 'Wolf Sheep Predation' model, version sheep-wolves-grass

In this model, you see some similarities with an agent-based model, the '*Flocking*' model; it has agents: wolf and sheep. You also see some similarities with a CA model. The grass has two states: *grass* or *no-grass*, and the way the grass regrows is triggered by rules that are similar to the transition rules typical of a CA model.

**Task 11** Open the *NetLogo Code* tab, read the code and to determine how the grass regrows in the '*Wolf Sheep Predation*' model.

**Attention: Question.** Run the '*Wolf Sheep Predation*' model once more and answer the following questions:

- Why is this model an example of a model with *delayed effects*?
- Is this model descriptive, prescriptive or predictive? Explain your answer.
- What happens when you increase the energy the sheep gain from eating grass?

- Explain why this model is an example of *counter-intuitive effects*? For example, what happens when you double the number of wolves at the beginning of the model? Does such change improve their chances of survival?

**Task 12** Open the *NetLogo Code* tab again and make changes to the model. To apply the changes to the model, click on *Recompile Code* > *model-version* choose *sheep-wolves-grass* > *setup* > *Go*. Try to make the following changes one at the time.

- Change the colour of sheep to *yellow*.
- Make the wolf *twice as big*.
- Make the grass *blue*. You will have to change the code in multiple places.

After all the changes, the ‘*Wolf Sheep Predation*’ model should look like this:

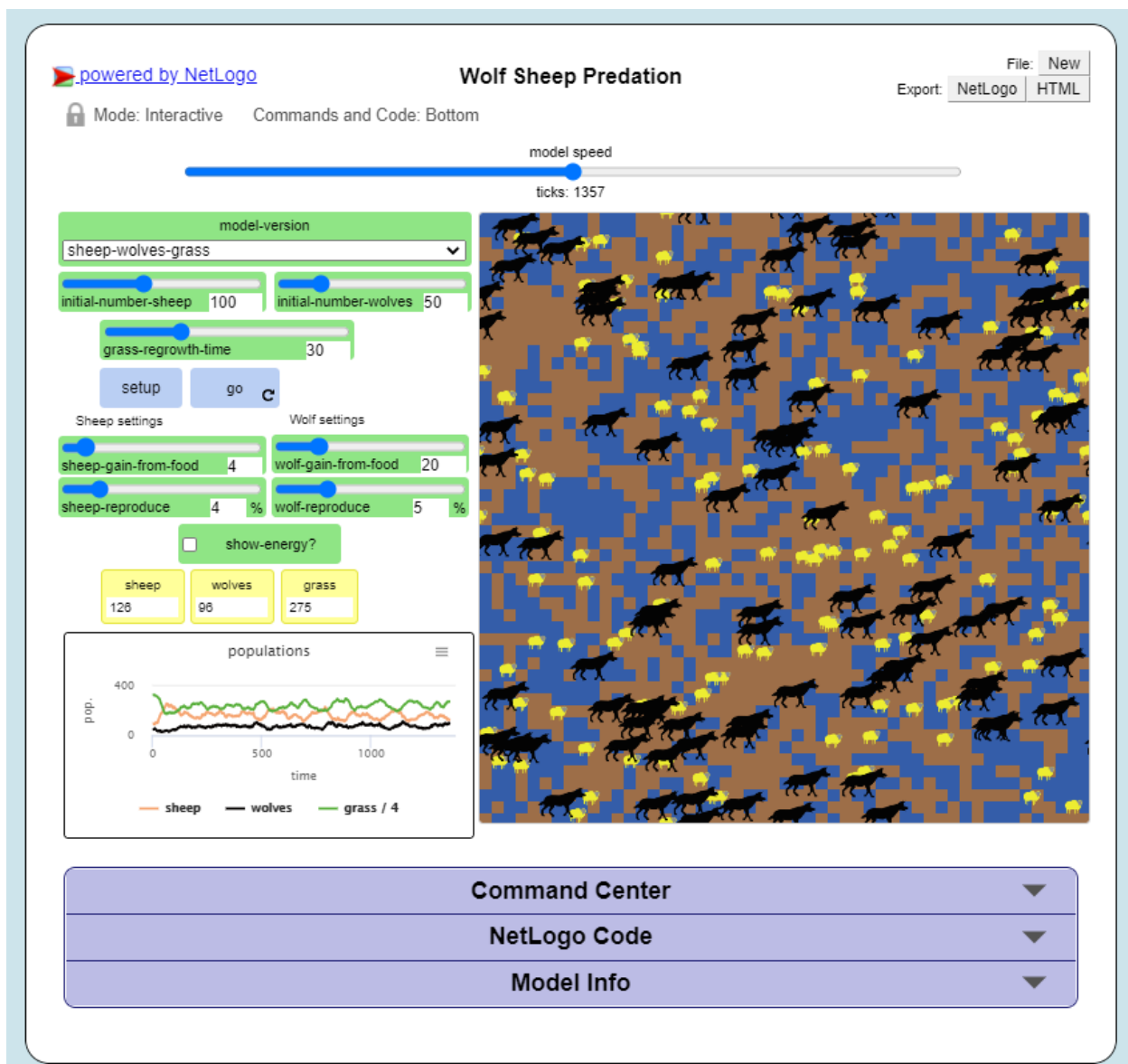



Fig. 14.5: The ‘*Wolf Sheep Predation*’ model showing yellow sheep, bigger wolves, and blue grass

*Section author: Ellen-Wien Augustjin, André Mano & Manuel Garcia Alvarez*



## SPATIAL DATA INFRASTRUCTURE

A  **Spatial Data Infrastructure** (SDI) refers to the collection of technologies and policies to make geospatial data accessible across institutions and countries.

An SDI aims to coordinate the collection of various geospatial data, and to foster their integrated use at different spatial levels: *local, regional or global*.

It is essential to realise that implementing an SDIs is not just a matter of technological requirements. The societal dimension is equally important, such as making geospatial data accessible to the general public so that everyone can benefit from it.

**Geoportals** are a key component of SDIs; they provide a way to find and access geospatial data via the Internet.

---

### 15.1 The PDOK Geoportal

In this section, we will explore the *Publiek Dienstverlening Op de Kaart* (PDOK). These geoportals provide access to geospatial data in the Netherlands. The datasets available through these geoportals can be accessed mostly as OGC-based services (Open Geospatial Consortium) services, but sometimes also as direct data downloads.

The PDOK integrates the metadata from the **National Registry** (NGR), and several tools and web services that easier to find and use geospatial data from many government agencies in the Netherlands.

For a quick overview on the PDOK geoportal, you may want to watch the **PDOK video** (audio in Dutch). Activate the subtitles *CC*. Then, enable auto translation: *Settings > Subtitles/CC > Auto-translate > English*.

---

**Important: Resources.** You will require the latest LTR version of LTR version of **QGIS**.

This exercise does not require any data files. However, you need a web browser. We recommend **Google Chrome** because you will need to translate some webpages to English.

---

**Task 1** Start a web browser and navigate to the **PDOK website** and explore its contents. Take some time to explore the website's interface.

**Task 2** In the PDOK home page, click on *Bekijk alle dataset* or *See all datasets*. See **Fig. 15.1** This will take you to a webpage where you can search the NGR metadata repository.

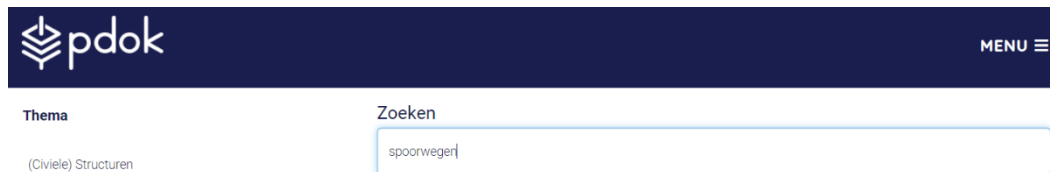


Fig. 15.1: The PDOK home-page

**Task 3** Suppose we need data about the *railroad network* in the Netherlands. Then, we can use the search bar to find all available datasets.

**Note: Reflection.** Any good search starts with the right choice of keywords. For this task, it is very relevant to know that, the railroad datasets are part of the Dutch national transport database, called “*Nationaal Wegenbestand*”; abbreviated as ‘*nwb*’. The Dutch word for railroads is ‘*spoorwegen*’.

Type **spoorwegen** in the search bar and press *Enter*.



**You should see the result of the search below the search bar.** Whenever relevant, translate the pages to English. In Chrome, look for the **translate** icon, then *translate to English*.

**Task 4** Under *Spoorwegen*. Click on *Details bekijken* or *View details*, to see detailed information about the dataset. See Fig. 15.2

## Spoorwegen

Deze service bevat de spoorverbindingen, inclusief kilometeringspunten, stations en overwegen.

### Thema

Transport

### Organisatie

ProRail

### Ontsluitingen

Geo Services

**Details bekijken**

Fig. 15.2: Search result for ‘spoorwegen’

**Attention: Question.** Why do you think the description of dataset (*Omschrijving*) is only available in Dutch?

**Task 5** Click on the *Geo Services* tab. You will see a list of geo-services available for this dataset. Every geo-service mentioned a **URL**. See Fig. 15.3. This is the most important piece of information you will need. You can use such URL to access the data using several tools: a *web browser*, a *GIS software*, or a *program* that you code yourself.

For example, if you want to see the description of one of the geo-service in the web browser. You can simply copy and paste the URL in a new tab and hit *Enter*.

## Spoorwegen (WMS)

Type	wms
URL	<a href="https://geodata.nationaalgeoregister.nl/spoorwegen/wms?request=GetCapabilities&amp;service=wms">https://geodata.nationaalgeoregister.nl/spoorwegen/wms?request=GetCapabilities&amp;service=wms</a>
Metadata Service	<a href="#">Bekijk de metadata</a>
Metadata Data	<a href="#">Bekijk de metadata</a>

Fig. 15.3: Geo-service of the 'Spoorwegen' dataset in the PDOK geoportal

**Task 6** *Open* a new tab in your web browser, *Copy* the URL for the Spoorwegen (WMS) service, *Paste* the URL in the *Address bar* and hit *Enter*. A page containing a description of the geo-service will appear. What you see is a description of **capabilities** of the geo-service in XML (*Extensible Markup Language*). See Fig. 15.4

Do not worry if you do not understand what you see. The *capabilities document* is not meant to be read by humans; rather it should be understood by computers across the Internet.

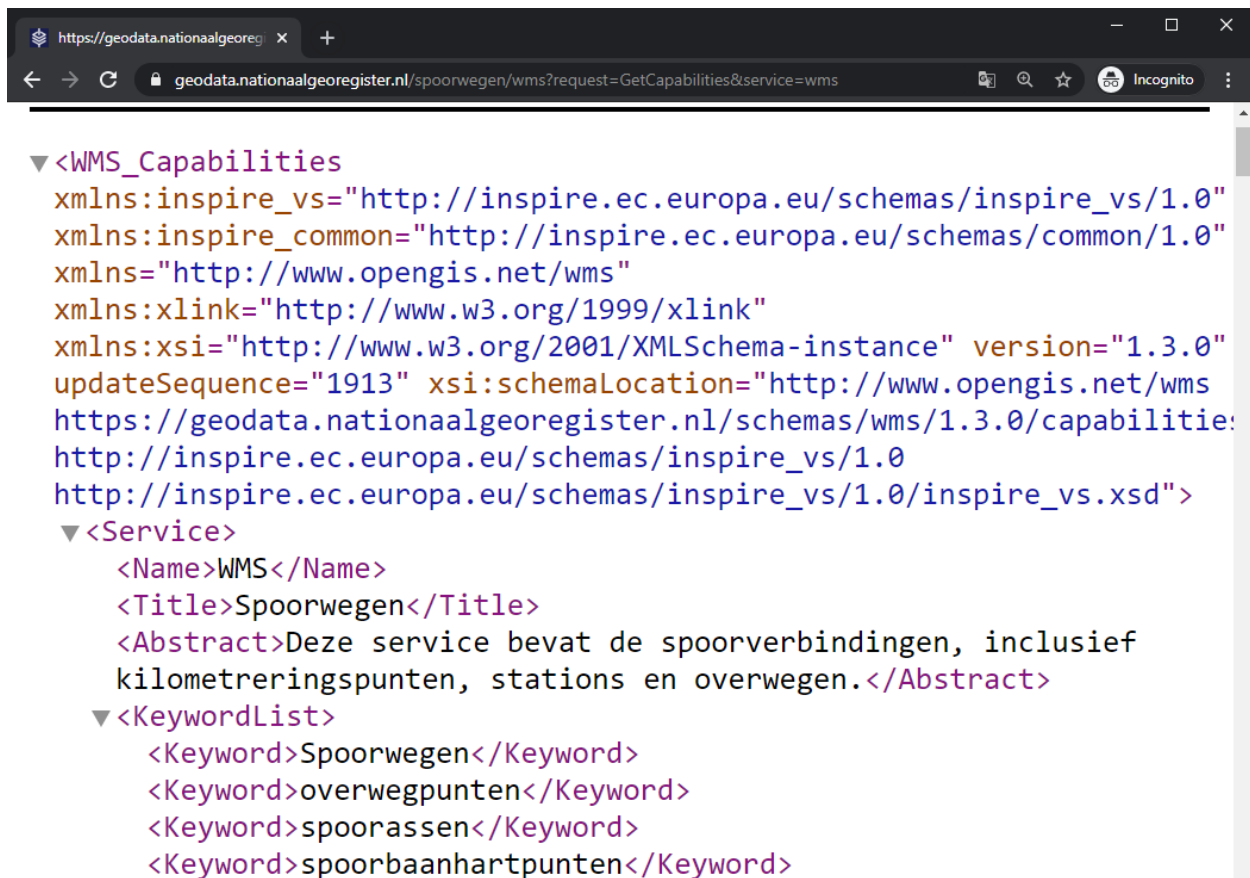


Fig. 15.4: 'Get-capabilities' XML document for the 'Spoorwegen' dataset

**Note: Reflection.** The eXtensible Markup Language (XML) is used to transfer information in the web. Such information is organised using **tags**. In XML, opening tags are declared using `< >`, and closing tags using `</ >`.

Information is enclosed between tags, such as `<TagName>info to be transfer</TagName>`. For example, if we were to write the name of this course in XML, we would do something like this:

```

<CourseName>
  Data Integration: Principles, Approaches and User Perspectives
</CourseName>

```

**Attention: Question.** Check the information in the *get capabilities* document for the 'spoorwegen' and answer the following questions:

- In which CRSs (Coordinate Reference Systems) can this be retrieved?
- What is the spatial extend (i.e. *Bounding Box*) of the data available in EPSG:28992?
- How many layers can this service provide? What are the names of such layers?

We visualise the data available in this geo service, by using adding additional parameters to the root URL, such that the web browser requests the data in the proper format.



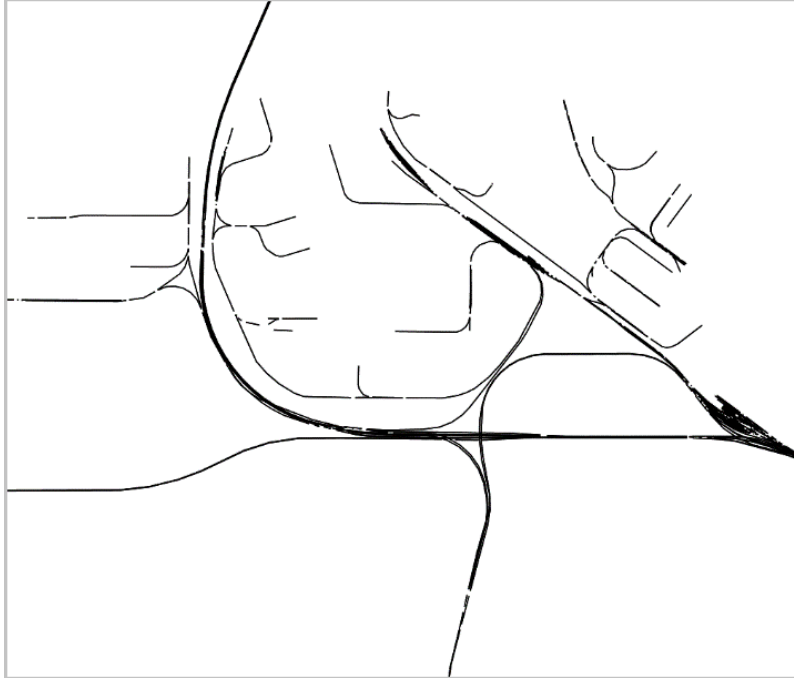
**Task 7** Paste the following URL (an HTTP request) in the web browser and hit *Enter*.

```

1 https://geodata.nationaalgeoregister.nl/spoorwegen/ows?
2 SERVICE=WMS&&SERVICE=WMS&VERSION=1.3.0&
3 REQUEST=GetMap&FORMAT=image%2Fpng&TRANSPARENT=false&
4 LAYERS=spooras&STYLES=spoorwegen%3ASpooras&
5 CRS=EPSG%3A28992&
6 WIDTH=1020&HEIGHT=866&
7 BBOX=113541.12%2C487085.75999999995%2C120395.51999999999%2C492905.27999999997

```

The web browser should show you the following map:



The request above retrieves the railways from an area in The Hague. The `LAYERS` parameter defines the name of the data layer that is displayed, here `spooras`. The extent of the map is controlled by the `BBOX` parameter, which is declared by defining an extent such as  $\min(x), \min(y), \max(x), \max(y)$ . For example:

```

BBOX=113541.12%2C487085.75999999995%2C120395.51999999999%2C492905.
↪27999999997

```

The `%2C` in this case means a `,`.

If you change the values of the parameters in the request, you will get a visualisation for a different part of the dataset.

Change the value for the parameter `LAYERS`, for another one. Use the *get capabilities* document to find a valid name and see what happens. You can also try to modify the extent of the `BBOX`. If you do so, remember to keep the ratio between the coordinates to avoid deforming the image.

**Task 8** Use QGIS to visualise the data provided by this WMS service. To do so, watch the video tutorial on [Using WMS Services](#).

**Attention: Question.** Go back to the PDOK geoportal and answer the following questions.

- What is the purpose of the PDOK facility?
- Who are the targeted user groups?
- Which organisation is ultimately responsible for the Portal?
- What services can PDOK provide for your discipline? Explain.

*Section author: André Mano & Manuel Garcia Alvarez*

## USE AND USERS

The video-lectures stressed the importance of being aware of the requirements of **use and users of geo-information**, in various applications. Only with that knowledge and awareness can geo-information scientists do their research, and develop the models and tools that are needed to tackle the problems related to the Earth as a system.

---

**Important: Resources.** This is a reading exercise. You will need the document [use and users.pdf](#), which you can download from Canvas.

---

The objective of the exercise is to make you aware of the relevance of **users** and **uses** of geo-information. Here, you will read about a case study where geo-information plays a key role. Then you will answer some questions. Finally, you will discuss your answers in the during the *virtual classroom*.

**Task 1** Read the document on *Use and Users*. Answer the questions below, and take your answers to the virtual classroom.

**Attention: Question.**

- Who are the users and stakeholders? What are their roles?
- What are they using? (e.g. what data, information, tools, systems, etc.)
- What are the spatial and temporal scales of the selected application?
- From the perspective of use and users: what are the (potential) difficulties involved? What hinders (or may hinder) an effective, efficient and satisfactory use?

*Section author: André Mano*



## RESAMPLING

In data integration, **resampling** of images is an important pre-processing step, especially when we are dealing with datasets with different spatial resolutions and scales. When the differences in the spatial resolution between the input and the output are *small*, simple interpolation methods are good enough for estimating pixel values. However, when such differences are significant, then we need to apply *aggregation*.

The image you will be working on is a subset taken from band 1 of Landsat 5 sensor, and it has a pixel size of 30 *m*.

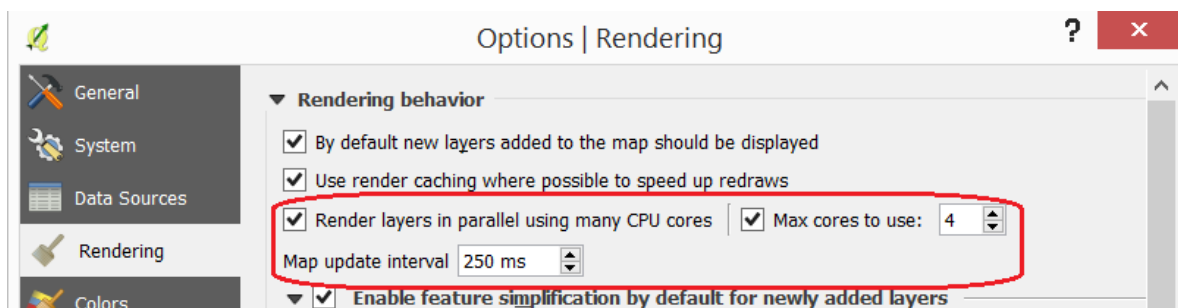
---

**Important: Resources.** You will require the latest LTR version of LTR version of **QGIS**, plus the dataset [resampling.zip](#). When you unzip the dataset, you will find the following files inside:

- 15\_20100627\_band1.tif - an image subset from band 1 of Landsat 5 and spatial resolution of 30 *m*.
  - grid.shp - a vector dataset representing a grid for the original pixel size of the Landsat 5 image.
  - average.txt - a file with a kernel filter definition.
  - Resampling.qgs - a QGIS project file loaded with data layers and styles.
- 

**Task 1** Make sure you have the **Map swipe tool** and **Value Tool** plugins installed.


**Task 2** Configure QGIS to render layers using multiple CPU cores. Go to *Settings > Options > Rendering* and make sure the option *Render Layers in parallel using many CPU cores* is on. Set *Max Cores* to the number of CPU cores in your computer, use at least 4 for better performance. See below.



## 17.1 Basic Resampling

Assume that we are going to use the Landsat image as input for a model. The model requires a grid size of 50 m. Thus we have to resample the image to a pixel size of 50 m before proceeding with the modelling.

**Note: Reflection.** In this section, we make the assumption that applying a resampling before the primary data processing operations (modelling in this case) is the correct procedure. *Note that in many cases it may be more accurate to process the image first and then do the resampling.*

Read the details about resampling on the  [Geocoding](#) page.

### 17.1.1 Resampling Techniques

**Task 3** Start QGIS and open the ‘Resampling’ project. From the *Processing toolbox*, open the tool **r.resamp.interp** and resample the image `15_20100627_band1.tif` to 50 m. Use the **nearest** (neighbour) interpolation method. See Fig. 17.1 Provide a self-descriptive name for the output file, for example: *nearest\_50*

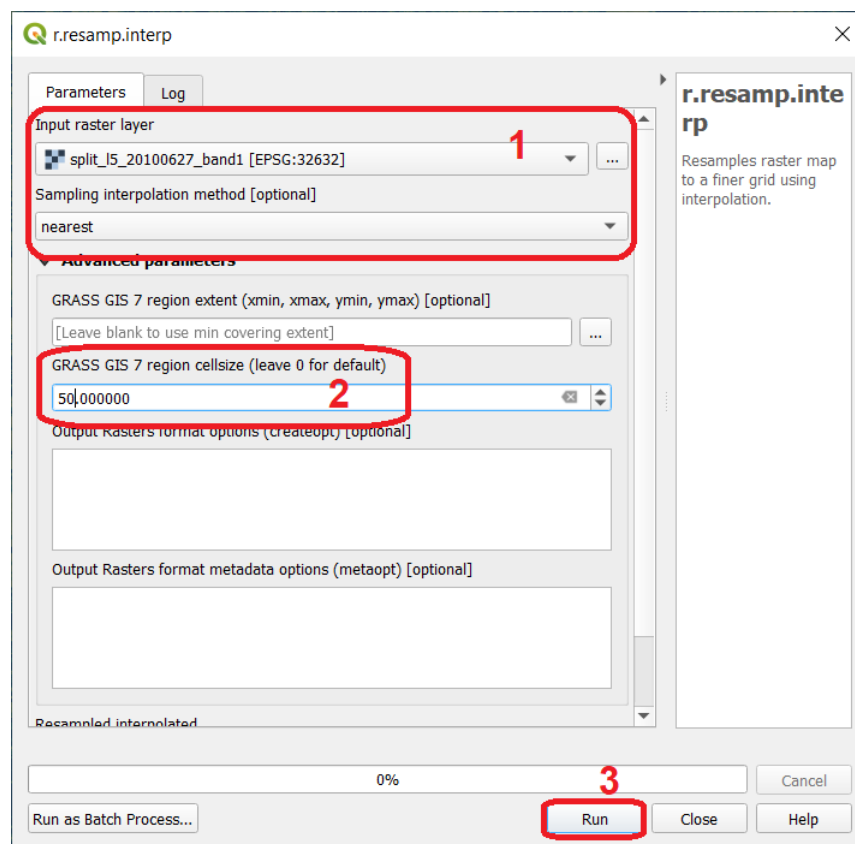


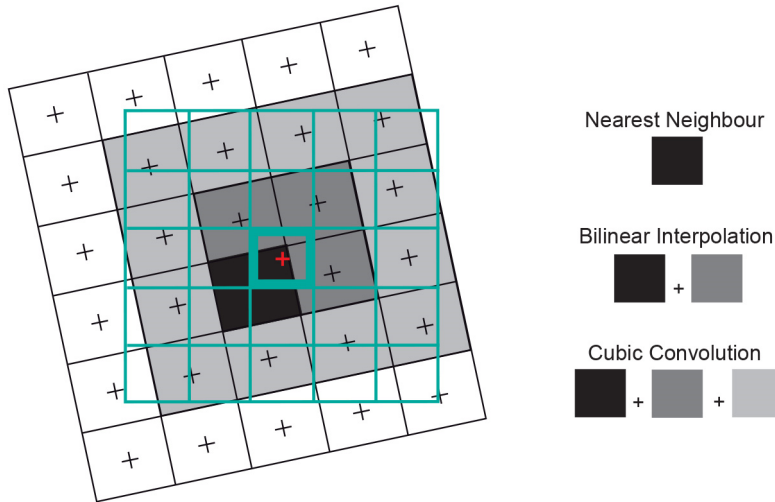
Fig. 17.1: The ‘r.resamp.interp’ tool. Resampling to 50 m using the ‘nearest’ method

**Task 4** Repeat the previous task. This time use the **bilinear** and **bicubic** interpolation methods. Make sure to provide meaningful names for the outputs. For example: *bilinear\_50*, and *bicubic\_50* respectively.

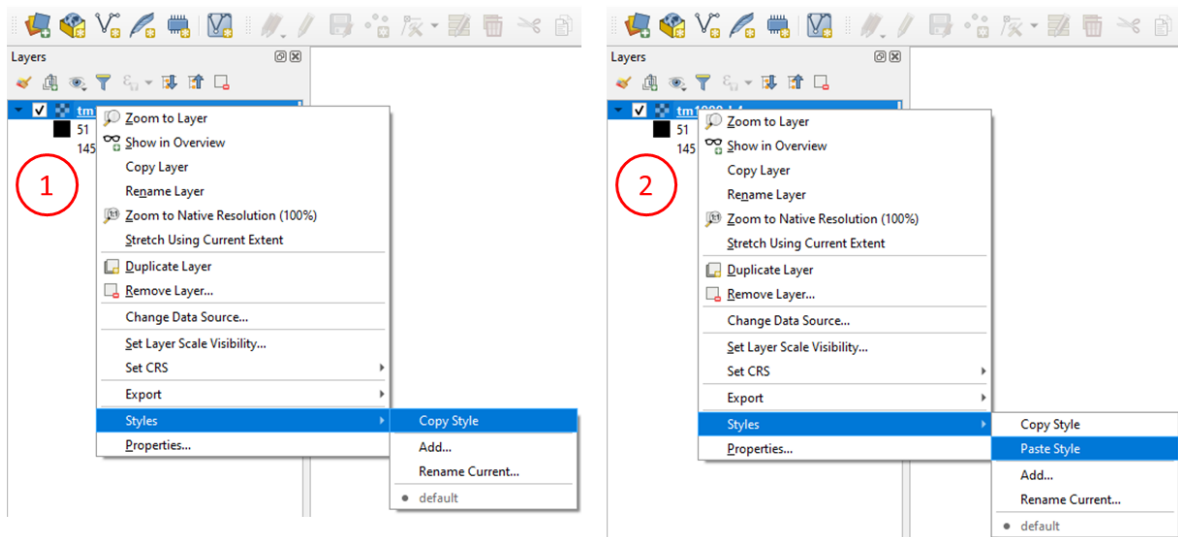
**Attention: Question.** Resampling relies on interpolation techniques, and therefore it depends on Tobler's first law of geography: *locations that are closer together are more likely to have similar values than locations that are farther apart.*

Keeping that law in mind and look at the image below. It shows (for the pixel marked in red) which neighbouring pixels are considered in every interpolation technique.

Can you think of a use case where one resampling techniques should be preferred over the others? Why?



**Task 5** Copy the style of layer 'l5\_20100627\_band1' and paste it on the *resampled* layers you created. To copy the style: *Right click on 'l5\_20100627\_band1' > Styles > Copy style.* See below. To paste the style: *Right click on the layer where you want to paste the style > Styles > Paste style.*



Having the original and the resampled images with the same style will make it easier to compare the result of the different resampling techniques.

### 17.1.2 Comparing the Results

You should now have four raster layers in your project. The original *Landsat band 1* and three additional images; the results from the resampling. See Fig. 17.2

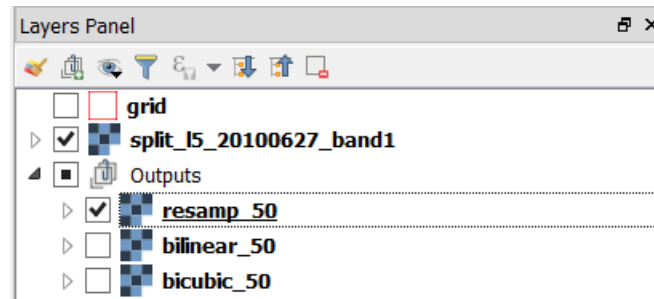


Fig. 17.2: Layers resulting from the resampling of 'I5\_20100627\_band1' using different techniques

---

**Note: Reflection.** When resampling is applied to an image to produce a version with *lower resolution*, no new radiometric information is produced. That is resampling changes the pixel size, but it does not produce further information. On the contrary, resampling often implies a loss of information, and in the case of *lower resolution* a loss of spatial precision. Despite that, resampling is a technique that is required in many cases to integrated datasets with different spatial resolution. What should be carefully considered is if the loss of information and precision are acceptable for the analysis or not.

---

**Task 6** Perform a visual comparison of the size and values of the pixels of the four raster layers. *Zoom* into to 'grid' layer and explore the raster layers using the **Value tool** and **Swipe map tool** plugins. Watch the video tutorial on [visually comparing rasters](#).

---

**Note: Reflection.** Relate the differences you observe in *resampled layers* with the theory you learned in resampling and [Geocoding](#).

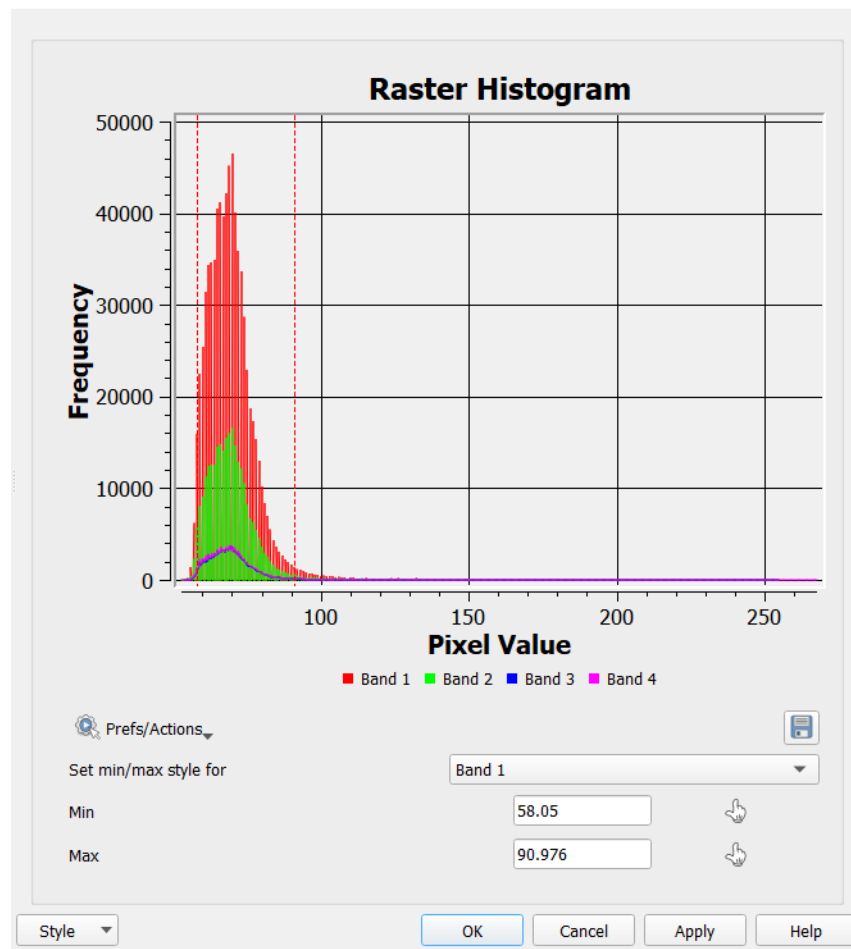
---

Another way to compare the resampling results is by using a histogram. A histogram will show the differences in the distribution of the values. To do this, we need to stack the individual resampling results in a single layer stack.

**Task 7** From the *Processing Toolbox*, open the **Build Virtual Raster** tool. For *Input layers* select: 'I5\_20100627\_band1', 'nearest\_50', 'bilinear\_50' and 'bicubic\_50'. Name the resulting stack as **stack\_50**. See Fig. 17.3







**Attention: Question.** How do you explain the differences in the distribution of values in the histogram? Especially for *band 2* (nearest\_50) and *band 4* (bicubic\_50).

## 17.2 Advanced Resampling

For many practical applications, you have to resample an image to *much larger pixel sizes* than the original. In this section, you will resample the image to a pixel size of 200 m. For the sake of comparison, you will use an **optimal** and a **sub-optimal** method.

### 17.2.1 Optimal Resampling: with Aggregation

Resampling an image to a *relatively larger pixel size* means that the **radiation** values, in the original image, must be integrated from a much larger surface area than the original. That is, for this case, to integrate the radiation values from an area of 30 m × 30 m to an area of 200 m × 200 m. To do so, we first have to do an *aggregation* (i.e. a convolution filter) and do the resampling only after that.

**Task 9** From the *Processing toolbox*, open the tool **r.mfilter** and apply a *low-pass kernel* of 7x7. Such kernel will average the data over a 7 by 7 pixels area, that is  $30 \times 7 = 210$  m. Therefore, the kernel filter aggregate the radiation value for an area of 210 m by 210 m.

As *Input layer* choose '15\_20100627\_band1' > for *Filter file* use the average.txt > for *Filtered type* average\_7x7.tif > *Run*. See Fig. 17.4

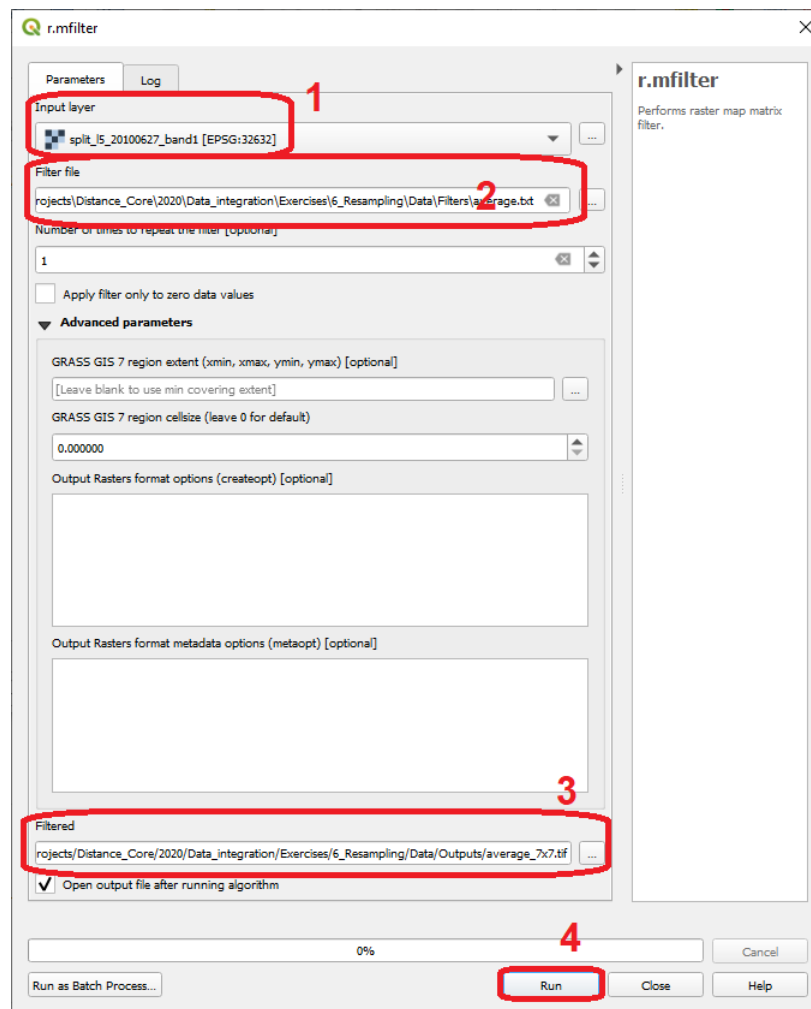


Fig. 17.4: Aggregation of radiation values using the 'r.mfilter' tool

**Task 10** Use the **r.resamp.interp** tool and resample the `average_7x7.tif` to a pixel size of 200 *m*. Use the **nearest** interpolation method. Refer to Task 3 if you need to.

### 17.2.2 Sub-Optimal Resampling: no Aggregation

To understand the reason why we should aggregate prior a resampling when the resampling resolution is *much larger* than the original pixel size. Now, you will apply only a resampling of 200 *m* to the Landsat image.

**Task 11** Use the **r.resamp.interp** just like you did in the previous tasks. Use the `'15_20100627_band1'` as input layer, **nearest** as interpolation method, and 200 *m* for pixel size.

### 17.2.3 Comparing Optima and Sub-Optimal Results

**Task 12** Compare the *resampled* layers with and without aggregation. Use the technique you used in Task 6.


**Task 13** Plot the histograms for the *resampled* layers with and without aggregation. If necessary, save the histogram(s) to a file so that you can look at both of them at the same time.

**Attention: Question.**

- When comparing the resampled images with and without aggregation. **Which one shows a ‘smoother’ image? Why?**
- Which resampled images has a smaller value range? Why?

*Section author: Zoltán Vekerdy, André Mano & Manuel Garcia Alvarez*

## MONO AND MULTI-TEMPORAL CLASSIFICATION

In section *Digital Image Classification*, we apply image classification to classify land cover, to a dataset containing images for a single date (mono-temporal). However, not all classes of land cover are equally distinguishable any time of the year. For example, seasonal crops like *maize* can only be distinguished during their growing season. Some types of trees (e.g., deciduous trees) change their spectral reflectance thru the year, because of changes in their foliage. In such cases, we can use images from multiple dates ( multi-temporal) to increase the level of information in the classification process.

In this exercise, we demonstrate the advantage of using multi-temporal data in image classification, to increase the separability between classes. We use the case of land cover mapping as an example.

---

**Important: Resources.** You will require the latest LTR version of QGIS, plus the dataset [mono\\_bi\\_temporal\\_classification.zip](#). When you unzip the dataset, you will find the following directories and files:

- **Landsat** - Images from the Landsat-TM, from 2010-06-23:

- L5197024\_02420100627\_B10.tif
- L5197024\_02420100627\_B20.tif
- L5197024\_02420100627\_B30.tif
- L5197024\_02420100627\_B40.tif
- L5197024\_02420100627\_B50.tif
- L5197024\_02420100627\_B60.tif
- L5197024\_02420100627\_B70.tif

- **ASTER** - Images from the ASTER, from 2010-04-23:

- VNIR\_Swath\_1.tif
- VNIR\_Swath\_2.tif
- VNIR\_Swath\_3B.tif
- VNIR\_Swath\_3N.tif
- SWIR\_Swath\_4.tif
- SWIR\_Swath\_5.tif
- SWIR\_Swath\_6.tif
- SWIR\_Swath\_7.tif
- SWIR\_Swath\_8.tif
- SWIR\_Swath\_9.tif

- TIR\_Swath\_10.tif
- TIR\_Swath\_11.tif
- TIR\_Swath\_12.tif
- TIR\_Swath\_13.tif
- TIR\_Swath\_14.tif

- **Shapefiles:**

- Evaluation.shp - areas with three distinct land cover classes.
- AreaOfInterest.shp - spatial extent for band subsets.

- **Output:** - An empty directory to store your results.

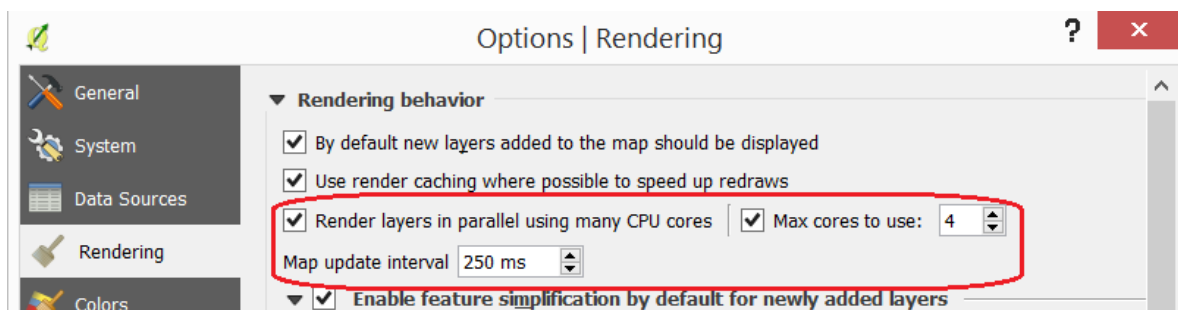
- mono\_bi-temporal\_classification.qgs - a QGIS project loaded with the files listed above.
- 

## 18.1 Setting Up the Analysis Environment

Before starting with the processing and analysis of the data in this exercise, we need to add specific tools for remote sensing to QGIS.

**Task 1** Make sure that the **Semi-automatic Classification (SCP)** and **Value tool** plugins are installed.

**Task 2** Configure QGIS to render layers using multiple CPU cores. Go to *Settings > Options > Rendering* and make sure the option *Render Layers in parallel using many CPU cores* is on. Set *Max Cores* to the number of CPU cores in your computer, use at least 4 for better performance. See below.



In the following sections, we will conduct data analysis to compare the differences between mono-temporal and bi-temporal digital image classifications. The workflow in Fig. 18.1 shows the steps (operations) for such data analysis.

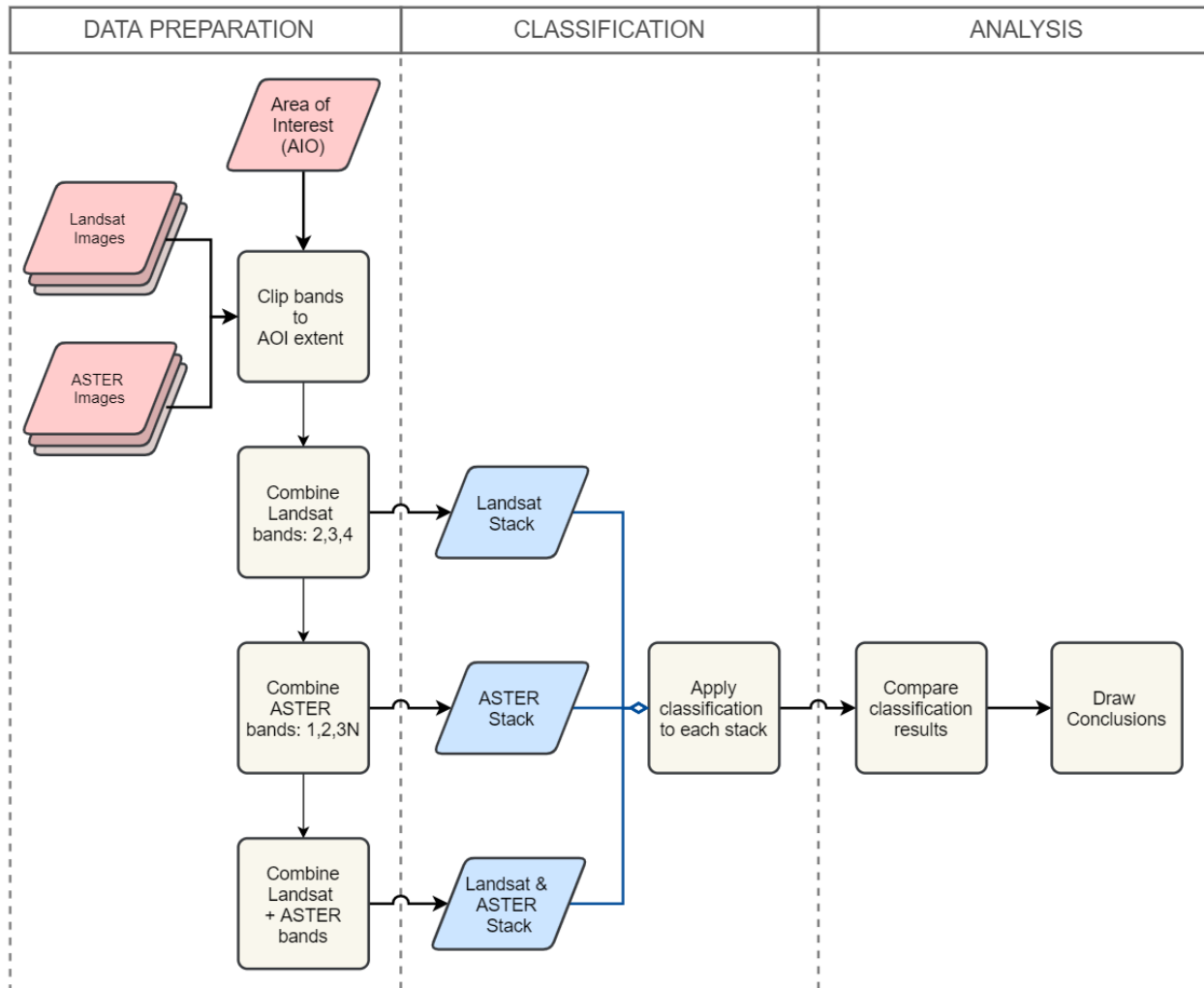


Fig. 18.1: A flowchart for the comparison of ‘mono’ and ‘bi’ temporal image classification

## 18.2 Data Preparation

### 18.2.1 Band Subsetting

You will start by stacking bands in three different combinations. One stack will combine three bands from Landsat-TM from the 23-Apr-2010. Another stack will combine three bands from ASTER, taken on the 23-Jun-2010. Finally, the last stack will combine all the bands in the previous two stacks: three Landsat bands and three ASTER bands.

**Task 3** Start by learning about the characteristics of the Landsat TM and the ASTER sensors. Put special attention to the wavelengths of the different bands. Note that *the SWIR bands are not available in the ASTER images since 2008 due to the deterioration of the cooling system of the sensor.*

Use the **ITC Sensor Repository** for this task:

- <https://www.itc.nl/Pub/sensordb/getsen.aspx?name=ASTER>

- <https://www.itc.nl/Pub/sensordb/getsen.aspx?name=TM>

**Attention: Question.** Which three bands of Landsat TM and ASTER are very similar?

Now that you know which bands are identical. We will create a subset for the area in the `AreaOfInterest.shp`. That is, we will clip each band to the extent of the polygon in such file.

**Task 4** Open the QGIS project `mono_bi-temporal_classification.qgs`. Go to *SCP > Band set*. Many of the operation available in **SCP** required to define a *band set* in advance. Then, the operation will be applied to all the bands in a set. **SCP** uses numbers to differentiate between band set. The current project already contains a band set (**1**) listing the all the bands we will need in this exercise. See Fig. 18.2

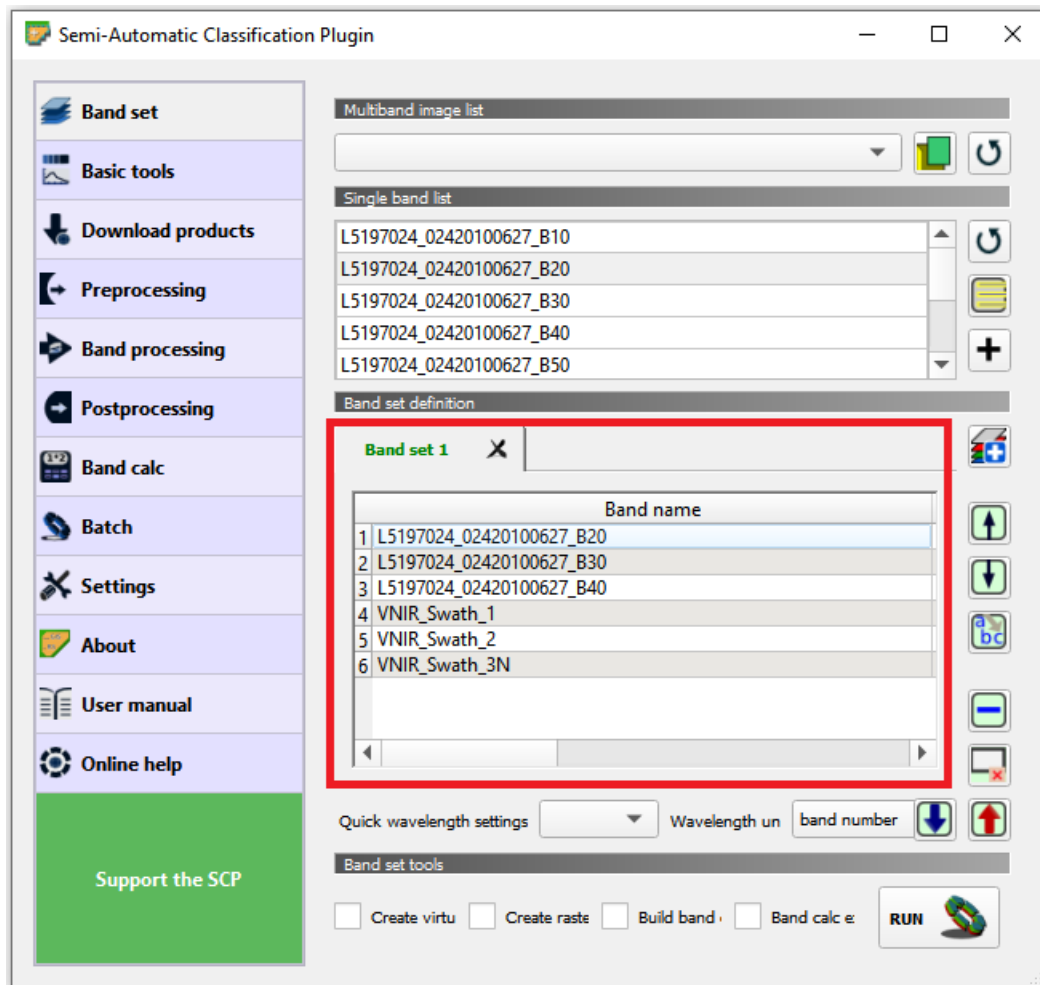



Fig. 18.2: Band set 1 in the 'mono\_bi-temporal\_classification' project

Go to *Preprocessing > Clip multiple rasters*. For *Select input band set* select **1**. Then, make sure the *Output name prefix* is set to `clip`. Tick *Use shapefile for clipping* and select the 'AreaOfInterest' layer.

See Fig. 18.3 Finally, click on .

The tool will ask you to provide a directory to store the output files. For convenience, use the empty *Output* folder inside root directory for this exercise.



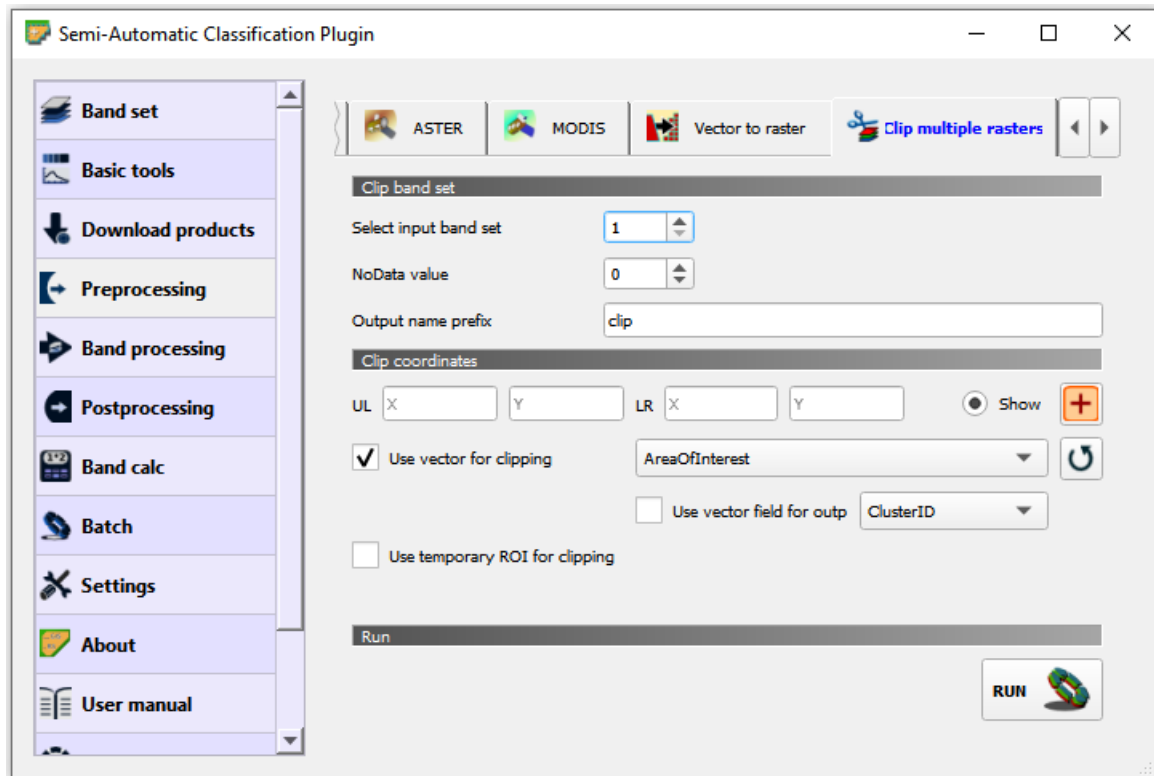
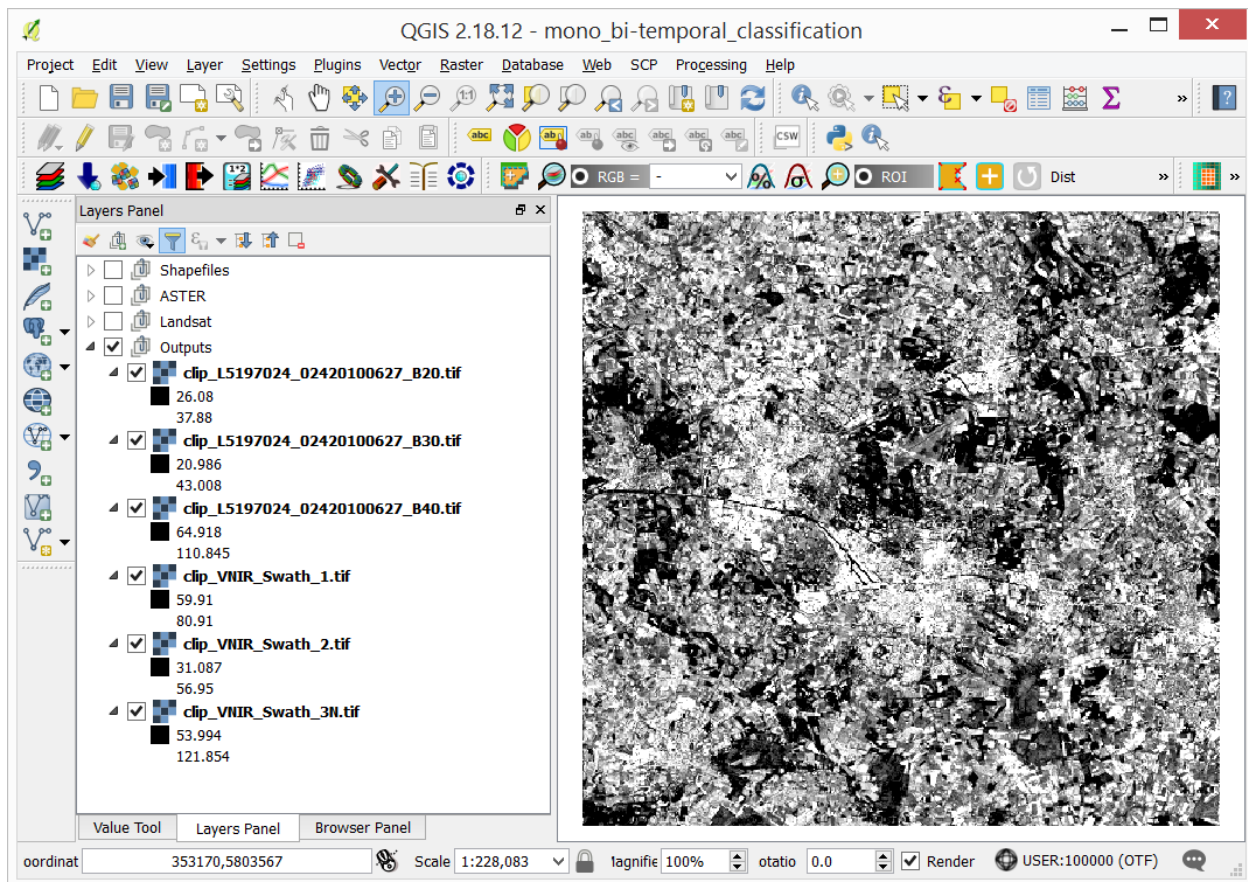


Fig. 18.3: The 'Clip multiple rasters' tool in the SCP plugin

**Note: Reflection.** Keep your QGIS project organised. The '*mono\_bi-temporal\_classification*' project has a *layer group* named "Outputs". Place the outputs of this exercise under this group. Alternatively, you can create more groups to keep the project even more organised.



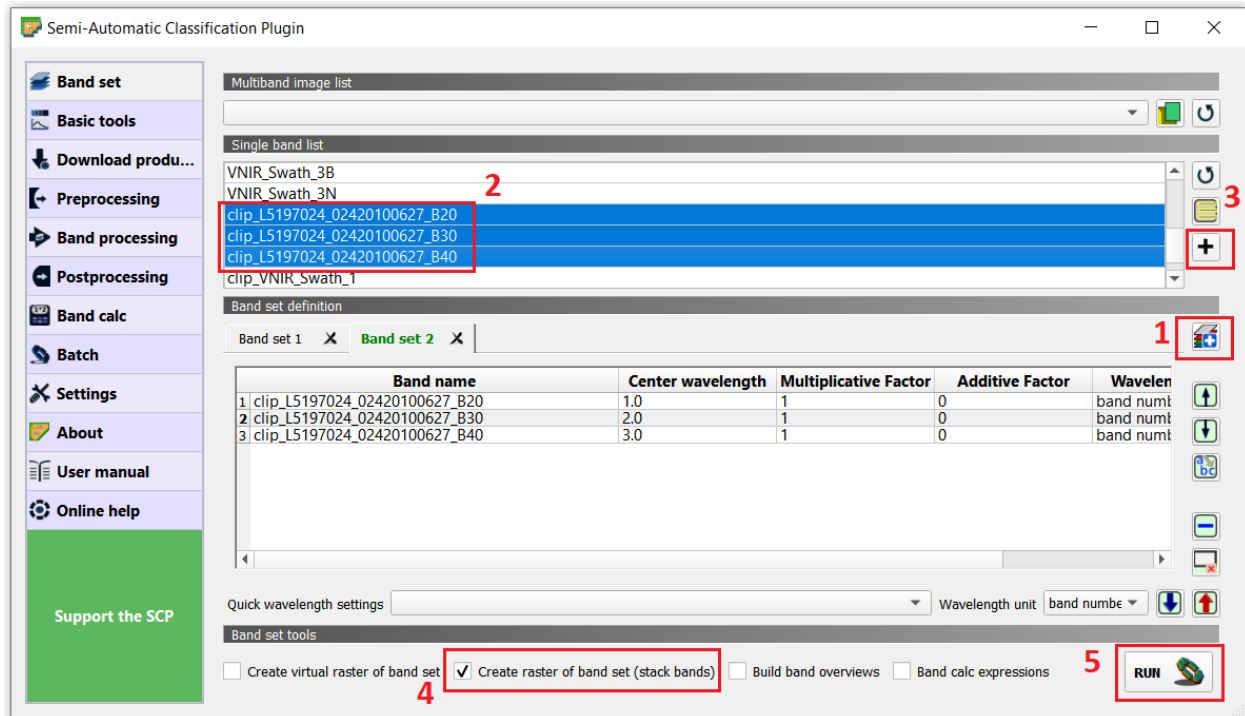
Moreover, avoid having all the layers turned on. Especially the original Landsat and ASTER layers; they will consume resources every time you zoom or pan over the map view.

## 18.2.2 Band Stacking

As a next step, you will create three band stacks using the subsets created above:

- The first stack with Landsat bands 2, 3 and 4
- The second stack with ASTER bands 1, 2 and 3N
- The third stack, a multi-spectral and multi-temporal stack with Landsat bands 2, 3, 4 and ASTER bands 1, 2, 3N

**Task 5** Create a new band stack for Landsat. Go to *SCP > Band set*. Then *Add a new band set > Select bands 2, 3, 4 from Landsat (clip version) > Add band to Band set > check that bands are in the correct order > tick Create raster or band set > Run*. See Fig. 18.4. Save the stack to the **Output** directory.

Fig. 18.4: Creating a new *band stack* in the SCP plugin

**Note:** QGIS. The QGIS does not preserve the original numbering of the bands in the new stack. This means you have to keep track of which bands in the *new stack* correspond to the *original* dataset.

For the stack you just created, that means:

Original band number (Landsat)	Band number (New Stack)
2	1
3	2
4	3

You can verify the *new stack* combined the correct bands and in the right order, using the **Value tool** plugin.

**Task 6** Open the **Value Tool** panel. Go to *View > Panels >* tick the *Value Tool*. Enable the panel; make sure only active layers are the three Landsat subsets and the *new stack*. Hover your mouse over the image and check that the pixel values correspond between the subsets and the new stack. See Fig. 18.5

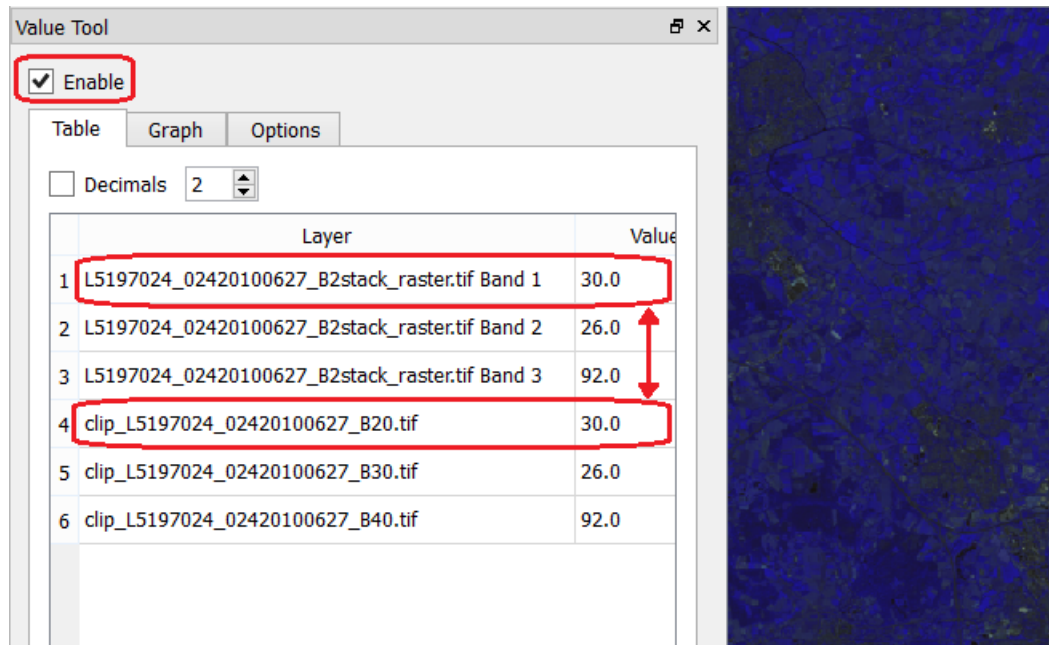


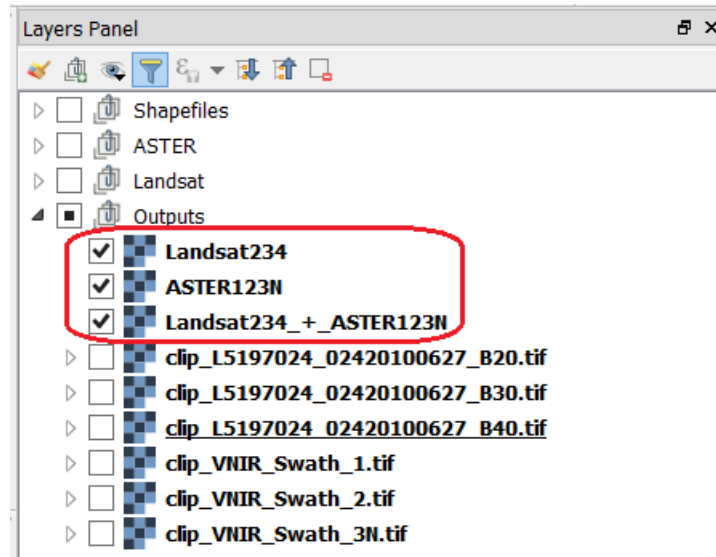
Fig. 18.5: Comparing pixel values between the 'Lansat subsets' and the 'Landsat 2,3,4 stack'.

**Task 7** Repeat the procedure in the previous Task, and create two more stacks. One for ASTER bands 1, 2, 3N. And one for the 'multi-spectral and multi-temporal', bands Landsat 2, 3, 4 and ASTER 1, 2, 3N. Remember to keep track of the order of the bands in the stacks. We suggest the following arrangement:


For ASTER band stack	
Original band number (ASTER)	Band number (ASTER stack)
1	1
2	2
3N	3

For Multi-spectral and Multi-temporal stack	
Original band number (ASTER)	Band number in Stack
Landsat 2	1
Landsat 3	2
Landsat 4	3
ASTER 1	4
ASTER 2	5
ASTER 3N	6

Now, you should have three band stacks in your project. We recommend you rename the stacks in the **Layer panel** so that you can easily distinguish them. See below.



## 18.3 Classification

Next, you will perform  **unsupervised classification** using the band stacks you created. We will use a *k-means* clustering algorithm to identify different classes of land cover. *For the sake of simplicity, we will name each class of land cover using numbers.*

You will apply unsupervised classification to each band stack: *Landsat stack*, *ASTER stack*, and *m\*ulti-spectral and multi-temporal stack\**.

**Task 8** Unclutter your project by removing the layers you will not need anymore. Remove the groups '*Landsat*' and '*ASTER*'; keep only the shapefiles and the band stacks.

**Task 9** From the **Processing** toolbox, open the tool **K-means clustering for grids**, and provide the following parameters (Fig. 18.6):

- For *Grids* select the '*Landsat234*' stack;
- For *Method* keep *Hill-Climbing*, set *Clusters* to 15, and *Maximun Iterations* to 50.
- Save the classification results to a file named '*Landsat234\_class*', in the '*Output*' folder;
- Click *Run* to execute the classification.

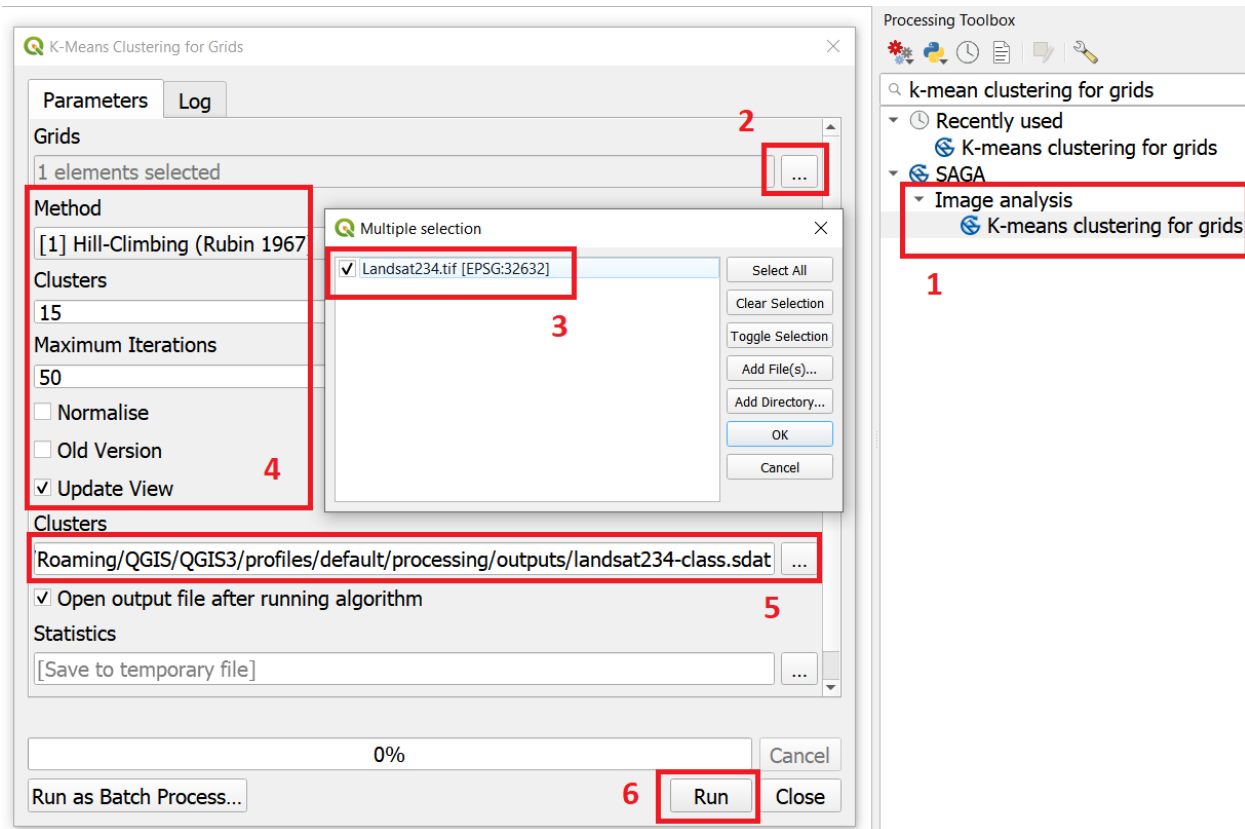


Fig. 18.6: Unsupervised classification using the 'K-Means clustering' tool

**Task 10** Repeat the *Unsupervised classification* for the 'ASTER123N' and the 'Landsat234\_+ \_ASTER123N' stack. Name the resulting files as *ASTER123N\_class* and *Landsat234\_ASTER123N\_class*, respectively.

You should now have three distinct Classifications.

## 18.4 Analysis

At this point, you have results of the *unsupervised classification* for three band stacks. Namely,

1. **Landsat234\_class** - a mono-temporal and multi-spectral classification map, using Landsat bands from April 2010.
2. **ASTER123N\_class** - a mono-temporal and multi-spectral classification map, using ASTER bands from June 2010.
3. **Landsat234\_ASTER123N\_class** - a multi-temporal and multi-spectral classification map, using bands from Landsat and ASTER for April and June 2010.

In this part of the exercise, we answer the question: **Which of the 'classification maps' distinguishes better between classes of land cover?**

### 18.4.1 Calculating Basic Statistics

**Zonal statistics** will help us to know what is the predominant pixel value (i.e. class) within for a specific area. We will use the polygons in the 'evaluation' layer to compute *zonal statistics* and find out the predominant class for each polygon.

**Task 11** From the **Processing toolbox**, open the **Zonal Statistics** tool. Compute *zonal statistics* for each of the classification maps. For *Statistics to calculate* select only majority. Type a meaningful and *distinct* prefix in *Output column prefix* each time. See Fig. 18.7 The tool will create columns with names that start with such prefixes.

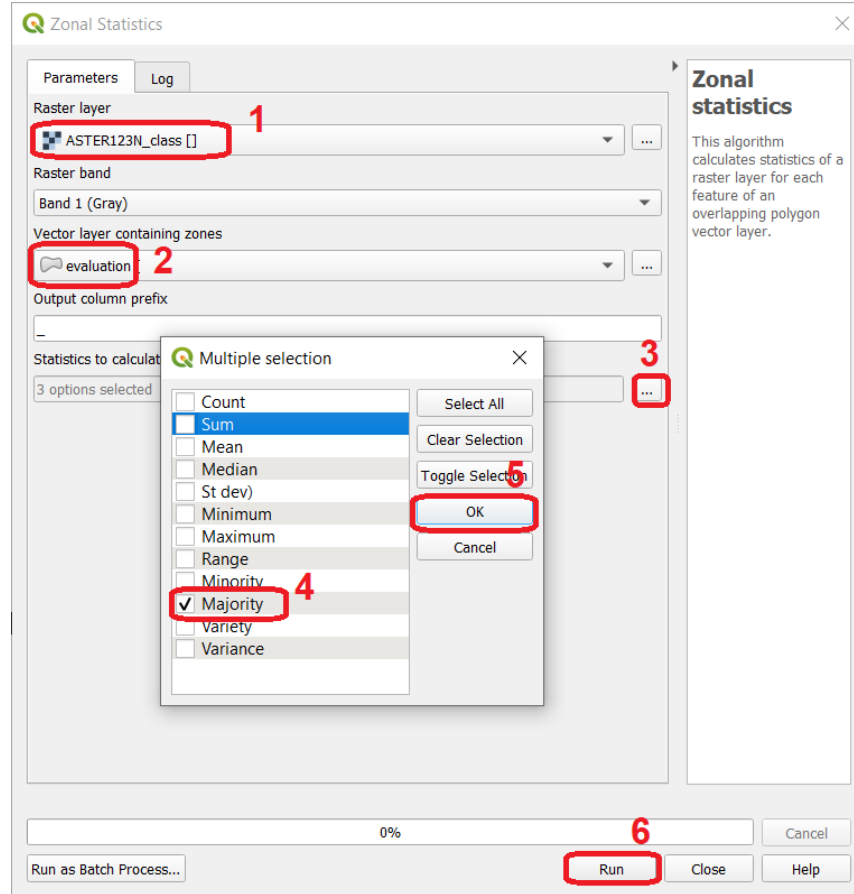


Fig. 18.7: Computing zonal statistics for the classification maps

The statistics (majority) should now be in the attribute table of the 'evaluation' layer:

evaluation :: Features total: 3, filtered: 3, selected: 0

	AREA	PERIMETER	ID	ASTERmajor	Lsatmajori	LsatAsmajo
1	58218.438	1045.015	1	7.0000000000000000	8.0000000000000000	7.0000000000000000
2	-50447.592	961.373	2	4.0000000000000000	8.0000000000000000	13.0000000000000000
3	-36090.712	763.432	3	4.0000000000000000	4.0000000000000000	4.0000000000000000

Show All Features

## 18.4.2 Comparison of Classification Results

**Task 12** Open the attribute of the ‘*evaluation*’ layer and examine the columns containing the result of *zonal statistics*. Remember, that such columns show the predominant class for each evaluation polygon (e.i., the class with more pixels inside each polygon) and each polygon in the ‘*evaluation*’ layer represent a distinct class of land cover.

Write in the table below, the total number of **distinct classes** that were identified in each *classification map*.


Classification Map	Number distinct Classes
Landsat234_class	
ASTER123N_class	
Landsat234_ASTER123N_class	

**Attention: Question.** Which ‘classification map’ provided a better separation between classes? Why?

*Section author: Zoltán Vekerdy, André Mano & Manuel Garcia Alvarez*



## CHANGE DETECTION

In this section, you will apply Remote Sensing to integrate data to detect changes in geographic phenomena. Read about it on the learning path  [Change Detection](#).

### 19.1 The Sistan Basin

In this exercise, you will work with spatial data from the Sistan basin; a wetland area on the borders of Iran and Afghanistan. The watersheds this basin consist of a system of rivers that flow from the Hindu Kush Mountains in Afghanistan through freshwater lakes, and then to a saline depression in Afghanistan; the final destination of the rivers. See [Fig. 19.1](#)



Fig. 19.1: The Sistan basin in Afghanistan and Iran [WC2010]

The lowest parts of the basin are covered by wetlands, known as *Hamoun wetlands*. The wetlands cover about  $4500 \text{ Km}^2$ , they are located in one of the driest regions in the world, such an area suffers from prolonged droughts. The droughts endanger the livelihood of about half a million inhabitants, including fishers, farmers, and others. Therefore, the monitoring of the water and vegetation is of prime importance for that ecosystem [WC2020].

---

**Note: Reflection.** Read the report [Monitoring Environmental Change in the Sistan Basin](#). It is an optional activity, but it will give you a broader context about the Sistan basin in this exercise.

---

The Sistan basin is a large and complex ecosystem. In this exercise, we will only focus on the geographic area occupied by the **Hamoun wetlands**. Specifically, we aim to answer the question **How the presence of water and vegetation changes over time in the Hamoun wetlands?**

To answer such a question, we will perform a data analysis consisting of four steps:

1. Determine where the overall presence of water, wet soil, dry soil and vegetation concentrate.
  2. Identify and compare seasonal changes in the presence of water, wet soil, dry soil and vegetation.
  3. Quantify the differences in the presence of water between specific dates.
  4. Quantify seasonal changes in the overall presence of water and vegetation.
- 

**Important: Resources.** You will require the latest LTR version of LTR version of [QGIS](#), plus the dataset *change\_detection.zip* <data\_change\_detection\_>. When you unzip the dataset, you will find several folders containing the following files:

- `vegetation.img` - time series for the presence of vegetation in the Hamoun wetlands.
- `wet_soil.img` - time series for the presence of wet soil in the Hamoun wetlands.
- `dry_soil.img` - time series for the presence of dry soil in the Hamoun wetlands.
- `water.img` - time series for the presence of water in the Hamoun wetlands.
- `limits.shp` - a boundary of the Hamoun wetlands.
- `Quantifying_change.model3` - a model to automate a data processing routine.
- `Difference_style.qml` - QGIS style file.
- `Dry_soil_sum.qml` - QGIS style file.
- `Vegetation_sum.qml` - QGIS style file.
- `Water_sum.qml` - QGIS style file.
- `Wet_soil_sum.qml` - QGIS style file.
- `date-time.csv` - a table with the acquisition dates for the time series.
- `change_detection.qgis` - a QGIS Project loaded with some of the datasets listed above.

The spatial data for this exercise consists of four image dataset with 37 bands. Each of the bands contains a *snapshot* of the *Hamoun wetlands* for different acquisition dates; between January 2, 2005, and April 22, 2006. Each image dataset describes an **index of the presence** of one of four types of land cover used, for this analysis, as variables of environmental change:

- *vegetation*,
- *dry soil*,
- *wet soil*, and
- *water*.

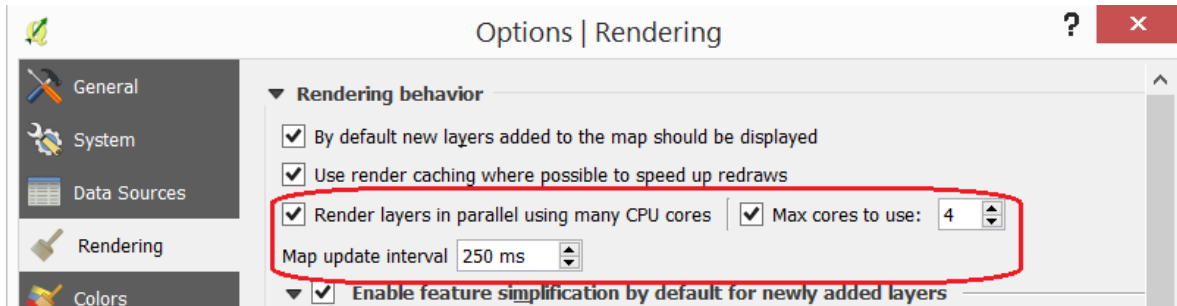
The **pixel values** in each band range from 0 to 1, and they represent the **percentage** of the area of a pixel covered by a specific type of land cover. For example, a pixel value of 0.42 in the '*vegetation*' dataset means that such a pixel is 42% covered with vegetation.

The appendix *Sistan Basin Dataset* contains a table with the acquisition dates for each band. The dates are also available in the `date-time.csv`.

---

**Task 1** Make sure you have the **Value Tool** and the **Temporal/spectral profile** plugins installed.

**Task 2** Make sure that QGIS is configured to render layers using multiple CPU cores. Go to *Settings > Options > Rendering* and make sure the option *Render Layers in parallel using many CPU cores* is on. Set *Max Cores* to the number of CPU cores in your computer, use at least 4 for better performance. See below.



## 19.2 Understanding the Data

The first step in every data analysis is to build enough understanding of the data involved. In this exercise, we will start by looking at the dates for the *change detection analysis*.

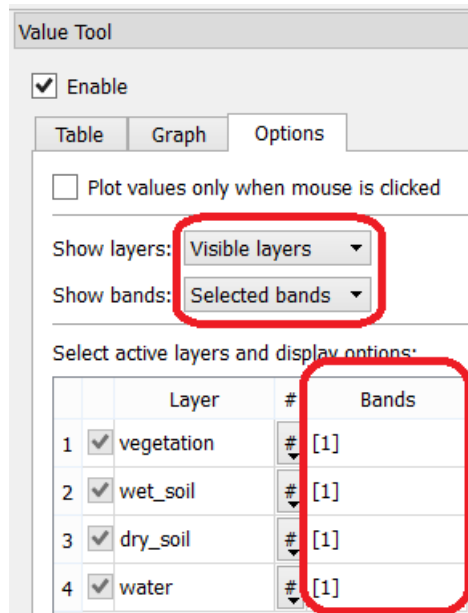
The datasets: `vegetation.img`, `wet_soil.img`, `dry_soil.img` and `water.img`, represent a time series with 37 snapshots, each snapshot is stored as a band, and each band contains values from 0 to 1 which represent indices of coverage.

**Task 3** Open the QGIS project `change_detection.qgis` and make sure you have the **Value Tool** plugin visible and active.

You will see the band 1 of each time series dataset displayed as pseudocolours. Band 1 contains values for January 2 of 2005; the starting date of the time series.

**Note: Reflection.** For the sake of comparison, switch the layers on and off and observe the values. For example, observe how the areas with high *dry soil* values have low *wet soil* values, and vice-versa. The **Value Tool** can help in such comparisons. Please do not rush this step; you must understand your datasets before proceeding with any analysis. Put special attention to the range of values in each layer and their spatial distribution.

**Note: QGIS.** The Value Tool allows you to control for which bands to plot the values. Make sure you are plotting only the values for the band 1 in each of the images; otherwise you will be plotting values for 148 bands ( $4 \times 37 = 148$ ).



By now, you should have an idea of where the values for a particular variable are higher or lower on *02/Jan/2005*. In the next step, we will build an understanding of where the values are high or low during the total length of the time series. This is, **we will identify where such values reach global maximums and maximums**.

## 19.3 Overall Concentration of Environmental Variables

To know where the presence of water, vegetation, dry and wet soil tend to concentrate during the time series; we will aggregate the values of all 37 bands.

**Task 4** Go to *Raster > Raster Calculator...* and **add** the 37 bands of each *time series*. Construct an *Expression* for the **Raster Calculator** using the formula below. Give meaningful names for each output file, for example *vegetation\_sum*, *water\_sum*, etc. See Fig. 19.2

```
"vegetation@1" + "vegetation@2" + "vegetation@3" + ... + "vegetation@36" +  
→ "vegetation@37"
```

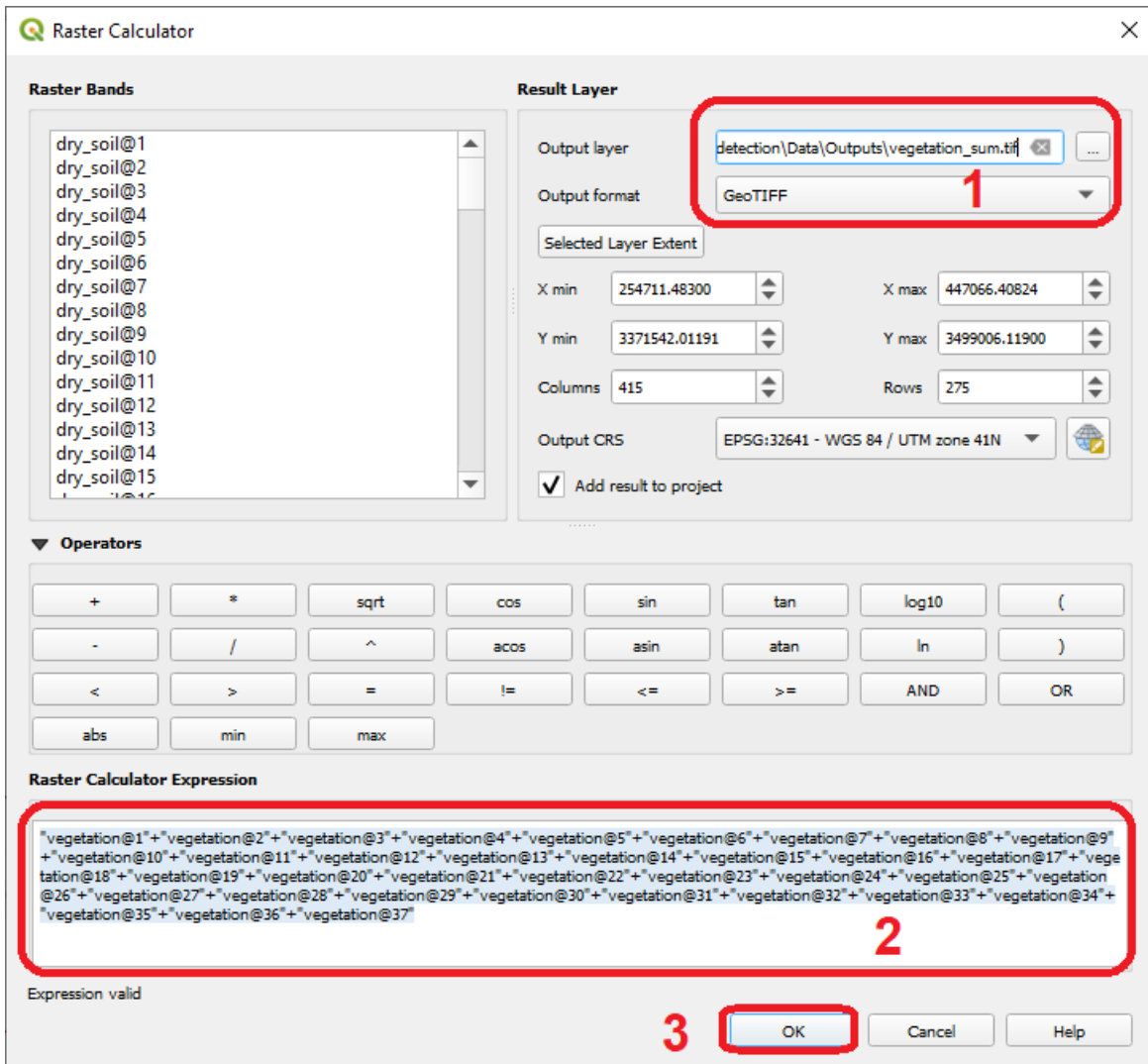


Fig. 19.2: Aggregation of values using the 'Raster Calculator'

**Note: QGIS.** For convenience, you can copy the expressions listed below to the *Raster Calculator Expression*.

- 'vegetation' time series:

```
"vegetation@1"+"vegetation@2"+"vegetation@3"+"vegetation@4"+"vegetation@5"+
"vegetation@6"+"vegetation@7"+"vegetation@8"+"vegetation@9"+"vegetation@10"+
"vegetation@11"+"vegetation@12"+"vegetation@13"+"vegetation@14"+"vegetation@15"+
"vegetation@16"+"vegetation@17"+"vegetation@18"+"vegetation@19"+"vegetation@20"+
"vegetation@21"+"vegetation@22"+"vegetation@23"+"vegetation@24"+"vegetation@25"+
"vegetation@26"+"vegetation@27"+"vegetation@28"+"vegetation@29"+"vegetation@30"+
"vegetation@31"+"vegetation@32"+"vegetation@33"+"vegetation@34"+"vegetation@35"+
"vegetation@36"+"vegetation@37"
```

- 'wet\_soil' time series:

```
"wet_soil@1"+"wet_soil@2"+"wet_soil@3"+"wet_soil@4"+"wet_soil@5"+"wet_soil@6"+
"wet_soil@7"+"wet_soil@8"+"wet_soil@9"+"wet_soil@10"+"wet_soil@11"+"wet_soil@12"+
```

(continues on next page)

(continued from previous page)

```
"wet_soil@13"+"wet_soil@14"+"wet_soil@15"+"wet_soil@16"+"wet_soil@17"+"wet_soil@18"+  
"wet_soil@19"+"wet_soil@20"+"wet_soil@21"+"wet_soil@22"+"wet_soil@23"+"wet_soil@24"+  
"wet_soil@25"+"wet_soil@26"+"wet_soil@27"+"wet_soil@28"+"wet_soil@29"+"wet_soil@30"+  
"wet_soil@31"+"wet_soil@32"+"wet_soil@33"+"wet_soil@34"+"wet_soil@35"+"wet_soil@36"+  
"wet_soil@37"
```

- ‘dry\_soil’ time series:

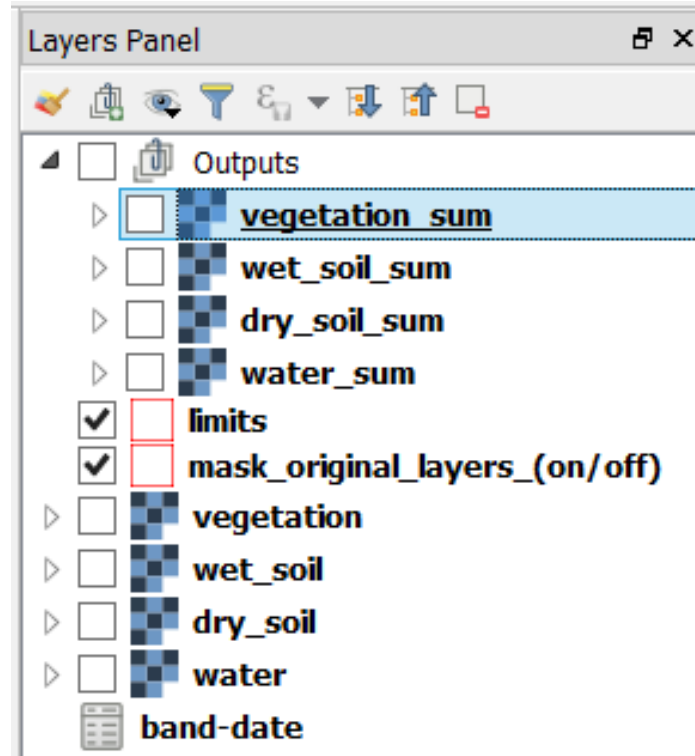
```
"dry_soil@1"+"dry_soil@2"+"dry_soil@3"+"dry_soil@4"+"dry_soil@5"+"dry_soil@6"+  
"dry_soil@7"+"dry_soil@8"+"dry_soil@9"+"dry_soil@10"+"dry_soil@11"+"dry_soil@12"+  
"dry_soil@13"+"dry_soil@14"+"dry_soil@15"+"dry_soil@16"+"dry_soil@17"+"dry_soil@18"+  
"dry_soil@19"+"dry_soil@20"+"dry_soil@21"+"dry_soil@22"+"dry_soil@23"+"dry_soil@24"+  
"dry_soil@25"+"dry_soil@26"+"dry_soil@27"+"dry_soil@28"+"dry_soil@29"+"dry_soil@30"+  
"dry_soil@31"+"dry_soil@32"+"dry_soil@33"+"dry_soil@34"+"dry_soil@35"+"dry_soil@36"+  
"dry_soil@37"
```

- ‘water’ time series:

```
"water@1"+"water@2"+"water@3"+"water@4"+"water@5"+"water@6"+"water@7"+"water@8"+  
"water@9"+"water@10"+"water@11"+"water@12"+"water@13"+"water@14"+"water@15"+"water@16"  
↪ "+"  
"water@17"+"water@18"+"water@19"+"water@20"+"water@21"+"water@22"+"water@23"+"water@24"  
↪ "+"  
"water@25"+"water@26"+"water@27"+"water@28"+"water@29"+"water@30"+"water@31"+"water@32"  
↪ "+"  
"water@33"+"water@34"+"water@35"+"water@36"+"water@37"
```

---

**Note: QGIS.** Keep your project organized. The ‘change\_detection’ project has a layer group named “Outputs”. Place the outputs you generate under this group or create more groups to keep the layer in your project organized. Also, keep the two vector layers always visible.



**Task 5** Change the **Style** for each of the layer you produced in the previous task, so that you can easily visualise where the values concentrate (i.e., where they reach their highest and lowest values). For the 'vegetation\_sum' layer, go *Right-Click > Properties... > Symbology > Style > Load Style... > search and select for the vegetation\_sum.qml file > Open > OK*. See [Fig. 19.3](#)

The styles you applied are only to facilitate a visual analysis. *All the layers are divided into 5 classes but only the highest 20% of values are visible*. Such values identify areas where the presence of each index (variable) accumulates over period depicted in the time series.

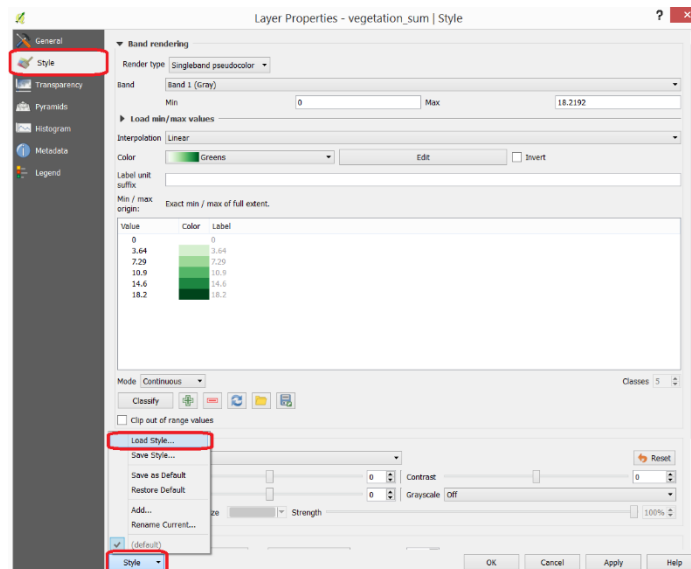


Fig. 19.3: Apply a style to the 'vegetation\_sum' layer using a style file

**Task 6** Repeat the procedure above to change the styles of ‘wet\_soil\_sum’, ‘dry\_soil\_sum’, and ‘water\_sum’ layers. Look for the correct style files in your data directory. Your project should now have the four aggregation layer properly styled. See Fig. 19.4

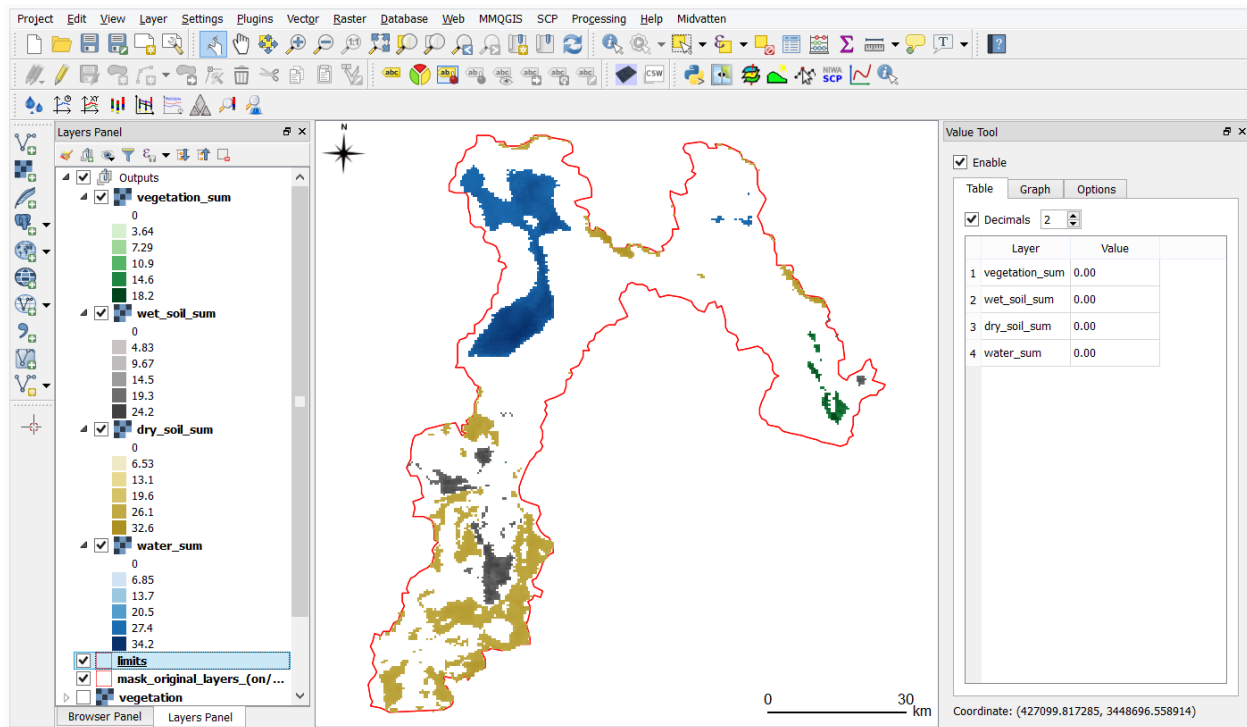


Fig. 19.4: The ‘vegetation\_sum’, ‘wet\_soil\_sum’, ‘dry\_soil\_sum’, and ‘water\_sum’ layers with custom styles

## 19.4 Identification and Comparison of Seasonal Changes

Now that you have an overview of the range and spatial distribution of value for each of the ‘index’ image. We will take a look at how the values change over time.

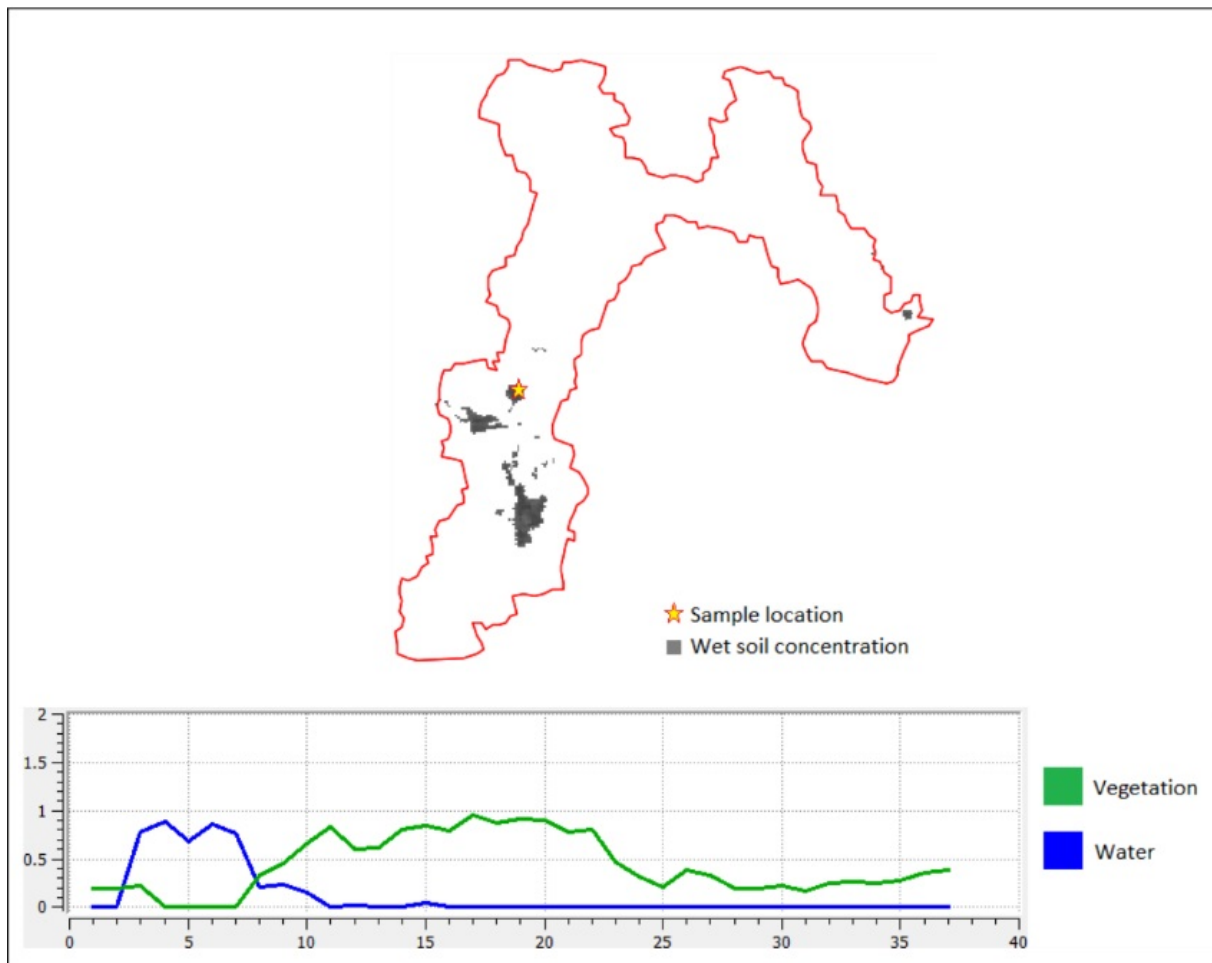
**Task 7** Use the **Temporal/spectral Profile** plugin to inspect how the values in the ‘water’ layer change over time. Sample two or three points close to the areas where the values in the ‘aggregated’ layers are the highest. Watch the video tutorial on [inspecting time series](#).

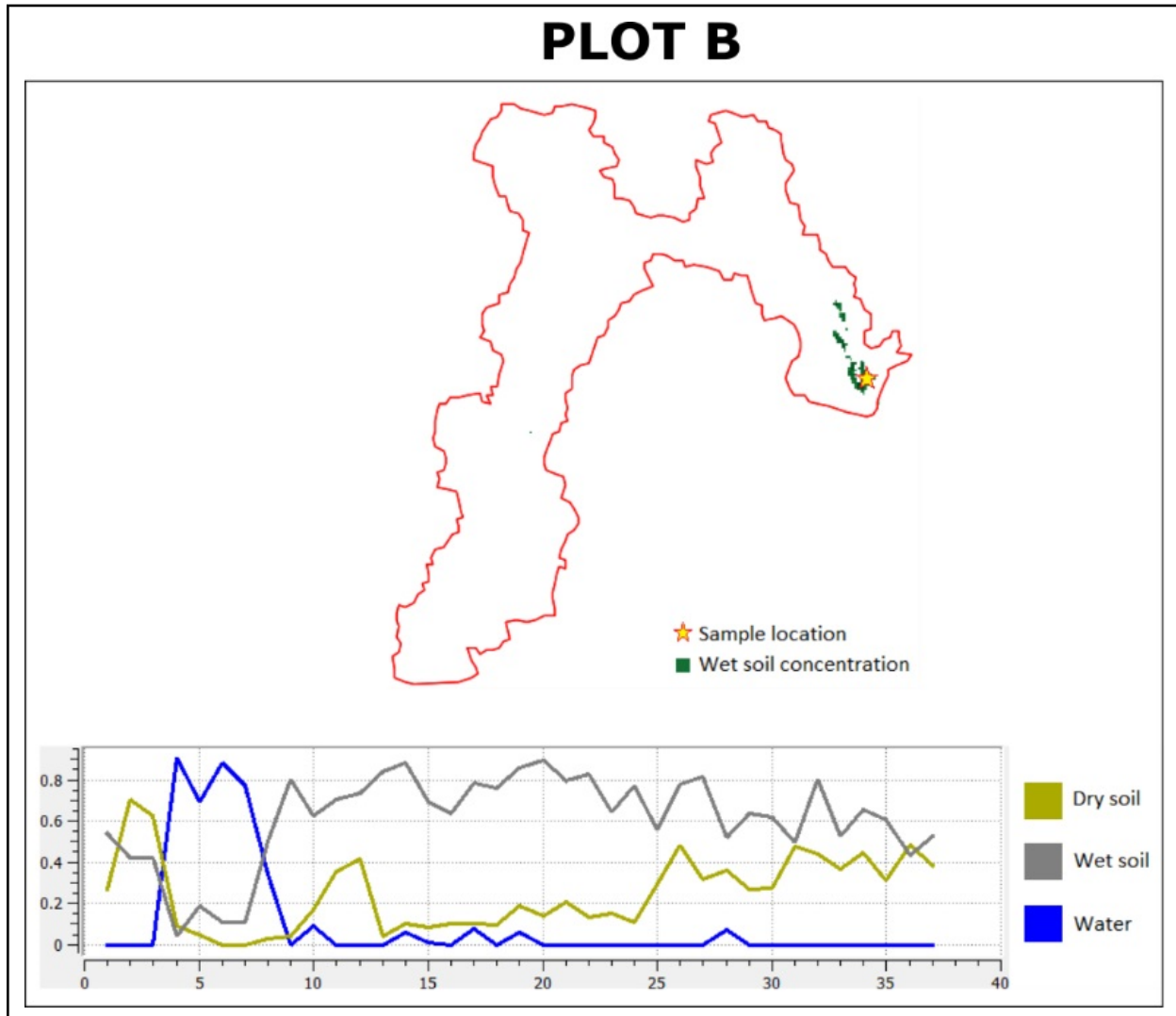
**Task 8** Use the **Temporal/spectral Profile** plugin to explore how the other *variables* change or compare over time.

**Attention: Question.** Observe the two plots below. For each plot, **can you explain how the changes in each variable are related?**



## PLOT A





## 19.5 Quantifying Differences in the Presence of Water

In this section, we will quantify the changes in the presence of water. Specifically, we will look at how the values of the *presence of water* increase or decrease between dates. This variable is crucial because its behaviour influence the other three variables.

**Task 9** From the *Processing Toolbox*, Right-click on the tool **Raster calculator** > *Edit Rendering Styles for Outputs...* See [Fig. 19.5](#) Click the ellipses (...) > select the `Difference_style.qml` file > *Open* > *OK*.

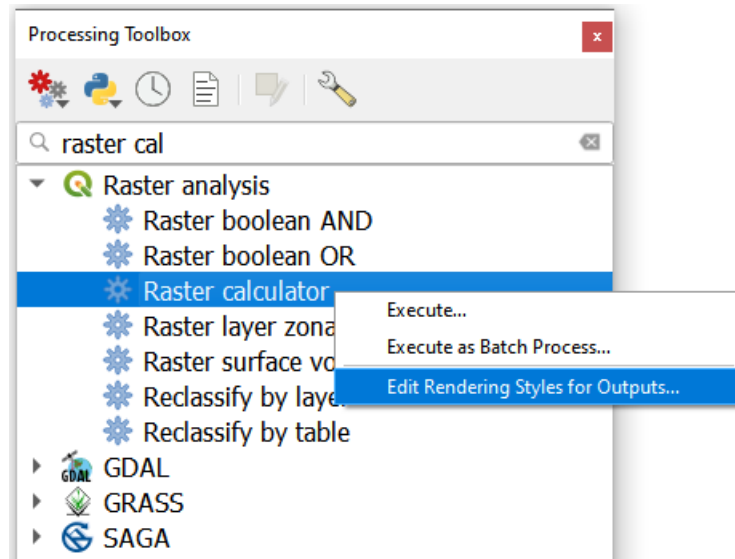


Fig. 19.5: The 'Raster calculator' in the Processing Toolbox

**Note: QGIS.** In the following tasks, you will use the **Raster calculator** to generate layers that compute the difference between two acquisition dates. Setting the tools to use the same style to render the output layers will make it easier to compare and understand the results, and it will save time.

**Task 10** Use the **Raster calculator** in the *Processing Toolbox* to compute the difference between **band 8** (21/04/2005) and **band 12** (22/06/2005) from the 'water' layer. Use the following formula to compute the difference map:

$$\Delta_{map} = Map_{(final\ state)} - Map_{(initial\ state)}$$

Do not forget to set the *Reference layer(s)* parameter to the 'water' layer, Fig. 19.6

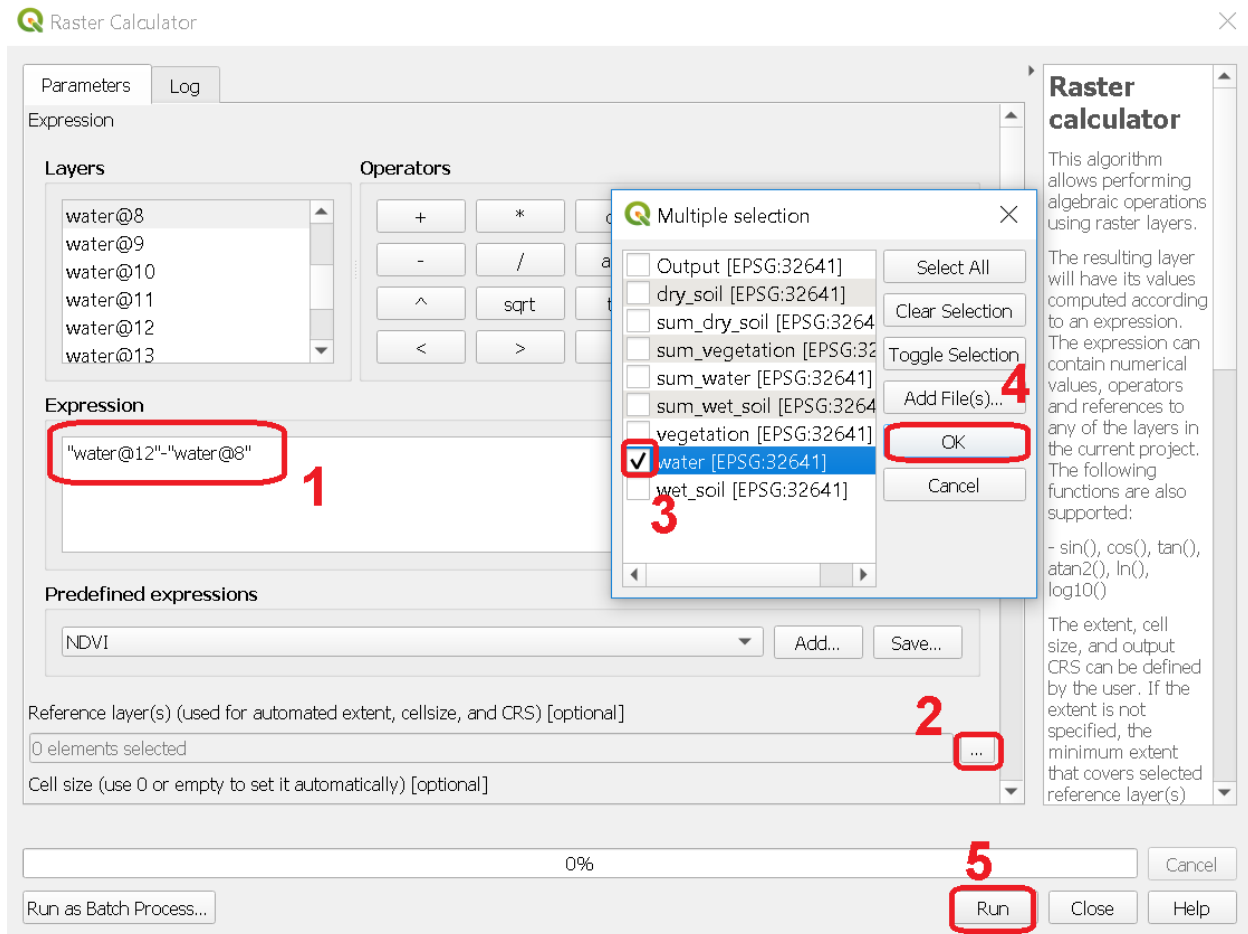
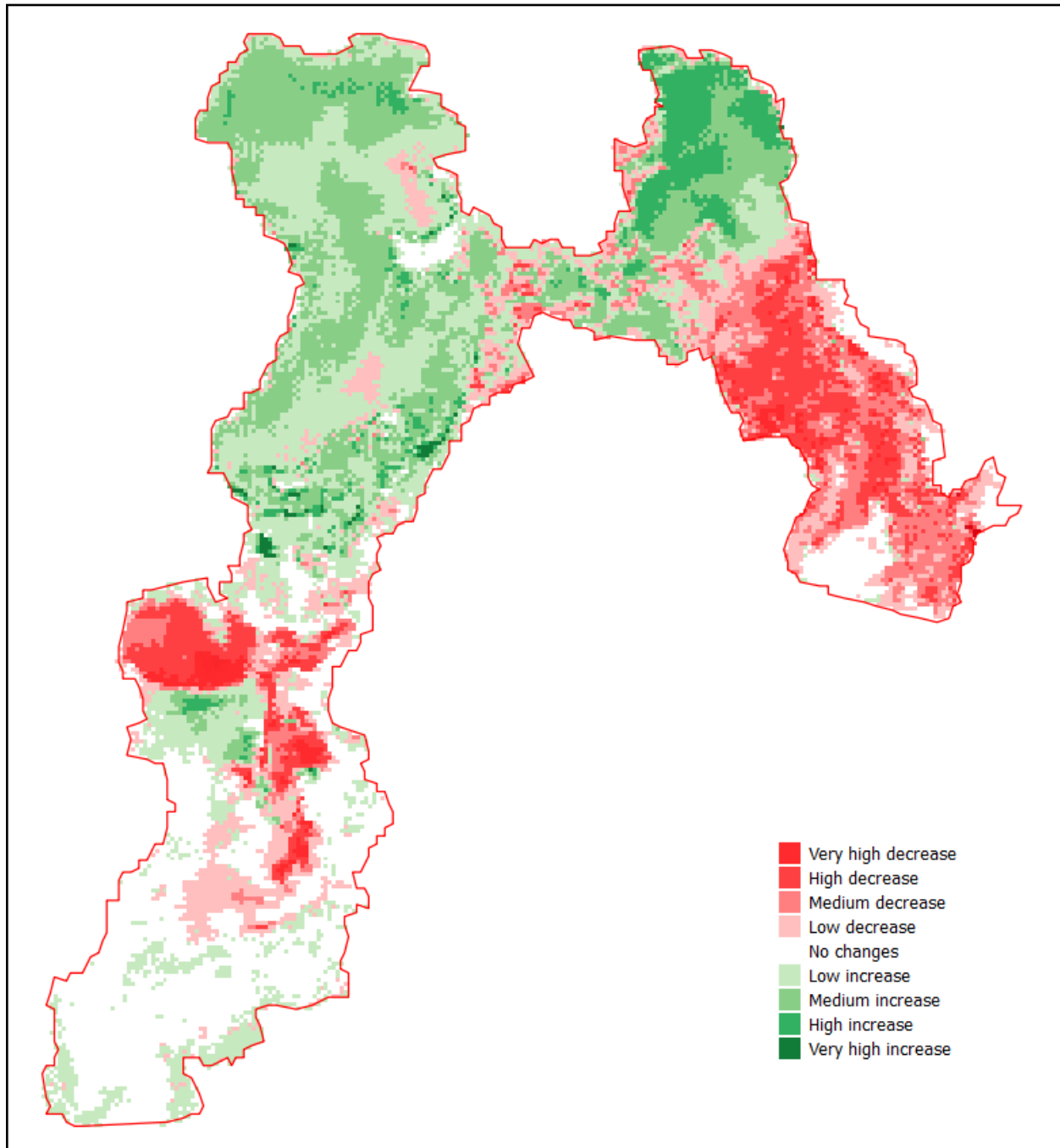


Fig. 19.6: Computing a difference map between bands of the 'water' layer

**Attention: Question.** Look closely at map resulting from the previous task. See below. **What do the values of the difference map mean?**



**Task 11** Repeat the procedure described in the previous task. This time compute the difference between **band 8** (21/Apr/2005) and **band 20** (12/Sep/2005) from the 'water' layer.

**Attention: Question.** Look closely at difference maps from the previous tasks. **What changes occurred between April 21 and September 12 of 2005?**

## 19.6 Quantifying Changes in Water and Vegetation

In the last part of this exercise, we will assess how the values change globally in the Hamoun wetlands over ten months; from January to October 2005. For this analysis, *we consider the percentage of the total area of the basin covered by water*. Earlier in the exercise, we explained that the *index* (0 to 1) represent the percentage of the area of a pixel covered by a specific geographic phenomenon, in this case, *water*. Therefore, the *total percentage of area covered by water* in a band  $T_{water}$ , can be computed using the following equation:

$$T_{water} = \frac{B \times 100}{A}$$

Where  $A$  is the total number of pixels in a band, and  $B$  is the sum of the pixel values in that same band. The constant 100 converts the values to percentage.

To complete this part of the analysis, we would have to apply the equation above 10 times. One time per each month according to the table below.

Band (water layer)	Date (Y-M-D)	Date (M-Y)
1	05-01-02	Jan-05
3	05-02-21	Feb-05
4	05-03-12	Mar-05
7	05-04-03	Apr-05
10	05-05-22	May-05
11	05-06-13	Jun-05
13	05-07-08	Jul-05
16	05-08-09	Aug-05
19	05-09-03	Sep-05
23	05-10-01	Oct-05

Instead of repeating the procedure to compute  $T_{water}$  over and over, you will use a predefined QGIS model which automate such a task.

**Task 12** Go to *Processing Toolbox > Models > Add Model to Toolbox...* See Fig. 19.7 . Select the model *quantifying\_change.model3*.

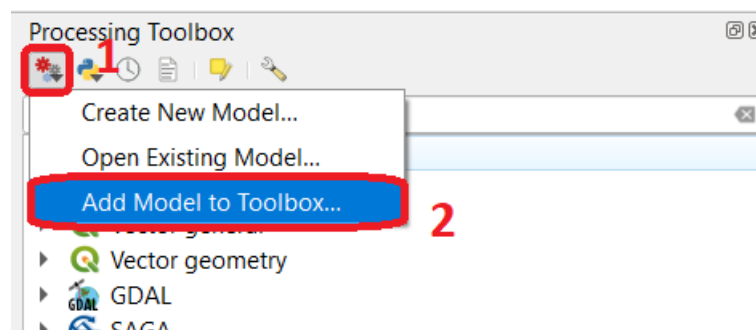


Fig. 19.7: Add model to 'Processing Toolbox'

**Task 13** Use the model you just add to the **Processing Toolbox** to compute  $T_{water}$  for the bands listed in the table above. Go to *Processing Toolbox > Models > quantifying\_changes*, Fig. 19.8 . Double click on the model to open the model. For *Extent* and *Indicator*, select the 'water' layer. Click *Run*

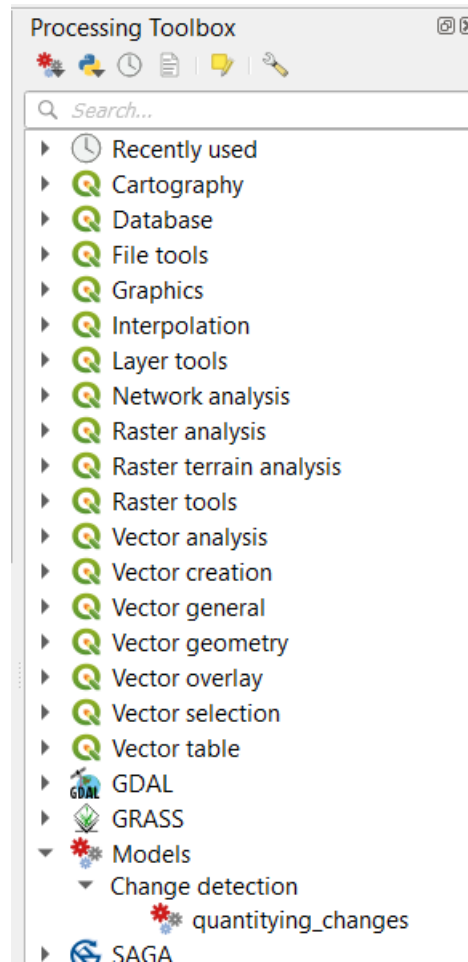


Fig. 19.8: The model 'quantifying changes' in the Processing Toolbox

**Note: QGIS.** The '*Quantifying\_change*' model outputs a band stack containing the values for  $T_{water}$ .  $T_{water}$  is the total percentage of area covered by water per band (month). Therefore, you will notice that for a particular band, *all the pixels inside the boundary of the Hamoun wetlands have the same value*.

Besides computing the value of  $T_{water}$  for a band; the '*Quantifying\_change*' model creates a single-band raster to which it copies the value of  $T_{water}$  to all pixels inside the `limits.shp`. The model repeats this for each band in the table above, and then it combines all bands in a band stack called '*OUTPUT*'. After running the model, you should see the '*OUTPUT*' stack in the QGIS project.

**Task 14** Use the **Temporal/Spectral Profile** plugin to inspect the values of the band stack created by the model, [Fig. 19.9](#) Refer to **Task 8** if you need to.

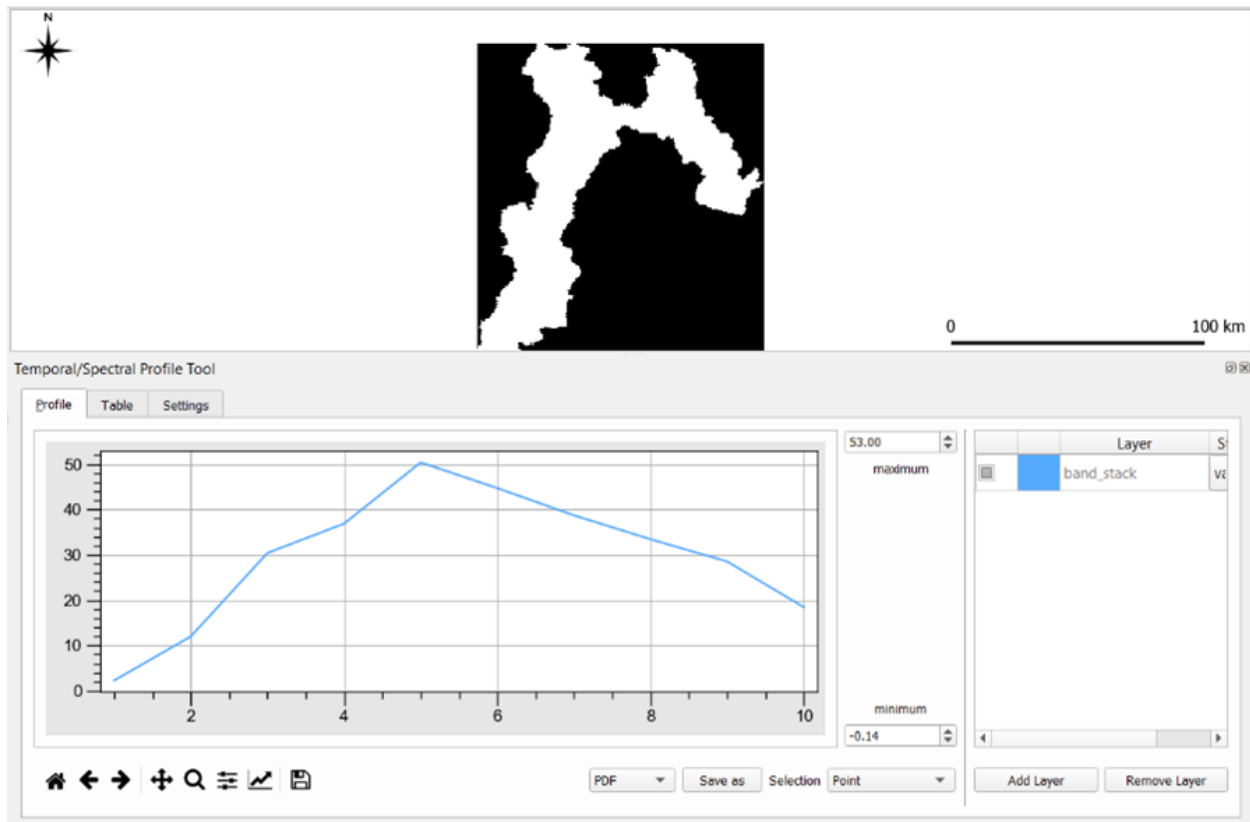


Fig. 19.9: Exploring the values of 'total percentage of area covered by water' with the 'Temporal/Spectral Profile' tool

**Note: QGIS.** QGIS will not preserve the original number of the bands in the output band stack. This means that you have to keep track of which band in the output stack corresponds to the bands in the original dataset. In this case, they correspond as following:

Original Band No.	Band No. in Output Stack	Month
1	1	January
3	2	February
4	3	March
7	4	April
10	5	May
11	6	June
13	7	July
16	8	August
19	9	September
23	10	October

**Attention: Question.** Look at the line plot of the values of  $T_{water}$  in the Temporal/Spectral Profile tool. What does the profile curve show? How do we interpret the values?

**Task 15** Repeat Tasks 13 and 14 using another variable, for example, *vegetation*. Plot the profile curves in the



**Temporal/Spectral Profile** plugin. Write down your observations and take them to the virtual classroom.

*Section author: Zoltán Vekerdy, André Mano & Manuel Garcia Alvarez*



## COMMON GIS FILE FORMATS

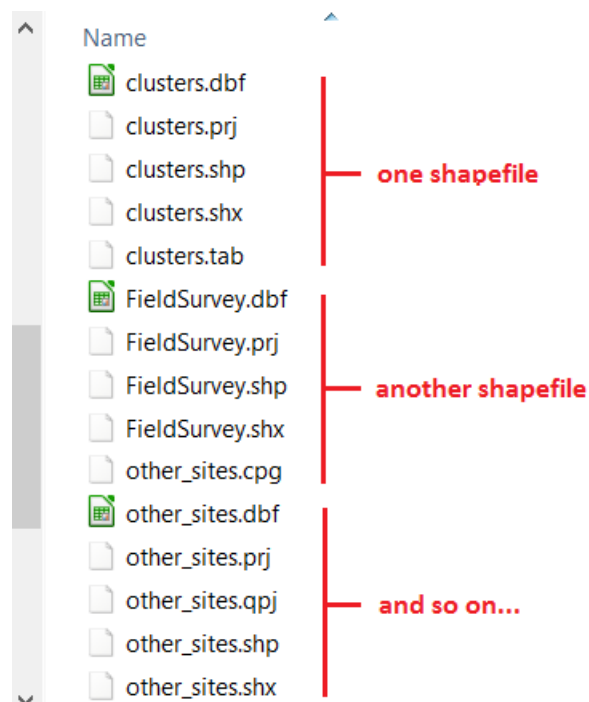
**Warning:** The information contained on this section **IS NOT part of the assessments** in this course. We do not expect you know all this. It is provided as reference for practical reasons.

### 20.1 Vector GIS File Extensions

#### 20.1.1 Shapefiles (.shp)

The most common file type amongst vector data is the Shapefile (.shp). It is maintained by ESRI and because its specification is public, shapefiles have become a de facto standard towards spatial information storage. Most of the time you will be working with shapefiles, therefore it's important to understand what exactly are they.

The shapefile is in fact a collection of files and not a single file. Some of these files are mandatory, others are optional. The first thing you have to remember is that for shapefiles to work properly all mandatory files have to be in the same directory and share the same name (only the extension changes). Remember this when copying shapefiles from one folder to another – ALL the files with the same name are components of a given shapefile and therefore have to be together for the shapefile to work properly.



When you create a new shapefile, your GIS software will automatically create all or at least the mandatory components of a shapefile (the following is based on Wikipedia entrance on the shapefile format):

### **Mandatory files:**

- `.shp`— shape format; the feature geometry itself
- `.shx`— shape index format; a positional index of the feature geometry
- `.dbf`— attribute format; columnar attributes for each shape, indBaseIV format

### **Optional files (most common):**

- `.prj`— projection format; the coordinate system and projection information.
- `.sbn` and `.sbx`— aspatial index of the features.
- `.cpa`— for identifying the character encoding to be used.

Another important characteristic of the shapefile is that it only supports one type of geometry for shapefile and the possibilities are point, line or polygon. That is to say that if you have a shapefile with polygons representing, say houses, you cannot have features represented by a line or by a point.

### **20.1.2 Geopackage (.gpkg)**

Geopackage is essentially a mini database in the form of one single file. It can contain several layers of different geometries, and also raster layers. In addition to that it offers some of the capabilities of databases like relationships, advanced constraints and triggers. It is a format based on an open standard maintained by the [Open Geospatial Consortium](#) (OGC).

Currently, this is the default vector format of QGIS and it is the vector format we will use during the course. Because of the advantages it offers in comparison with shapefiles, it is possible that geopackage will become as common as shapefiles.

### **20.1.3 Keyhole Markup Language (.kml)**

KML is a XML file that was originally developed for use within Google Earth back in 2004. In 2008, it became an Official Open Geospatial Consortium Standard (OGC), and since then it became a widely adopted format especially for use with Google Earth and web mapping applications.

KML allows to store more than one geometry per file, and even rasters if we use the compressed version of it – KMZ. Contrary to shapefiles, a KML file is composed of only one file, but on the other hand the alphanumeric information associated to the geometries (i.g. the attribute table) is not as easy to manage as in the case of shapefiles.

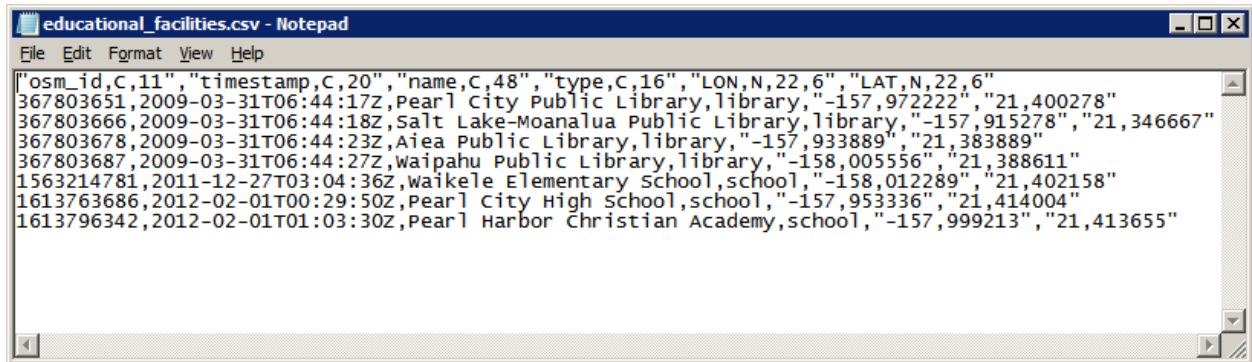
### **20.1.4 GPS eXchange Format (.gpx)**

Another common GIS data format is the GPS eXchange Format (.gpx), also an XML schema designed specifically for exchange data between GPS devices. Due to the increasing popularity and availability of GPS devices, GPX format became a very common format. Although not an official standard, due to its wide adoption to carry GPS log data, this format became a de facto standard. The latest specification is 1.1 and it was released in 2004.

GPX has two particularities. The first is that it only stores points and lines and with these geometries you can have waypoints, routes, route points, tracks and track points. The second is that the attribute table associated to the features has a fixed structure, determined by the XML schema that describes the GPX format.

## 20.1.5 Comma Separated Values (.csv)

Comma Separated values (.csv) is not a format specific to GIS but rather a general purpose format that allows to carry data in tabular format using commas or other pre-defined separator like semi-colon or Tab to delimitate columns. If you open the file `educational_facilities.csv` in a text editor (do not use Word or similar programs) this is what you will see:



The first line is called the 'header' of the CSV and it is the equivalent to column names in a regular table. The following lines represent different rows just like you would see in a regular excel-like table.

## 20.2 Raster GIS file extensions

### 20.2.1 Tagged Image File Format (.tif)

Although not the most efficient, Tagged Image File Format (.tif), is one of the most widespread raster formats. The TIFF specification is public and therefore largely supported by every software package. The latest version was released in 1992 and maintained by Adobe. What makes TIFF so popular to distribute GIS data is its flexibility. It is not in the scope of this exercise to cover technical details, but here is quick overview of TIFF capabilities:

- Supports 8 to 16 bit (1 or 2 bytes) colour range per band, meaning a range of values of 0 to 255 or 0 to 32768 respectively;
- In a multi-band raster, the above capability allows TIFF to store a range of values that stretches to the millions.
- Allows negative and decimal values.
- Supports loss and lossless compression.
- Can store tags with alphanumeric information.
- The tags can be used to store georeferencing information. When that happens the TIFF is called GeoTIFF.

### 20.2.2 Other Raster and Vector formats

The list of GIS formats is very extensive. The formats we covered so far are the ones that you are more likely to encounter. In the case of QGIS, the complete list of supported formats depends on the formats supported by the GDAL library. You can check the complete list of supported formats in the links below:

- Raster formats - [http://www.gdal.org/formats\\_list.html](http://www.gdal.org/formats_list.html)
- Vector formats - [http://www.gdal.org/ogr\\_formats.html](http://www.gdal.org/ogr_formats.html)

In addition there are closed proprietary formats that are also fairly common, however because they are closed they do not ensure data interoperability. If you come across these formats you might need a special license to read them or convert them to one of the GDAL supported formats. The most important formats among these include:

- DWG (drawing) – a proprietary vector format native of AutoCAD
- .MrSID (Multiresolution Seamless Image Database) – a proprietary raster format owned by LizTech, possibly the most efficient format to store raster data;
- ECW (Enhanced Compression Wavelet) – another proprietary format for rasters developed by Hexagon Geospatial.

**Caution: Do not overload yourself in knowing all these formats and their details.** The main point here is to understand that different file formats provide different possibilities. Whenever you come across a data format you are not familiar with, do a little research on it to know if that format is adequate for your analysis needs.

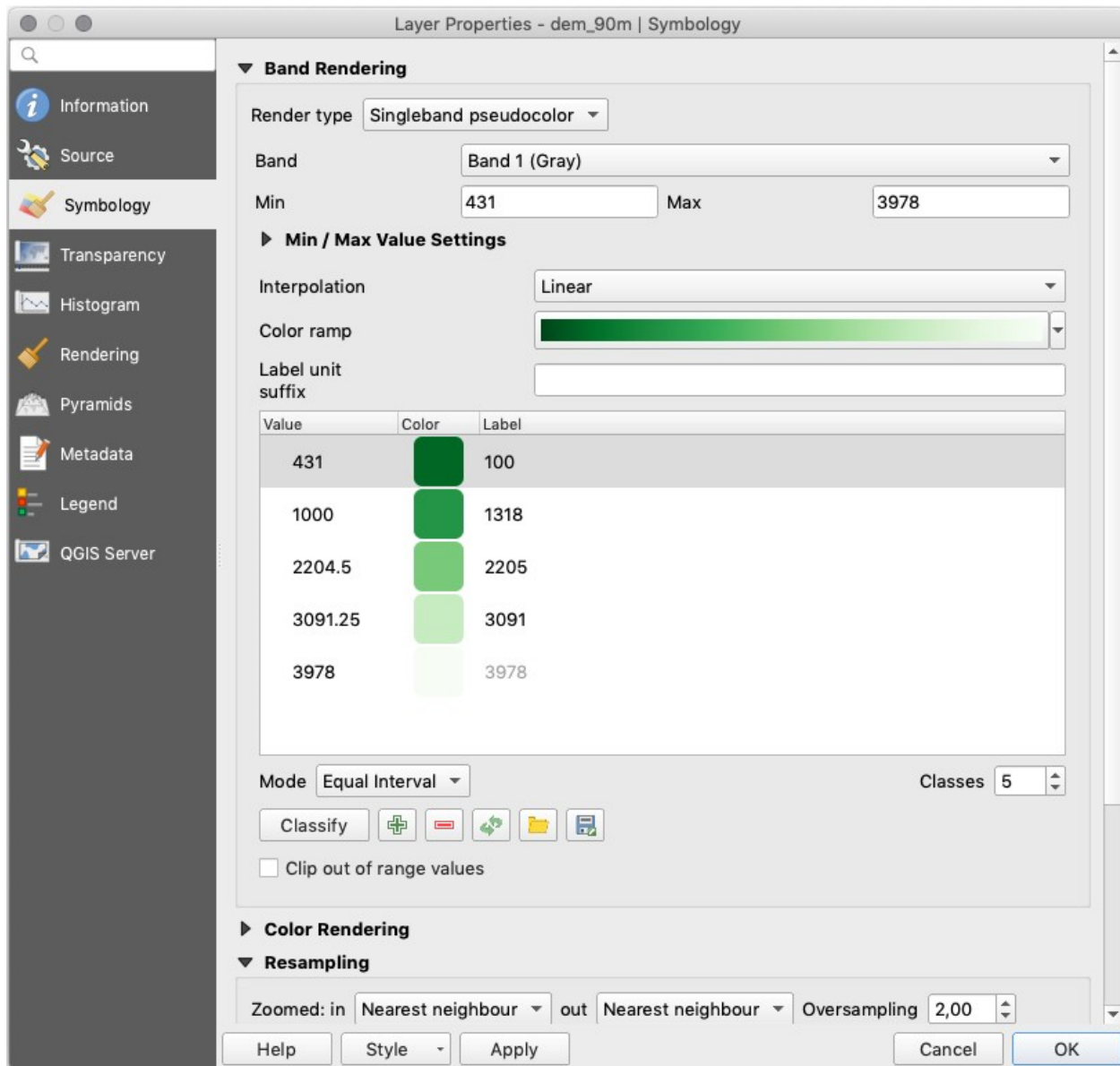
*Section author: Andre Mano Da Silva*

## SHADED TERRAIN MODELS

### 21.1 Depicting a DEM as Layer Tints

In QGIS a DEM will be shown as a grey-scale map by default, with black for the lowest and white for the highest values. In a map, we prefer to classify the heights and show them in so-called layer tints. You do this in the **Layer Properties > Symbolology** dialog:

1. As *Render Type* choose “Singleband pseudocolor”; and **Interpolation** = Linear ;
2. Choose a nice **Color ramp**, e.g. the “Greens”, as shown below. Make it go from dark for the lowest values to light for the highest values. For this, you have to click the menu button on the right of the “*color ramp*”, and choose the “*Invert colour ramp*” item.
3. Experiment with the colours, the classification steps, and other settings to get a nice layer tinting. After changing settings, do not forget to click the “*Classify*” button.



## 21.2 Using a DEM to Create a Hillshade

You can make a hillshade from the DEM as follows. Go to **Raster > Analysis > Hillshade** menu. See the image below. The tool will create a hypothetical illumination of a surface model by determining illumination values for each cell in an elevation raster. The shadows are modelled using the position of a light source, and considering the direction and steepness of the terrain. A hillshade can greatly enhance the visualisation of a surface for analysis or graphical display. By default, shadow and light are visualised as shades of grey associated with integer values from 0 to 255 (from black to white).

You can experiment with some of these options:

- The *Z factor* settings provide an optional exaggeration of the heights.
- The *Scale* option should be used if the height data units are not the same as the horizontal coordinate units. In our case, both are in meters. Still, if for example, one would have a DEM in meters and a projection in



latitude-longitude decimal degrees, one should use a *scale* = 111120.

- *ZevenbergenThorne*, *Combined* and *Multidirectional* are different algorithms to produce shadows.

Hillshade

Parameters Log

Input layer  
dem\_90m [EPSG:32632]

Band number  
Band 1 (Gray)

Z factor (vertical exaggeration)  
1,000000

Scale (ratio of vertical units to horizontal)  
1,000000

Azimuth of the light  
315,000000

Altitude of the light  
45,000000

☐ Compute edges

☐ Use ZevenbergenThorne formula instead of the Horn's one

☐ Combined shading

☐ Multidirectional shading

► **Advanced parameters**

Hillshade  
[Save to temporary file]

☒ Open output file after running algorithm

GDAL/OGR console call

```
gdaldem hillshade "/Users/barendkobben/Documents/BEWAARMAP/M-GEO/Online GIS/Data_BernerOberland/dem_90m.tif" /private/var/folders/zy/jv8h0qxs5blgxm57bn2fk7q00000gp/T/processing_4b235f36aba642c89767433c1e250c4f/1dac4cdc905b4e35833c6a7c912d6e77/OUTPUT.tif -of GTiff -b 1 -z 1.0 -s 1.0 -az 315.0 -alt 45.0
```

0%

Cancel

Help Run as Batch Process... Close Run

---

**Note:** In QGIS, the DEM Terrain models are part of the GDAL toolbox. Therefore the best place to find help and information about the hillshade option is at <http://www.gdal.org/gdaldem.html>.

---

*Section author: Barend Köbben & Andre Mano Da Silva*



## SISTAN BASIN DATASET

The table below contains the dates of acquisition for the time series used in the *Change Detection* exercise.

Band	Date (Y-M-D)	Date (M-Y)
1	05-01-02	Jan-2005
2	05-01-15	Jan-2005
3	05-02-21	Feb-2005
4	05-03-12	Mar-2005
5	05-03-21	Mar-2005
6	05-03-22	Mar-2005
7	05-04-03	Apr-2005
8	05-04-21	Apr-2005
9	05-04-26	Apr-2005
10	05-05-22	May-2005
11	05-06-13	Jun-2005
12	05-06-22	Jun-2005
13	05-07-08	Jul-2005
14	05-07-24	Jul-2005
15	05-07-31	Jul-2005
16	05-08-09	Aug-2005
17	05-08-16	Aug-2005
18	05-08-25	Aug-2005
19	05-09-03	Sep-2005
20	05-09-12	Sep-2005
21	05-09-19	Sep-2005
22	05-09-26	Sep-2005
23	05-10-01	Oct-2005
24	05-10-08	Oct-2005
25	06-01-05	Jan-2006
26	06-01-06	Jan-2006
27	06-01-11	Jan-2006
28	06-01-23	Jan-2006
29	06-01-26	Jan-2006
30	06-02-02	Feb-2006
31	06-02-19	Feb-2006
32	06-03-09	Mar-2006
33	06-03-15	Mar-2006
34	06-04-01	Apr-2006
35	06-04-11	May-2006

continues on next page

Table 22.1 – continued from previous page

Band	Date (Y-M-D)	Date (M-Y)
36	06-04-15	Jun-2006
37	06-04-22	Jul-2006

## BIBLIOGRAPHY

- [OB2009] Olenka Bilash. (2009, June). Best of Bilash. Improving second language education. Retrieved October 7, 2020, from <https://bestofbilash.ualberta.ca/inductivedeductive.html>
- [WC2010] Kmusser. (2010, May 13). Map showing the Helmand River drainage basin. In Wikipedia, The Free Encyclopedia. Retrieved 12:08, October 5, 2020, from <https://upload.wikimedia.org/wikipedia/commons/4/48/Helmandrivermap.png>
- [WC2020] Wikipedia contributors. (2020, September 14). Sistan Basin. In Wikipedia, The Free Encyclopedia. Retrieved 12:08, October 5, 2020, from [https://en.wikipedia.org/w/index.php?title=Sistan\\_Basin&oldid=978368838](https://en.wikipedia.org/w/index.php?title=Sistan_Basin&oldid=978368838)