POLITECNICO DI MILANO

Scuola di Ingenieria Civile, Ambientale e Territoriale

POLO TERRITORIALE DI COMO

MASTER of Science in
Environmental and Geomatic Engineering

Comparisons and problem identification on Local and Global Digital Elevation Models

Supervisor: Prof. Ludovico Biagi
Assistant Supervisor: Dott.ssa Alba Lucchese

MASTER Graduation Thesis by: Angela Patricia Ramos Clavijo
         Student Id. Number: 779550

Academic Year 2011/2013

Contents

Summary

Years of evolution have allowed the mankind to develop different tools and technologies that facilitate and improve several human activities, remarking that the technology cannot replace the human labor and that always this is needed to be supervised by an individual; it is the case of **Satellite and Airbone acquisition systems** which measure large areas in a matter of hours while a traditional method (theodolite or total station topographic survey) could take even months.

Nevertheless, these new methods present a common dilemma, namely, the lost of accuracy. This fact depends not only on the equipment that performs the capture but the resolution of the acquisition and the area extension which references to an acquisition in a global or local scale. Based on these facts, it is originated the topic of this document, aimed to compared different *global* Digital Elevation Models (DEMs) such as ASTER, SRTM and GMTED2010 with *local* Digital Elevation Models such as LiDAR, and then identified and analyze the problems local or a global model.

Being useful that all these data are downloaded for free (OpenData), more quantity of tests can be executed. Moreover, from the LiDAR page source it is possible to download either Digital Surface Model (DSM) and the Digital Terrain Model (DTM) which add to general scope of the study an extra scope. It consists in create a map of differences between them which likely shows the features present in the study area.

The whole process is developed in 10 chapters that are briefly explained below

Chapter 1. This chapter explains the general characteristics of the Digital Elevation Models, the way how they are acquire and the principal representations of these models.

Chapter 2. This chapter will explain three of the most important and well-known Digital Elevation Models regarding to a global scale. They are clustered or categorized low resolution when compared with a LiDAR model. Nevertheless, these models (especially SRTM and ASTER) in a general point of view generate a high resolution database taking into account that the acquisition is done for almost the entire Earth surface and the purchase of these products is low cost or simply for free.

Chapter 3. In the previous chapter global DEMs at a low resolution are mentioned, remarking that the acquisition is done for the entire Earth surface. However, when speaking locally, depending on the instruments that are used to make the measurements, this means high resolution as well accuracy. This chapter explains one of the most accurate methods to acquire information of local surfaces throughout remote sensing advantages.

Chapter 4. Since it deals with all this Geographic Information it is important to concentrate it in a unique Geographic Platform as GRASS GIS, in such a way that this information can be

read, visualized, edited and analyzed by means of FREE and Open Source software. That's why in this chapter will be explained the main functions of GRASS.

Chapter 5. This chapter will expose the general features of the study region which correspond to the Val di Sol area located in Trentino region. This general features correspond to climate, topographic, vegetation, etc.

Chapter 6. After characterized the study area, four processes are followed to compare Digital Surface Model with the respective Digital Terrain Model of the Val di Sole area. First of all, the download process and some general characteristics of the data set are explained then algebraic operations between DSM and DTM have been done with the idea of analyze statistically the obtained differences which in this case represent the different features (trees, buildings, bridge…) present in the study area and finally a categorization process is carried out based on a official land use map.

For this chapter two specific zones have been selected inside the whole Val di Sole area, namely, a small *mountain* sector and the main *valley*. This is done because one idea of the chapter is to compare the results between the Digital Models when the acquisition is performed in a mountain area and the plain.

Chapter 7. This chapter explains step by step the process of download SRTM, ASTER and GMTED2010 Digital Elevation Models from its respective websites and also the way to import them into the GRASS platform.

Chapter 8, Chapter 9 and Chapter 10. In this chapters are produce a final map that displays the differences between the global DEMs, namely, ASTER, SRTM and GMTED with respect to the local DEM. Furthermore, calculate a correlation index between the height (h) and the variation of height between the two models (Δh).

Keywords: Digital Elevation Model, Laser Imaging Detection and Ranging or Light Detection and Ranging (LiDAR), Global model, Local model.

Chapter 1

1. Digital Elevation Model (DEM) generation

The Digital Elevation Models (DEMs) are defined as a visual and mathematical representation of the surface that contains the altitudes of the terrain in a georeferenced frame, over an assigned reference surface.

To represent the mentioned frame, a gridding process is carried out through two well known methods, namely, metric or geographic, the former represents the surface with Cartographic Coordinates (N,E) and a constant step (ΔN, ΔE) while the further uses Geographic Coordinates (φ, λ) also with constant step ($\Delta \varphi$, $\Delta \lambda$). This regular grid allows storing the data in a **raster** file. An additional way to represent a Digital Elevation Model is **vector**-based triangular irregular network (TIN) which will be explained later.

But before going into details, it is important to highlight two areas of terrain representation that are directly related to the modeling of the Earth surface processes, namely, generation and interpretation of Digital Elevation Models. Their relationship to the overall context of digital terrain modeling is show in Figure 1. This figure clarifies the main functional connections between the tasks, particularly the interaction between Digital Elevation Model generation and Digital Elevation Model interpretation, and the overriding context provided by a wide range of applications.

Figure 1. The main tasks associated with digital terrain modelling
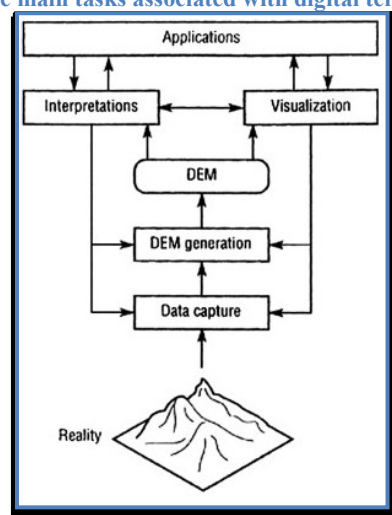
In order to understand better the previous figure, the main tasks are going to be briefly explained: it starts from the ***data capture*** that has a fundamental role during the Digital Elevation Models generation and that has improved through new developments in airborne and spaceborne remote sensing, such as laser and synthetic aperture radar systems and also

the development of the Global Positioning System (GPS) for ground data survey, until reach the applications of the mentioned Digital Elevation Model.

On the other hand, the elevation contours are not anymore the principal data source, as they were 20 years ago, for the interpolation of Digital Elevation Models, despite these data are widely available from existing topographic maps and that also can accurately reflect surface structure. Nowadays, models such as SAR/LIDAR are considered the main data source.

The next task is the ***Digital Elevation Model generation*** that involves the development of methods for interpolation and filtering of Digital Model data, these two processes continue to be a central area of digital terrain analysis, but the methods are now applied to a wider variety of data sources. These include traditional data sources such as points, profiles, contours, stream-lines, and break-lines, for which specific interpolation techniques have been developed, and remotely-sensed elevation data, for which various filtering procedures are required. Include in the task of Digital Elevation Model generation is a variety of associated Digital Model manipulation tasks for instance editing, resampling, and data structure conversion between regular grids and triangulated irregular networks (TINs), the two dominant forms of terrain representation. (Longley, Goodchild, Maguirre, & Rhind, 1999)

By other authors, the generation of Digital Elevation Models incorporates three interrelated tasks: (1) sampling the land surface (i.e. the gathering of height measurements); (2) creating a surface model from the sampled heights; and (3) correcting errors and artifacts in the surface model (Wilson, 2012).

Once the Digital Elevation Models is generated then ***interpretation and visualization*** is executed. In the interpretation of Digital Elevation Model are included scale analysis and resolution, terrain parameters such as elevation, slope, aspect, contour, drainage pattern, and a variety of terrain features i.e. mountain ranges, ridges, catchments, rivers, and valley, that can be constructed from Digital Elevation Models. With respect to visualization of Digital Elevation Model it can provide subjective assessments, such as perspectives views and intervisibility analyses for planning and monitoring applications. Interpretation and visualization of Digital Elevation Models can provide assessments of Digital Model quality which have direct implications for Digital Elevation Model generation and data capture, as indicated in Figure 1.

Mainly as each kind of model, Digital Elevation Models have multiples ***applications*** among a wide range of spatial scales, in civil engineering, planning and resource management, Earth sciences, and military studies, environmental and territorial field, creation of relief maps, terrain analysis in geomorphology and physical geography, flood or drainage modeling, etc.(Longley, Goodchild, Maguirre, & Rhind, 1999) In the following sections these different tasks will be exposed in detail.

1.1. Sources of elevation data

Three classes of source elevation data exist:
Surface-specific point elevation data: Surface specific point elevation, including high and low points, saddle points, and points on streams and ridges make up the skeleton of terrain that designate high and low edges, furnishing more information about the surface than just their coordinates; apart from streams and ridges, more types of surface-specific points and lines exist, such as peaks, pits, passes, ravines-lines and break-lines.(Longley, Goodchild, Maguirre, & Rhind, 1999)

Figure 2. Specific point elevation

They are an ideal data source for most interpolation techniques, including the previous mentioned triangulated methods and gridding methods. These data may be obtained by ground survey and by manually assisted photogrammetric stereo models which have been enhanced in the sense of accuracy of ground-surveyed data by the advent of the Global Position System. (GIScience at SFU, 2003)

A huge advantage of this data model is that provides topological information, though its construction can be time-consuming. Also another drawback of these accurate data is that they are available only for relative small areas. (Peucker & Douglas, 1975)

Countour line data: Contour data are still the most common terrain data source for large areas. Many of these data have been digitalized, in vector or raster format, from _existing topographic maps_ which are the only source of elevation data for some part of the world. As was aforementioned, this conversion of contour maps to digital formats follows these steps:

- Contour maps to vector format: The acquisition of a vector format can be performed through three possible techniques, namely manual digitizing, heads-up digitizing and automatic raster to vector conversion. The former works with a digitizing table that is managed by an operator. This person manually traces all the lines from the existing analog contour map using a cursor and creates a digital map on the computer that is characterized by several numbers of coordinated points, though the time-consuming and the low accuracy of the method makes it inefficient.

  The next technique has the same manually traced principle, but in this case the operator works directly on the computer screen using the scanned raster image as

backdrop. One advantage is a higher accuracy level since the raster image is scanned at high resolution; also disadvantages are present in terms of the time-consuming.

The further technique also so called automatic digitizing transforms the image already scanned and rasterized in a vector format using different algorithms based on image processing and pattern recognition techniques such as correct lines for scanning process, star and end point, line group of organized pixels and thickness of lines must be reduced to a single pixel ("skeletonization").(R2V, 1994)

- Contour maps to raster format: This process is simpler than the previous one because the scanner handles raster-map process. The spatial resolution and quality of the images depends on the resolution of acquisition device. (R2V, 1994)

Contour can be generated automatically from *photogrammetric stereo models either terrestrial or aerial*; but to understand better the system, some topics will be remarked about it. First of all, it is important to remember the definition of this technique as follows:

"Photogrammetry is the science of obtaining reliable information about the properties of surfaces and objects without physical contact with the objects, and of measuring and interpreting photographic image either analog or digital."(Schenk, 2005)

This science is based on artificial stereoscopic vision which employs the principle of the human vision process converting a 2D image in a 3D view. To obtain the third dimension it is necessary to capture an object from two different perspectives that follow a particular pattern or trend. (Lerma Garcia, 2002)

Figure 3. Stereo model

8
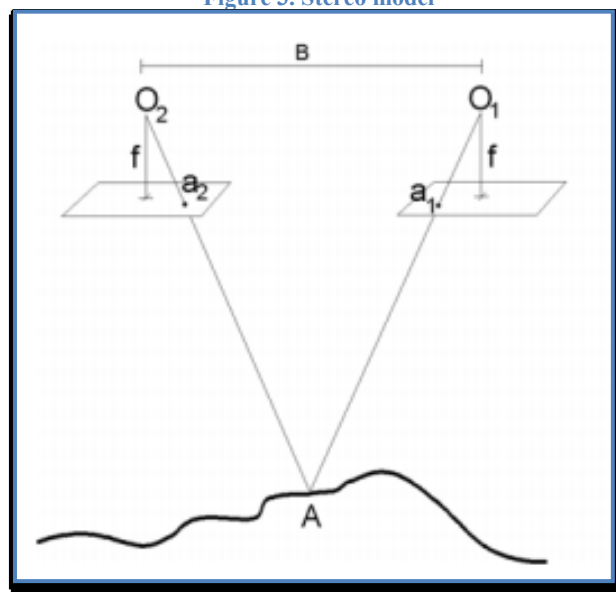
As was mentioned before the photogrammetric process can be performed on the terrain or in the air but in this section only the aerial method will be explained due to the common use and the convenience of it to produce maps or a Digital Elevation Models.

Figure 4. Areal survey
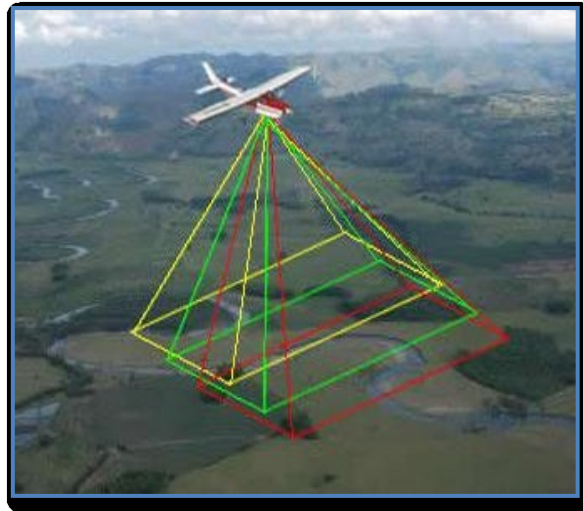
Then, two main steps have to be taken into account to elaborate those maps:

Image capture: It is necessary to program a flight plan which includes the direction of the airplane, number of photographs, flight altitude, etc. As is noticeable in

1. Figure 4 exist an overlapping between the captures, these must be both longitudinal and transversal with coverage among 60% and 20% respectively. During the time of exposure, a latent image is formed which is developed to a negative. At the same time diapositives and paper prints are produced.

2. Image processing: With the inputs obtained by the previous step the aero-triangulation and restitution processes take place. The aero-triangulation is the indirect determination of the orientation parameters of the captured images, namely spins (roll (ω), pitch(ϕ) and yaw(κ)), principal coordinates of each photograph (X_o, Y_o and Z_o) and also spatial coordinates of reference points in the surface; in Figure 5 this system is shown. As for the process of restitution, information from the stereoscopic model will be extracted by using a mobile index following the floating mark principle and finally a drawn map is produced.(Schenk, 2005)

Figure 5. Aero-triangulation
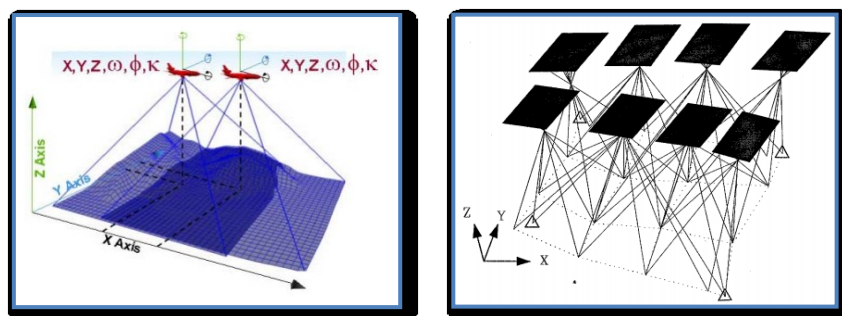
However, the analog images have been replaced by digital images and thus **digital photogrammetry** appears. The digital photogrammetry requires digital images which are taken by digital cameras; then these images go through an internal orientation where pixel coordinates (rows and columns) are transformed into image coordinates referenced to a coordinate system originated in the center of the projection. Finally, after an external orientation in which some control and homolog points are identified, the Digital Elevation Model is created when these two orientations are well combined by a software.

Remotely-sensed elevation data: Satellite sensors collect numerous images; through stereoscopic interpretation Digital Elevation Models are built. Stereoscopic methods have been applied, for instance, to ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) models, and to airborne and spaceborne synthetic aperture radar (SAR) such as SRTM (Shuttle Radar Topography Mission); these two acquisition techniques have a near-global coverage of the Earth surface at a lower resolution compare to the LiDAR (Laser Imaging Detection And Ranging) acquisition method that captures the images at a local scale.(Longley, Goodchild, Maguirre, & Rhind, 1999)

Remote sensing methods can provide broad coverage, but have present limitations in terms of acquisition and relief, which means: images acquired by this method are affected by systematic sensor and platform-induced geometry errors, introducing terrain distortions when the capture is taken outside the Nadir location of the sensor.

On the other hand, none of the sensor can measure the ground elevations underneath vegetation cover reliably. Even in the absence of ground cover, all methods measure elevations with significant random errors, which depend on the inherent limitations of the observing instruments, as well as surface slope and roughness. The methods also require accurately located ground control points to minimize systematic error.

Figure 6. Spaceborne acquisition
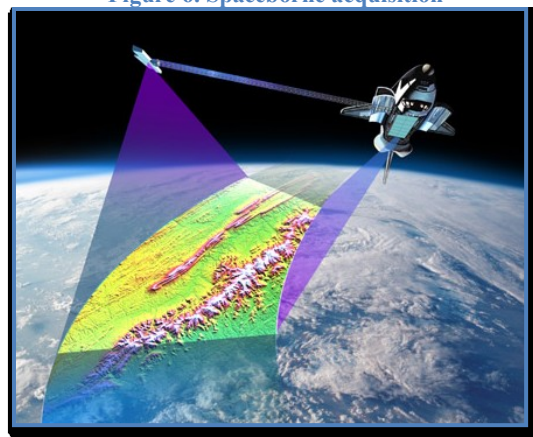
There are two classes of remote sensing technologies that are differentiated by the source of energy used to detect a target: passive systems and active systems. Passive systems detect radiation that is generated by an external source of energy, such as the sun, while active

systems generate and direct energy toward a target and subsequently detect the radiation (De Carolis, Active Microwave Instruments, 2012).

.

Figure 7. Active and Passive Systems
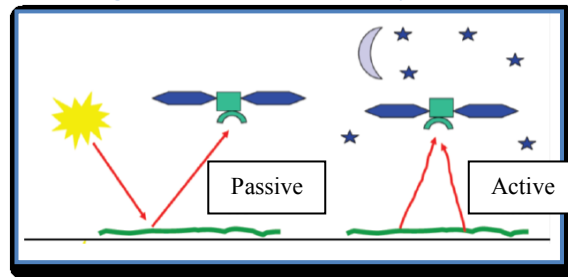
LiDAR systems are active systems because they emit pulses of light (i.e. the laser beams) and detect the reflected light. This characteristic allows LiDAR data to be collected at night when the air is usually clearer and the sky contains less air traffic than in the daytime. In fact, most LiDAR data are collected at night. Unlike radar, LiDAR cannot penetrate clouds, rain, or dense haze and must be flown during fair weather.

1.2 Digital Surface Model (DSM) vs. Digital Terrain Model (DTM)

Around these models are several definitions, however here are explained the most simple and practical ones, in such a way the lector does not get confused. To begin either the Digital Surface Model or the Digital Terrain Model are included in the definition of Digital Elevation Models the only difference between them is the data that each represents.

Digital Surface Model stands for the Earth's surface including all the objects on it, for instance, trees, plants, buildings, and other features elevated above the "Bare Earth". Figure 8. Airbornes and Spacebornes i.e. SRTM and ASTER Models by default capture Digital Surface data.

On the other hand, the Digital Terrain Model represents the bare ground surface without any object. An important fact that has to be taken into account in the DTM generation is the resolution attribute; higher the resolution capture, more feasible the DTMs generation which depends on the application of complex algorithms. LiDAR is a good example of high resolution capture, the advantages of using the further include the high density of sampling, high vertical accuracy, and the opportunity to derive these set of surface models given that some laser scanning systems can already provide at least two versions of the surface: the vegetation canopy (first returns) and ground surface (last returns).

Figure 8. DSM vs. DTM
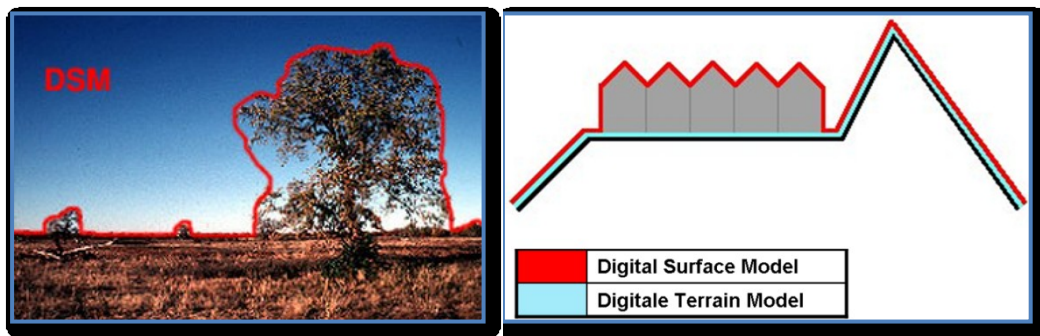
1.3 Digital Elevation Model representation

In order to understand better the representation of Digital Elevation Model, the grid concept is exposed below, due to the direct relation between the Digital Model represent and the grid file.

The principle of a **grid** is the division of a space into rectangular (regular tiles, grids squares, cells) that can be identified by two indices, one for the x-direction and the other for the y-direction.(Longley, Goodchild, Maguirre, & Rhind, 1999)

A grid coordinate system is defined in the map plane with axes parallel to the rows and columns of the grid and units equal to the sampling interval. The grid sample locations are at the whole integer grid coordinate points. To conform to mathematical array conventions, the grid coordinates (r,s) start at (0,0) in the upper left corner with r increasing to the right, and s increasing downward (this also conforms to digital image processing conventions). Note that the r coordinate corresponds to the grid column number j, and the s coordinate corresponds to the grid row number i. Figure 9 shows these notations.(Knowles)

Figure 9. Grid representation

Map coordinates   Grid coordinates
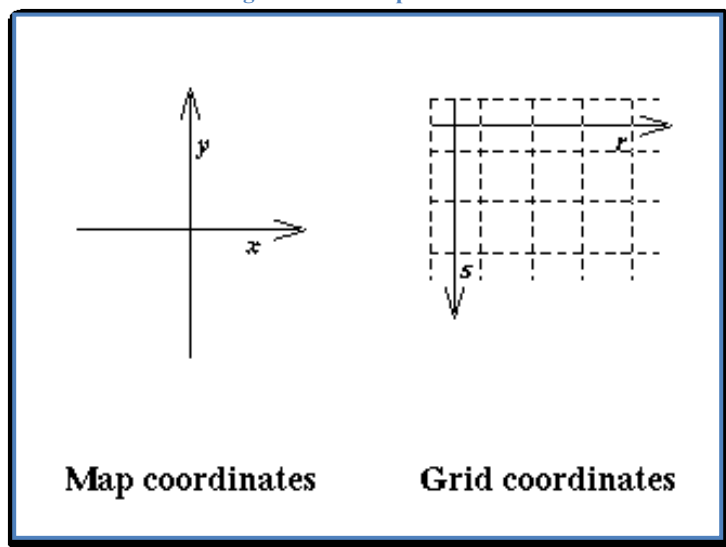
In the previous paragraph is mentioned the term coordinate system, which is the way of how to represent a surface. Two principal coordinate systems exist: Planar rectangular coordinates (x,y,z) and Geographic coordinates (ϕ, λ, h).

- Geographic coordinates (ϕ, λ, h): Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid. Lines of equal longitude are called meridians and they form ellipses (meridian ellipses) on the ellipsoid. Both lines form the graticule when projected onto a map plane. Note that the concept of geographic coordinates can also be applied to a sphere as the reference surface.

Figure 10. Geographic coordinates representation
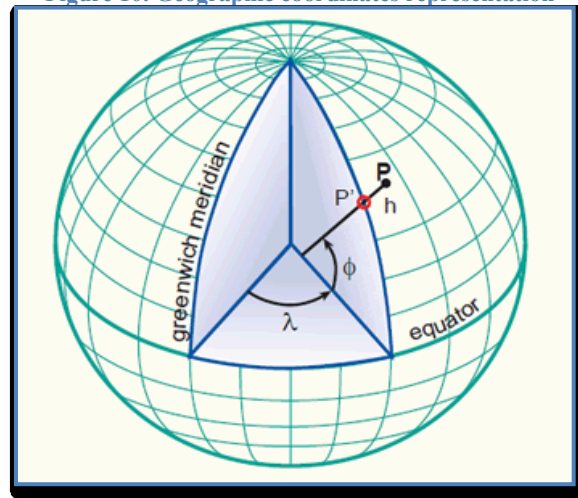
The latitude (ϕ), longitude (λ) and height (h) represent the 3D geographic coordinate system.

The **latitude** (ϕ) of a point P (Figure 10) is the angle between the ellipsoidal normal through P' and the equatorial plane. Latitude is zero on the equator ($\phi = 0°$), and increases towards the two poles to maximum values of $\phi = +90$ (90°N) at the North Pole and $\phi = -90°$ (90°S) at the South Pole.

The **longitude** (λ) is the angle between the meridian ellipse which passes through Greenwich and the meridian ellipse containing the point in question. It is measured in the equatorial plane from the meridian of Greenwich ($\lambda = 0°$) either eastwards through $\lambda = +180°$ (180°E) or westwards through $\lambda = -180°$ (180°W).

The **ellipsoidal height** (h) of a point is the vertical distance of the point in question above the ellipsoid. It is measured in distance units along the ellipsoidal normal from the point to the ellipsoid surface.(Knippers, 2009)

- Planar rectangular coordinates: Cartesian coordinates (x, y), are used to describe the location of any point in a map plane, unambiguously. It is a system of intersecting perpendicular lines, which contains three principal axes, called the X- , Y- and Z- axis.

The horizontal axis is usually referred to as the *X*-axis, the vertical the *Y*-axis and the Z-axis perpendicular to X- and Y- axis (note that the *X*-axis is also sometimes called **Easting**, the *Y*-axis the **Northing** and the Z-axis **Up**)(Figure 11). The intersection of the *X*- , *Y*- and Z-axis forms the **origin**. The plane XY is marked at intervals by equally spaced coordinate lines, called the map grid. Giving three numerical coordinates *x,y,z* for point P, it can be precisely and objectively specify any location P on the map.(Knippers, 2009)

Figure 11. Planar rectangular coordinates
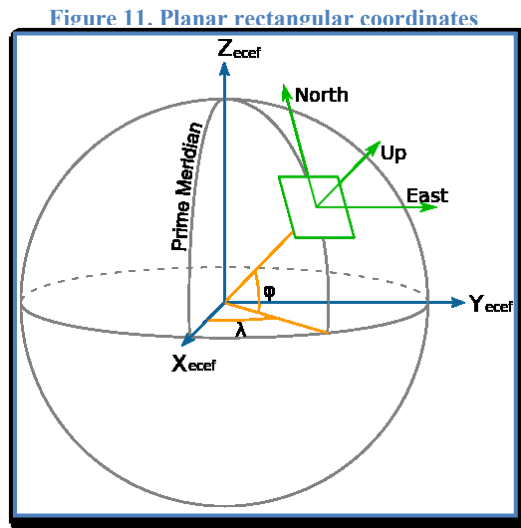
As was mentioned in the introduction of this chapter, two ways exist to represent Digital Elevation Models, namely, vector (TIN) and raster (geographic or metric), but first is necessary distinguish between vector and raster maps.

Vector map: Represented by means of coordinates. In this format, it is used the following geometric entities as primitives: point, line, surface. It can be represented the topological properties of data, allowing the description of spatial relationship between them (disjoint, touch, in, equal, contain, cross, overlap). (Brovelli, 2012)

RASTER map: Allows the representation of the world by means of a regular grid of small units, called pixels. The pixel contains information about a phenomenon; in the case that the phenomenon is not present in the pixel then information stored correspond to null data.(Brovelli, 2012)

Figure 12. Vector (left) & RASTER (right)

14
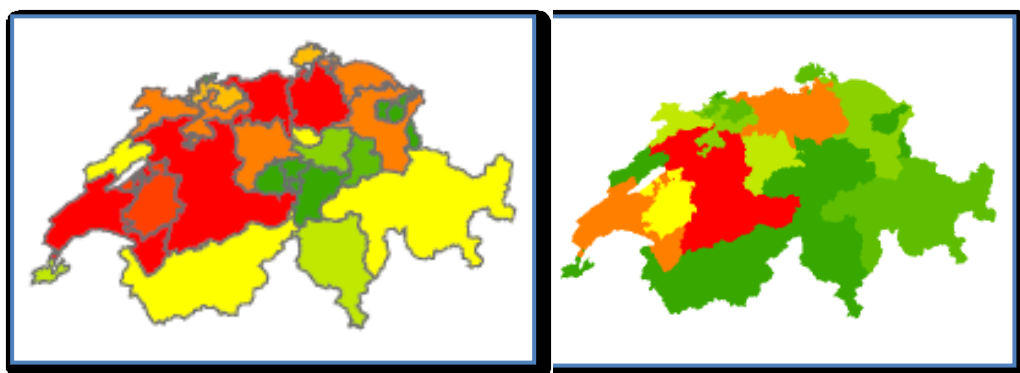
- Triangulated Irregular Network: Representation of the physical land surface or sea bottom, made up of irregularly distributed nodes and lines with three-dimensional coordinates (x, y, and z) that are arranged in a network of nonoverlapping triangles. These triangles are generated by combining the points detected in a network of triangular meshes that satisfy the criterion of Delaunay: a circle designed for the three points of a triangle must not contain other points.(Strutture Dati)

Figure 13. TIN representation – vector model
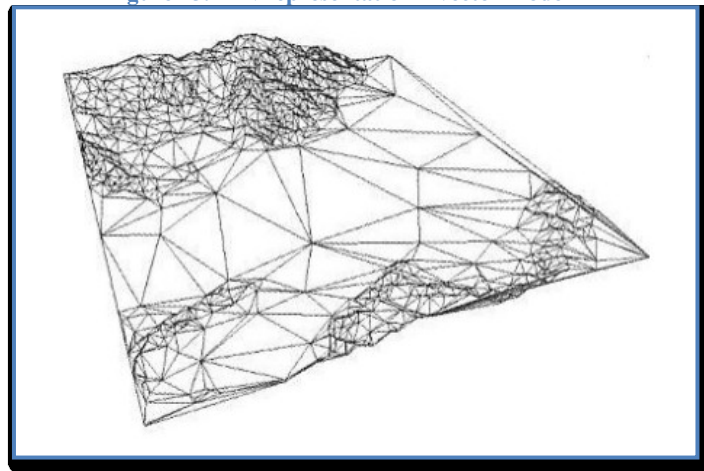
The main advantage of this representation is the ability of increasing the resolution only in the required areas, in the Figure 13 can be easily detected the aforementioned product.

- Geographic representation: The geographic grid is usually used for elevation models on a national, continental or global scale. It has the advantage of being "continuous", explicitly it does not have interruption due to the zonal change. The spacing grid nodes are usually expressed in sexagesimal seconds. The disadvantage is that this representation does not have a homogeneous metric density: indeed, the homogeneity in the metric and/or geographic sampling cannot be guaranteed. Thus switching from metric to geographic grid and vice versa is not just a transformation of the plane coordinates of the nodes of the grid, but interpolation methods must be used to recalculate heights at geographically or metrically equidistant points in the two respective cases.

- Metric representation: The metric grid is used to describe portions of limited territory or at a local scale (indicatively comprised within a zone of representation, Universal Transverse of Mercator, UTM), the sampling density is homogeneous in the two mapping directions X, Y, and also has the advantage, irrespective of the projection map used in a country, to be directly comparable, without coordinate transformations, with a local topographic database and / or digital cartography.(AAVV, 2009)

Figure 14. Geographic and/or Metric representation
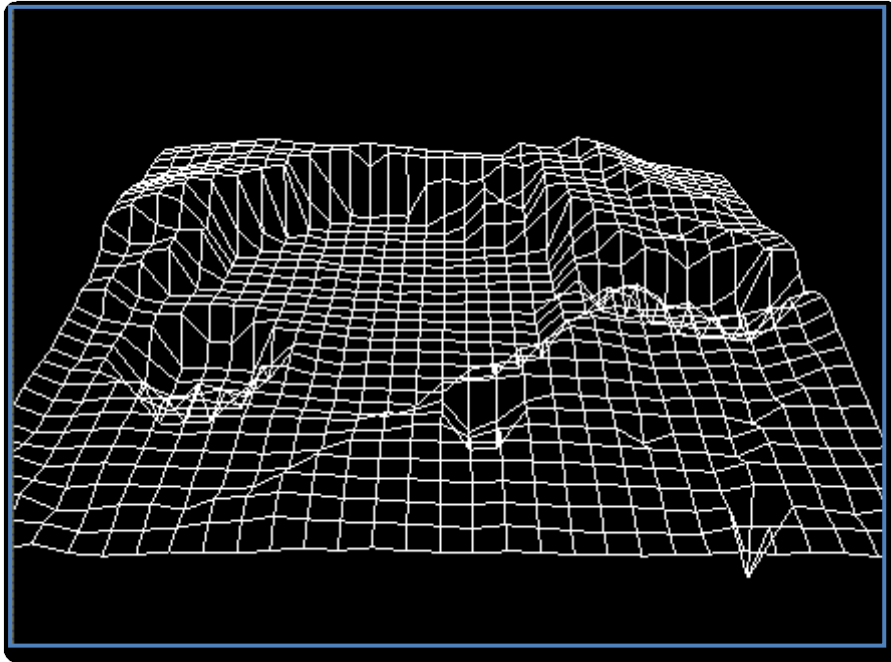
1.4 Resolution and accuracy

Many Digital Elevation Models can be generated and represented according to the aforementioned features; the awkward point is the choice of the proper model in order to gather all the information that is needed in a certain project. One important parameter that leads this choice is the spatial resolution.

The dimension of real surface represented by a single pixel determines the capability of the sensor to discriminate and reproduce a scene details. This is called Spatial resolution of a digital image, or **Pixel resolution**. Smaller size pixel means bigger detail level of a scene represented in digital images. Spatial resolution is linked to sensor characteristics and acquisition specifications. It determines the total level of information (detail) contained in a digital image, recorded in numerical form (De Carolis, From radiance to images: RS sensors characteristics, 2012).

High resolution	Low resolution
➢ Digital Elevation Models with a resolution better than 10m, i.e. LiDAR pixel resolution = 1m.	➢ Digital Elevation Models with a resolution greater or equal than 10m, i.e. ASTER pixel resolution = 30m.
➢ Digital Models give more detail as well as data to be stored.	➢ Digital Models give less detail and less data to be stored.
➢ Digital Models can cover less area.	➢ Digital Models can cover more area.

16

In the table are presented the principal DEMs with their associated resolution and accuracy (United State Geological Survey, 2010) & (Wilson, 2012).

Table 1. Resolution and accuracy DEMs

Model	Resolution (m)	Accuracy (m)
ASTER	30	7- 14 vertical / variable 7- 14 horizontal/ variable
SRTM	90	16 vertical 20 horizontal
LiDAR	1 -3	0.15 - 1 vertical 1 horizontal
SPOT	30	10 vertical 15 horizontal
GMTED	250 500 1000	26 - 30 vertical 29 - 32 vertical 25 - 42 vertical

Chapter 2

2. Global low resolution Digital Surface Model (DSM)

This chapter explains three of the most important and well-known Digital Elevation Models at the global scale. They are clustered or categorized low resolution when compared with a LiDAR model as can be seen in Table 1. Nevertheless, these models (especially SRTM and ASTER) in a general point of view generate a better resolution database taking into account that the acquisition is done for almost the entire Earth surface and the purchase of these products is low cost or simply for free.

2.1 Shuttle Radar Topography Mission (SRTM)

The Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate low-resolution digital topographic database of Earth. SRTM consisted of a specially modified radar system that flew onboard the Space Shuttle Endeavour during an 11-day mission in February of 2000. SRTM is an international project spearheaded by the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA). This GDEM version was released in 2003.(Jet Propulsion Laboratory, 2009)

The SRTM objective was to acquire a digital elevation model of all the Earth between about 60° north latitude and 56°south latitude, about 80 percent of Earth's land surface. In quantitative terms, the cartographic products derived from the SRTM data were to be sampled over a grid of 1 arc-second by 1 arc-second (approximately 30 m by 30 m) for USA area and 3 arc-second by 3 arc-second (approximately 90 m by 90m) for the rest of the world; with linear vertical absolute height error of less than 16 m and circular absolute geolocation error of less than 20 m. It is important to remark that all quoted errors are at 90% confidence level and also that they are referenced to the World Geodetic System 1984 (WGS84)/ Earth Gravitation Model 1996 (EGM96) geoid (Farr & others, 2007).

Moreover, SRTM provide a substantial upgrade over the primary source datasets used for GTOPO30, the older 3-arc-second DTED®1, and Digital Chart of the World (DCW) 1:1,000,000-scale cartographic data produced by NGA. The original SRTM data processing and editing is documented in Farr & others (2007). The void-filled SRTM data are a revised version of the original NGA dataset. The void-filled version includes additional spike/well removal using a threshold of 60 meters (instead of the original 100 meters) with respect to the surrounding terrain and the detection and removal of phase unwrapping errors that were remnants of the original raw radar data processing.(Danielson & Gesch, 2011)

SRTM employed two synthetic aperture radars, one C-band system (λ=5.6 cm; C-RADAR) and one X-band system (λ=3.1 cm; X-RADAR). The SRTM radars were designed to operate as single-pass interferometers, utilizing the SRL C- and X-band capabilities. For single-pass interferometry operations, each of the two SRTM radars was equipped with a supplementary receive-only antenna, in addition to the main transmit/receive antennas situated in the Shuttle's payload bay. The supplementary antennas were placed at the end of a retractable 60 m mast (Figure 15). During the Shuttle launch and landing the mast was stowed in a canister attached to the forward edge of the main antenna assembly. The technique employed is known as Interferometric Synthetic Aperture Radar (InSAR).(Bamler, 1999) The operational goal of C-RADAR was to generate contiguous mapping coverages, X-RADAR generated data along discrete swaths 50 km wide; X-RADAR was included as an experimental demonstration since the X-band radar had a slightly higher resolution and a better signal to noise ratio than the C-band system.

Figure 15. Shuttle Radar Topography Mission
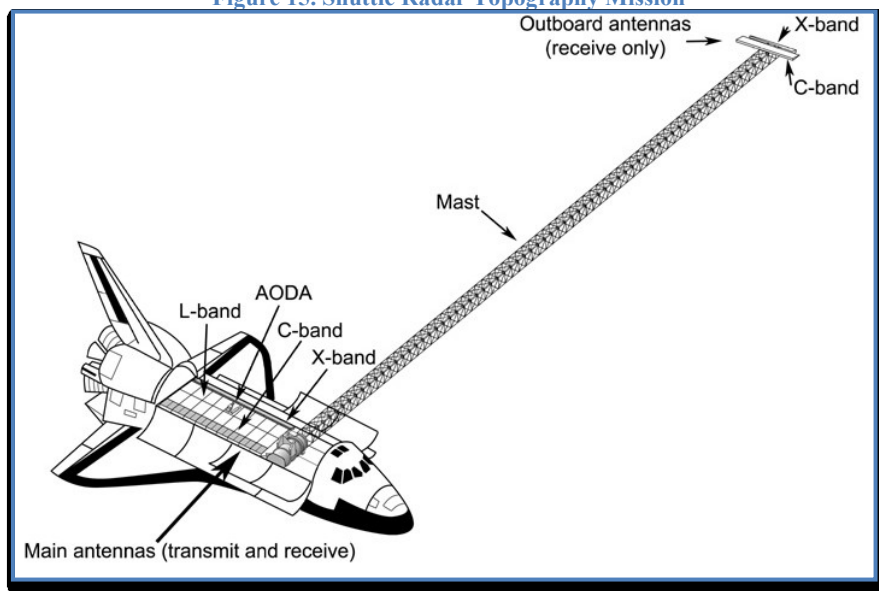
The C-band interferometer employed active antennas with electronic steering. To meet the goal of global gap-less mapping, the design goal was a swath width of 225 km; four sub-swaths are imaged quasi-simultaneously by periodically steering the beams from small to large look angles and back. At the same time two swaths (1 and 3 vs 2 and 4 in Figure 16) are imaged at different polarization (HH and VV). The X-SAR swath width is in the order of 45 km, the quality of X-band interferograms and DEMs is expected to be better than of those from C-band.(Lucca, Validation and Fusion of Digital Surface Models, 2011)
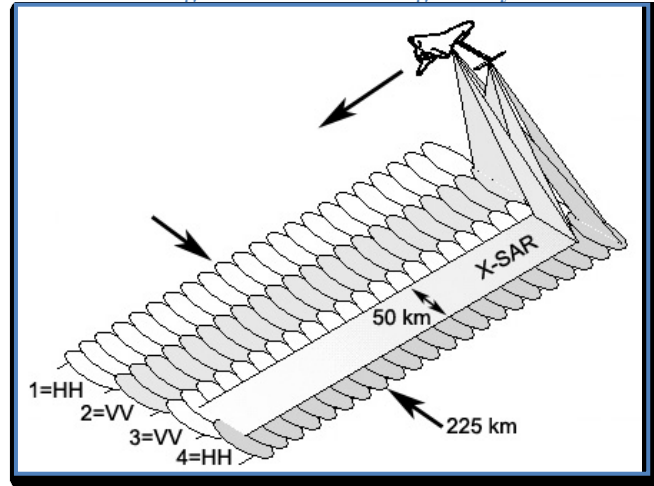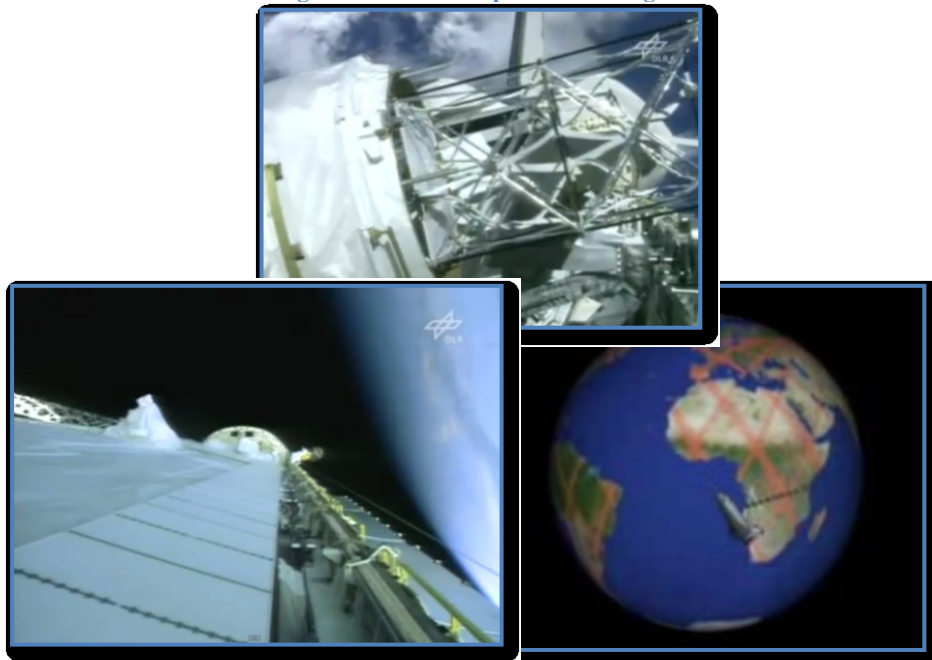
Figure 16. SRTM beam geometry

Figure 17. SRTM – Spaceborne images

An important application of this model is that SRTM is used as a base-model for the GMTED2010 generation. This fact is explained in section 2.3.

Drawbacks

- One disadvantage is that X-band data will not give full coverage as C-band.
- Presence of "no-data" in the elevation dataset due to mountain, desert, water or heavy shadow that prevented the quantification of elevation model. (Lucca, Validation and Fusion of Digital Surface Models, 2011)

2.2 Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is an imaging instrument built by Ministry of Economy, Trade and Industry (METI) of Japan and

operates on the NASA Terra platform which was launched in December 1999. Images are acquired in 15 spectral bands using three separate telescopes and sensor systems (Table 2).

Table 2. ASTER bands

Band	N. Bands	Resolution (m)
Visible	3 (Red, Green, Blue)	15
Near-Infrared (NIR)	1	15
Short-Wave-Infrared (SWIR)	6	30
Thermal Infrared (TIR)	5	90

It is important to remark that the Visible and Near-Infrared bands are acquired using a backward-looking telescope, thus providing along-track stereo coverage from which high-quality digital elevation models are generated as one of the ASTER standard data products. Moreover, this launch permitted the generation of two GDEM versions: the first released in 2009 and the second version published in 2011. (Aster Global DEM Validation, 2009)


Figure 18. Terra-platform (ASTER-sensors)
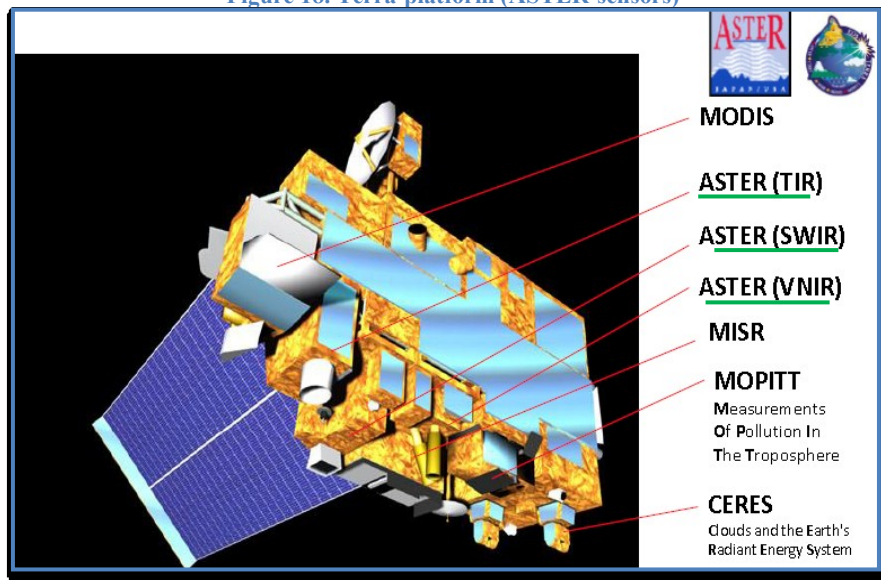
ASTER GDEM V1 coverage spans from 83 degrees north latitude to 83 degrees south, encompassing 99 percent of Earth's landmass. The methodology used to produce the ASTER GDEM involved automated processing of the entire 1.5-million-scene ASTER archive, including stereo-correlation to produce 1,264,118 individual scene-based ASTER DEMs, cloud masking to remove cloudy pixels, stacking all cloud-screened DEMs, removing residual bad values and outliers, averaging selected data to create final pixel values, and then correcting residual anomalies before partitioning the data into 1°-by-1° tiles.

The ASTER GDEM V1 is in GeoTIFF format with geographic lat/long coordinates and a 1 arcsecond (approximately 30 m) grid. It is referenced to the WGS84/EGM96 geoid. Pre-production estimated (but not guaranteed) accuracies for this global product were 20 m at 95% confidence for vertical data and 30 m at 95 % confidence for horizontal data.(ASTER GDEM Readme File – ASTER GDEM Version 1, 2009)

On the other hand, the improved GDEM V2 adds 260,000 additional stereo-pairs, improving coverage and reducing the occurrence of artifacts. The refined production algorithm provides improved spatial resolution, increased horizontal and vertical accuracy, and superior water body coverage and detection. The ASTER GDEM V2 maintains the GeoTIFF format and the same gridding and tile structure as V1, with 30-meter postings and 1°-by-1° tiles.(Jet Propulsion Laboratory , 2004)

Figure 19 shows the orbiting Terra-platform (left) and a 3D model produced by ASTER capture (right).

Figure 19. ASTER – Spaceborne images
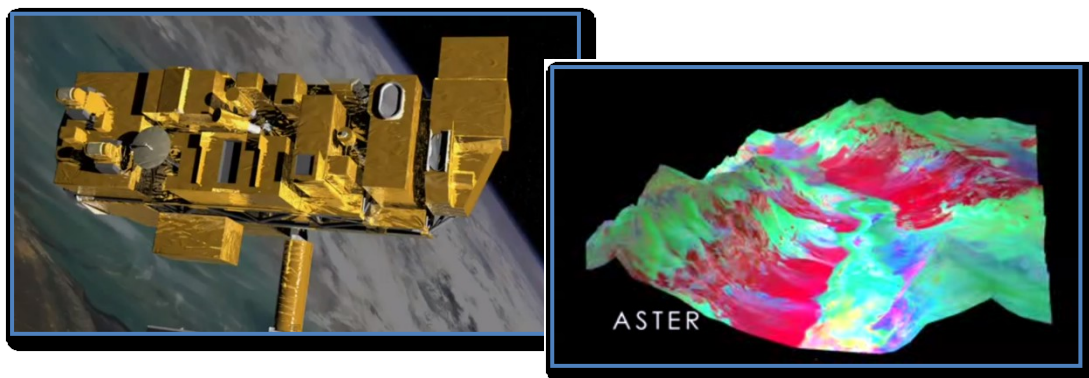
The main drawback in ASTER model is the presence of "no-data" in some areas due to constant cloud cover.(Lucca, Validation and Fusion of Digital Surface Models, 2011)

2.3 Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)

The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) is an enhanced replacement for Global 30 Arc-Second Elevation (GTOPO30), Global Land One-Km Base Elevation (GLOBE) model and other comparable 30-arc-second resolution global model, using the best available data. This work has been done by the United States Geological Survey (USGS) and the NGA. This new model has been generated at three separate resolutions (horizontal post spacing) of 30-arc-seconds (≈ 1 Km), 15-arc-seconds (≈ 500 meters) and 7.5-arc-seconds (≈ 250 meters).

GMTED2010 is based on data derived from the following data:

1. Shuttle Radar Topography Mission (Digital Terrain Elevation Data Version 2, DETD V.2)
2. Antarctica satellite radar and laser altimeter DEM
3. Shuttle Radar Topography Mission (Digital Terrain Elevation Data Version 1, DETD V.1)
4. Canadian Digital Elevation Data Version 3
5. Canadian Digital Elevation Data Version 1
6. Greenland satellite radar altimeter Digital Elevation Model

7. National Elevation Dataset - Alaska
8. 15-arc-second Satellite Pour l'Observation de la Terre (SPOT) 5 Reference 3D
9. Global 30-Arc-Second Elevation Dataset (GTOPO30)
10. National Elevation Dataset
11. GEODATA 9 second Digital Elevation Model Version 2

The main source is NGA's SRTM Digital Terrain Data; for the geographic areas outside the SRTM coverage area and to fill in remaining holes in the SRTM data, the rest of the data sources are used. (Danielson & Gesch, 2011)

To perform the mentioned operation, first of all it is necessary to detect the voids in the data, the corresponding elevation posts are voided out and then systematically replaced with an alternate source of elevation data primarily from non-SRTM DTED, NED, and SPOT5. In places where no acceptable alternate source data were available and where the size of the void and the surrounding terrain were appropriate, interpolation was used to fill the void. Although most of the data voids in the 1-arc-second SRTM data have been filled by NGA, some residual voids remain where suitable source data at the required spatial resolution were not available and no interpolation was done. For these areas, GMTED2010 production included filling the residual voids in the SRTM DTED 2 dataset with the vertical heights referenced to the EGM96 geoid.

One of the main advantages of GMTED2010 over GTOPO30 is that seven new raster elevation products are available at each resolution, namely:

1. Minimum elevation map: It has been produced using the minimum elevation as an aggregation method. Which means, subtract the minimum value from a processing window.

2. Maximum elevation map: It has been produced using the maximum elevation as an aggregation method. Which means, subtract the maximum value from a processing window. An example is shown in Figure 20, this example also can be applied for the minimum, mean and median aggregation methods.

3. Mean elevation: It has been produced using the mean elevation as an aggregation method. Which means, subtract the mean value from a processing window.

4. Median elevation: It has been produced using the median elevation as an aggregation method. Which means, subtract the median value from a processing window.

5. Standard deviation: It has been produced using a combination of two functions in ArcGIS. A Blockstd function, finds the standard deviation for the specified posts defined by the neighborhood blocks, and sends the computed standard deviation to the post location in the corresponding blocks on the output raster grid (Figure 21). The

Blockstd output was then generalized to the desired output resolution of 30, 15 or 7.5 arc-second using neighbor resampling.

6. Break-line emphasis: It was used to produce reduced resolution products that maintain stream (channel) and ridge (divide) characteristics as delineated in the full resolution source data. Breakline emphasis maintains the critical topographic features within the landscape by maintaining any minimum elevation or maximum elevation value on a breakline that passes within the specified analysis window. Remaining elevation values are generalized using the median statistic. The breakline emphasis methodology can be summarized into three steps:

   a. Topographic breaklines (ridges and streams) are extracted from the full resolution DEM and then used to guide selection of generalized values.
   b. Full resolution streams are automatically thresholded, which enables easy extraction of the level one through five Strahler stream orders.
   c. Full resolution ridges are extracted by selecting the flow accumulation values that are equal to zero. Using focal and block image processing functions, ridges are thinned so that only critical divides are maintained.

7. Systematic subsamples: It was used to produce a reduced resolution version at each of the output grid spacing. The systematic subsample product was computed using the Resample function in ArcGIS with the nearest neighbor option. (Danielson & Gesch, 2011)

Figure 20. Aggregate example using the maximum value (3 x 3 processing window)
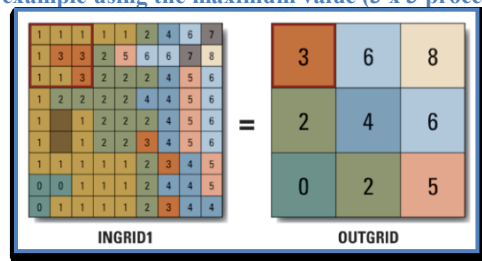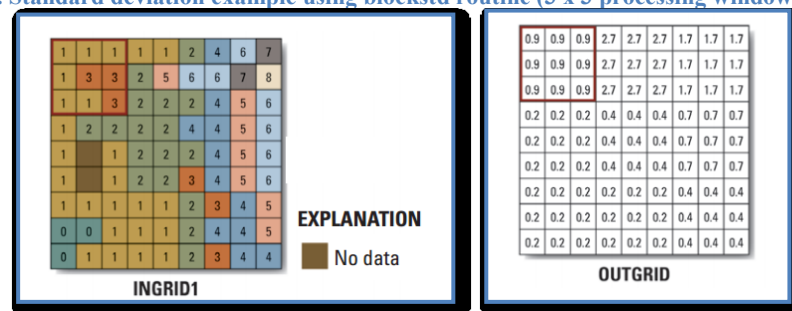

Figure 21. Standard deviation example using blockstd routine (3 x 3 processing window

24

Also the heights in the products are referenced to the WGS84/EGM96 geoid. Metadata have also been produced to identify the source and attributes of all the input elevation data used to derive the output products.

Remark: Aggregation methods used to generate the new maps produce reduced resolution data that represent the minimum, maximum, mean, and median of the full resolution source elevations within the aggregated output cell. The statistical-based products were generated using the Aggregate function within ArcGIS. The Aggregate function resamples an input raster grid to a coarser resolution based on a specified aggregation strategy (Minimum, Maximum, Mean, or Median). (Danielson & Gesch, 2011)

Then with these new products (21 Digital Elevation Models) the enhancement of GTOPO30 is done by realizing that it is passed from a single Digital Elevation Model, namely, GTOPO30 which has an horizontal grid spacing of 30 arc seconds (approximately 1 kilometer) to a multi-resolution and multi-product Digital Model as is the case of GMTED2010. In the Figure 22 these results are shown (USGS, 2012).

Figure 22. GMTED2010 vs. GTOPO30
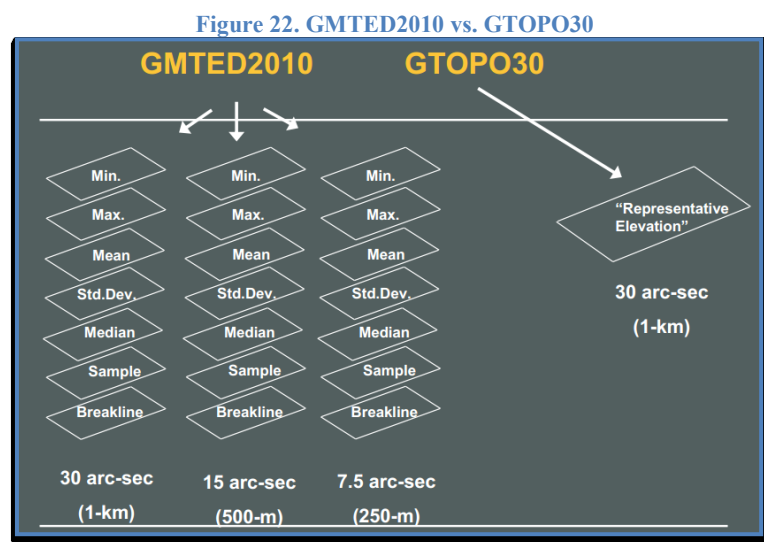
The new 21 products (7 products x 3 resolutions)(Figure 22) will be used in a variety of application situations. For example, the maximum elevation product could be used for the global calculation of airport runway surface heights or to determine the height of vertical obstructions. The minimum elevation product is useful for determining stream channel areas and the water surface. Comparison of the minimum and maximum products will provide a measure of the local relief in a given area. The standard deviation product provides a measure of the texture, or local variation in elevation, of the landscape surface. The breakline emphasis products will be useful for most hydrologic applications that involve watershed extraction and surface streamline routing. The remaining products, specifically the mean and systematic subsample products, will be useful for general visualization exercises and all-purpose morphological processing.

The three spatial resolutions allow the user to choose a level of detail and corresponding data volume that are appropriate for a specific application. (Danielson & Gesch, 2011)

Figure 23. GMTED2010 product
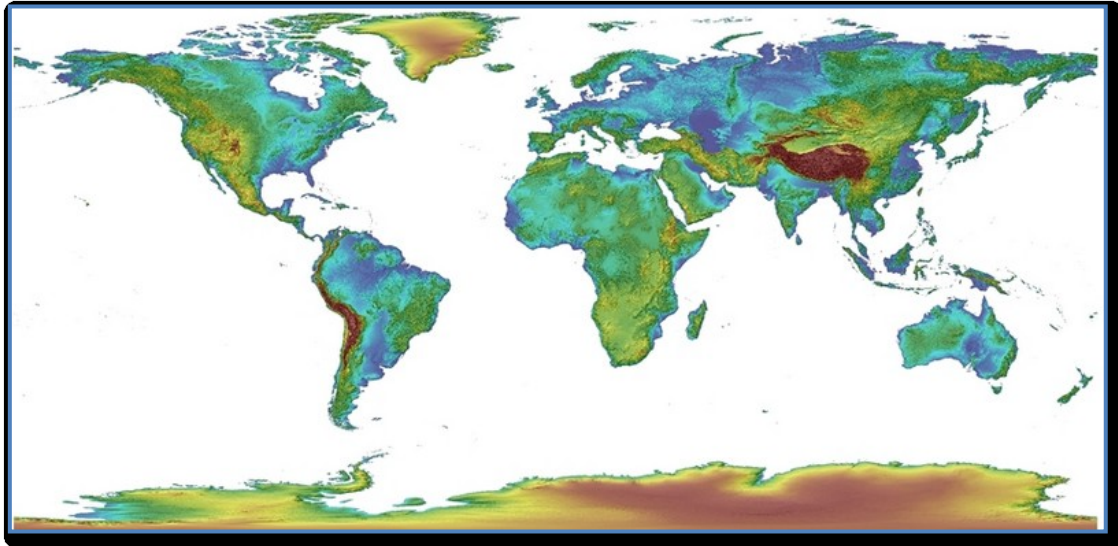
2.4 OpenData

The principal organization that deals with OpenData is Open Source Geospatial Foundation, or **OSGeo**, is a not-for-profit organization whose mission is to support the collaborative development of open source geospatial software, and promote its widespread use. The foundation provides financial, organizational and legal support to the broader open source geospatial community. It also serves as an independent legal entity to which community members can contribute code, funding and other resources, secure in the knowledge that their contributions will be maintained for public benefit. OSGeo also serves as an outreach and advocacy organization for the open source geospatial community, and provides a common forum and shared infrastructure for improving cross-project collaboration.(OSGeo, 2013)

http://www.osgeo.org/ → OSGeo Home Page

OSGeo Mission Statement: To support the collaborative development of open source geospatial software, and promote its widespread use.

OSGeo Goals:
The following more detailed goals support the overall mission:

- To provide resources for foundation projects - eg. infrastructure, funding, legal.
- To promote freely available geodata - free software is useless without data.
- To promote the use of open source software in the geospatial industry (not just foundation software) - eg. PR, training, outreach.

- To encourage the implementation of open standards and standards-based interoperability in foundation projects.
- To ensure a high degree of quality in foundation projects in order to build and preserve the foundation "brand".
- To make foundation and related software more accessible to end users - eg. binary "stack" builds, cross package documentation.
- To provide support for the use of OSGeo software in education via curriculum development, outreach, and support.
- To encourage communication and cooperation between OSGeo communities on different language (eg. Java/C/Python) and operating system (eg. Win32, Unix, MacOS) platforms.
- To support use and contribution to foundation projects from the worldwide community through internationalization of software and community outreach.
- To operate an annual OSGeo Conference, possibly in cooperation with related efforts (eg. EOGEO).
- To award the Sol Katz award for service to the OSGeo community.

(OSGeo, 2013)

Since the previous technologies have been implemented and improved to measure the Earth Surface and, in some way, to facilitate these data in regions where maps do not exist, the possibility to share them for FREE is considered; following the definition of Open Data that says:

"A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike."

The next picture shows what kinds of OpenData are available around the world.

Figure 24. Kinds of Open Data

Geodata: The data that is used to make maps – from the location of road and buildings to topography and boundaries.

Cultural: Data about cultural works and artefacts – for example titles and authors – and generally collected and held by galleries, libraries, archives and museums.

Science: Data that is produced as part of scientific research from astronomy to zoology.

Finance: Data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds etc).

Statistics: Data produced by statistics offices such as the census and the key socioeconomic indicators.

Weather: The many types of information used to understand and predict the weather and climate.

Environment: Information related to the natural environment such presence and level of pollutants, the quality and rivers and seas.

Transport: Data such as timetables, routes, on-time statistics. (Open Knowledge Foundation, 2012)

Moreover, as is said above for this study four Digital Elevation Models (DEMs) were required which have been downloaded from the subsequent links:

- ✓ SRTM 90 & ASTER 30

  http://gdex.cr.usgs.gov/gdex/
- ✓ GMTED2010

  http://earthexplorer.usgs.gov/
- ✓ LiDAR

http://www.territorio.provincia.tn.it/portal/server.pt/community/LiDAR/847/LiDAR/23954→Acceso al WebGIS pubblico

In the following sections, the download process (for each Digital Elevation Model) is explained step by step.

Chapter 3

3. Local high resolution Digital Surface Model (DSM)

In the previous chapter low resolution global DEMs are mentioned, remarking that the acquisition is done for the entire Earth surface. However, in a local scale, depending on the instruments that are used to make the measurements, they will result with a high resolution as well high accuracy, as is the case of LiDAR Digital Elevation Model. The next part explains one of the most accurate methods to acquire information of local surfaces throughout remote sensing advantages.

3.1 Laser Imaging Detection and Ranging or Light Detection and Ranging (LiDAR)

LiDAR is now commonly accepted as the most powerful technique for delivering highly accurate height data; these have become a vital component of many geospatial strategies and are widely used by government agencies and commercial sector for a variety of applications from flood risk modelling to wind farm site selection, archaeological exploration to urban planning.(Geomatics Group, 2013)

LiDAR is an optical remote-sensing technique that uses laser light to densely sample the surface of the earth, producing as is said before, highly accurate position (x,y,z) measurements. LiDAR, primarily used in airborne laser mapping applications, is emerging as a cost-effective alternative to traditional surveying techniques such as photogrammetry. LiDAR produces mass point cloud datasets that can be managed, visualized, analyzed, and shared using different GIS desktop softwares.

The major hardware components of a LiDAR system include a collection vehicle (aircraft, helicopter, vehicle, and tripod), laser scanner system, GPS (Global Positioning System), and INS (inertial navigation system). An INS system measures roll, pitch, and heading of the LiDAR system.

LiDAR is an active optical sensor that transmits laser beams toward a target (buildings, ground pipelines, highways, street furniture, power lines, railway tracks, vegetation…) while moving through specific survey routes. The reflection of the laser from the target is detected and analyzed by receivers in the LiDAR sensor. These receivers record the precise time from when the laser pulse left the system to when it is returned to calculate the range distance between the sensor and the target. Combined with the positional information (GPS and INS), these distance measurements are transformed to measurements of actual three-dimensional points of the reflective target in object space.(ArcGIS Resources, 2013)

Figure 25. LiDAR acquisition
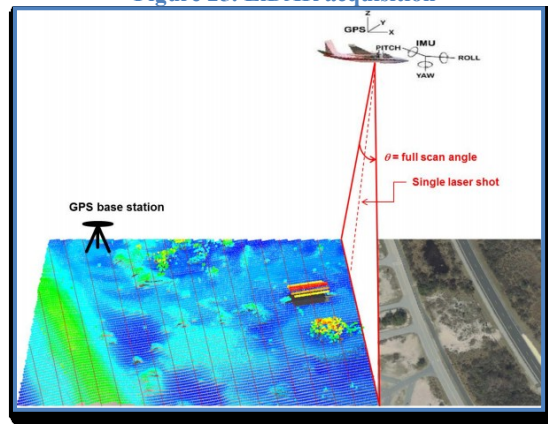
The three-dimensional coordinates (x,y,z) of the target objects are computed from: the time difference between the laser pulse being emitted and returned, the angle at which the pulse was "fired" and the absolute location of the sensor on or above the surface of the Earth.

3.2 LiDAR calculations

Two LiDAR typology exist: pulse and continuous laser emission [Baltsavias,1999a]; in the former, commonly used, laser pulses are emitted at regular time interval, the latter uses a continuously emitted signal basing the computation on phase difference between emitted and received ray beam.

In pulse laser the direct measure correspond to the time elapsed between laser emission and come back after terrain reflection (t_{travel}); the distance R between aircraft and ground is easily obtained considering the speed of the beam c (equal to light speed).(Lucca, Validation and Fusion of Digital Surface Models, 2011)

The emission system sends off a laser beam only when all the previously emitted ray have been recorded after ground reflection, from this derives a condition concerning emission frequency (*f*): the time elapsed from the emission and the reception must be less than signal period.

The direct consequence of this condition is the need of a compromise between flying height and point density:

- High flying height → low point density: few wide strips;
- Low flying height → high point density: high number of strips.

In continuous emission the signal may be sinusoidal, so the time of travel depends on phase difference ( ) between transmitted and received signal; the distance between aircraft and ground is computed as in the pulse system.

(Lucca, Validation and Fusion of Digital Surface Models, 2011)

LiDAR systems have advanced considerably. Early commercial units were capable of 10,000 points per second (10 kilohertz) and were large and bulky. Newer systems are more compact, lighter, have higher angular precision, and can process multiple laser returns in the air (i.e., a second laser shot is emitted before returns from the previous laser shot are received), allowing for pulse rates of over 300,000 per second (300 kilohertz).(Carter & others, 2012)

3.3 LiDAR attributes and cloud data

Additional information is stored along with every x, y, and z positional value. The following LiDAR point attributes are maintained for each laser pulse recorded: intensity, return number, number of returns, point classification values, points that are at the edge of the flight line, RGB (red, green, and blue) values, GPS time, scan angle, and scan direction. The following table describes the attributes that can be provided with each LiDAR point.

LiDAR attribute	Description
Intensity	The return strength of the laser pulse that generated the LiDAR point.
Return number	An emitted laser pulse can have up to five returns depending on the features it is reflected from and the capabilities of the laser scanner used to collect the data. The first return will be flagged as return number one, the second as return number two, and so on.
Number of returns	The number of returns is the total number of returns for a given pulse. For example, a laser data point may be return two (return number) within a total number of five returns.
Point classification	Every LiDAR point that is post-processed can have a classification that defines the type of object that has reflected the laser pulse. LiDAR points can be classified into a number of categories including bare earth or ground, top of canopy, and water. The different classes are defined using numeric integer codes in the LAS files.
Edge of flight line	The points will be symbolized based on a value of 0 or 1. Points flagged at the edge of the flight line will be given a value of 1, and all other points will be given a value of 0.
RGB	LiDAR data can be attributed with RGB (red, green, and blue) bands. This attribution often comes from imagery collected at the same time as the LiDAR survey.
GPS time	The GPS time stamp at which the laser point was emitted from the aircraft. The time is in GPS seconds of the week.
Scan angle	The scan angle is a value in degrees between -90 and +90. At 0 degrees, the laser pulse is directly below the aircraft at nadir. At -90 degrees, the laser pulse is to the left side of the aircraft, while at +90, the laser pulse is to the right side of the aircraft in the direction of flight. Most LiDAR systems are currently less than ±30 degrees.
Scan direction	The scan direction is the direction the laser scanning mirror was traveling at the time of the output laser pulse. A value of 1 is a positive scan direction, and a value of 0 is a negative scan direction. A positive value indicates the scanner is moving from the left side to the right side of the in-track flight direction, and a negative value is the opposite.

Post-processed spatially organized LiDAR data is known as point cloud data. The initial point clouds are large collections of 3D elevation points, which include x, y, and z, along with additional attributes such as GPS time stamps. The specific surface features that the laser encounters are classified after the initial LiDAR point cloud is post-processed. Elevations for the ground, buildings, forest canopy, highway overpasses, and anything else that the laser beam encounters during the survey constitutes point cloud data.(ArcGIS Resources, 2013)

3.4 From DSM to DTM

One of the main advantages of LiDAR is that it allows obtaining Digital Terrain Models (DTM) from Digital Elevation Models (DEM) through clever algorithms applying some filtering techniques, remembering that a DTM represents the terrain surface without any objects.

Several methods and algorithms have been presented to solve the previous task, here are mentioned the most principals:

- Mathematical morphology: Objects as buildings and trees are eroded from the surface using digital filters. A conventional method, called opening, has been implemented, improved and automated.(Dold, 2000).

   Mathematic morphology used to filter LIDAR data, is evolved from processing gray images. Based on different elevation, it is transform LIDAR data to corresponding grayscale raster images and then introduce mathematic morphology to process the data. Because of big elevation difference between surface feature point and ground, there will be great difference in image gray. With this difference can use mathematic morphology to detect surface features. Kilian processed LIDAR data with mathematic morphological opening operator and detect the lowest points in the filter window. Usually the lowest points are regarded as ground points. It is critical to set size of filter window. If the window is set too small, it can well maintain terrain in detail, but it can only filter out low shrubs, ground vehicles and other small objects. For large urban buildings, it can't work well. If the window is set too big, it can filter large object, but it can't maintain terrain in detail.

   Therefore, ideal window is not only be small enough to ensure that the details of the terrain, but also be large enough to remove the large groups of buildings. To solve this problem, Kilian use a series of variable windows to iterate on the operation with LIDAR data. He gave associated weights to laser pin points that were considered as ground points. In general the ground point will be given a higher weight; non-ground points will be given a lower weight (Kilian, 1996). Through repeated iterations, topographical trend surface is estimated by the corresponding value.(Li, Sun, & Yan, 2011). Figure 26 shows an example for a DSM and the generated DTM

Figure 26. Example – mathematical morphology unfiltered (left) and filtered (right)
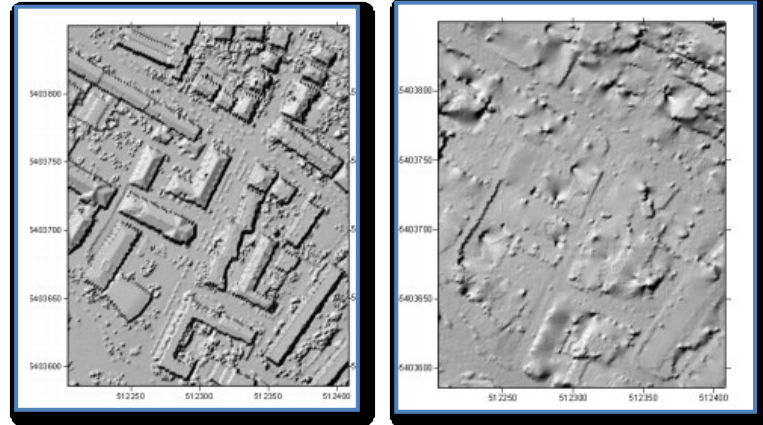
- Progressive densification algorithms: These filters at first identify some points belonging to the ground and then, depending on those, classify more points as ground. Usually the points used as seeds are the ones with lower height. Additional ground points are determined by investigating their neighbor in the reference surface.

- Surface based algorithms: these filters use a parametric surface that iteratively approaches to hypothetical bare earth. The surface is modified depending on the influence of the individual input points.

- Clustering-Segmentation algorithms: These filters are based on the idea that a cluster of points belongs to an object if it has height values greater than its neighbors. In these cases the classification is performed in two steps: at first segmentation is carried out and then the segments are divided in different classes depending on the differences in height between segments.(Lucca, Validation and Fusion of Digital Surface Models, 2011)

Chapter 4

4. Geographic Resources Analysis Support System (GRASS GIS)

Since it deals with all this Geographic Information it is important to concentrate it in a unique Geographic Platform as GRASS GIS, in such a way that this information can be read, visualized, edited and analyzed by means of Free and Open Source Software (FOSS).

4.1.   Definition

GRASS GIS, commonly referred to as GRASS (Geographic Resources Analysis Support System), is a **free** Geographic Information System (GIS) software used for geospatial data management and analysis, image processing, graphics/maps production, spatial modeling, and visualization. GRASS GIS is currently used in academic and commercial settings around the world, as well as by many governmental agencies and environmental consulting companies. GRASS GIS is an official project of the Open Source Geospatial Foundation.

GRASS GIS contains over 350 modules to render maps and images on monitor and paper; manipulate raster, and vector data including vector networks; process multispectral image data; and create, manage, and store spatial data. GRASS GIS offers both an intuitive graphical user interface as well as command line syntax for ease of operations. GRASS GIS can interface with printers, plotters, digitizers, and databases to develop new data as well as manage existing data. (GRASS GIS, 2012)

GRASS is released under the GNU General Public License (GPL), and it can be used on multiple platforms, including Mac OS X (Mountain Lion), Linux (GNOME, Ubuntu) and other UNIX compliant platforms (32/64bit), additionally MS-Windows.

4.1.1. Free and Open Source meaning
An Open Source first of all, offers full insights into the system, thus users can analyze the methods internally used, understand their functionalities, modify methods to their purpose, check error and, in case required, correct or update methods (having access to the source code).

Another general purpose of the Open Source release under GPL is the opportunity for users to implement their own ideas or to suggest modifications which could be implemented by everyone familiar with programming.

This fundamental tool has been used not only by university researchers but also by different user groups such as local and national governmental organizations, as well as privacy companies. Currently, GRASS is in the top-ten list of the biggest available Open-Source programs. (Neteler, 2000)

The official website to download GRASS

http://grass.osgeo.org/download/software/

4.1.2. Main capabilities

- RASTER 2D analysis: Automatic rasterline and area to vector conversion, Buffering of line structures, Cell and profile dataquery, Colortable modifications, Conversion to vector and point data format, Correlation / covariance analysis, Expert system analysis, Map algebra (map calculator), Interpolation for missing values, Neighbourhood matrix analysis, RASTER overlay with or without weight, Reclassification of cell labels, Resampling (resolution), Rescaling of cell values, Statistical cell analysis, Surface generation from vector lines.

- RASTER 3D (voxel) analysis: 3D data import and export, 3D masks, 3D map algebra, 3D interpolation (IDW, Regularised Splines with Tension), 3D Visualization (isosurfaces), Interface to Paraview and POVray visualization tools.

- Vector analysis: Contour generation from raster surfaces (IDW, Splines algorithm), Conversion to raster and point data format, Digitizing (scanned raster image) with mouse, Reclassification of vector labels, Superpositioning of vector layers .

- Point data analysis: Delaunay triangulation, Surface interpolation from spot heights, Thiessen polygons, Topographic analysis (curvature, slope, aspect), LiDAR.

- Image processing system: Canonical component analysis (CCA), Color composite generation, Edge detection, Frequency filtering (Fourier, convolution matrices), Fourier and inverse fourier transformation, Histogram stretching, IHS transformation to RGB, Image rectification (affine and polynomial transformations on raster and vector targets), Ortho photo rectification, Principal component analysis (PCA), Radiometric corrections (Fourier), Resampling, Resolution enhancement (with RGB/IHS), RGB to IHS transformation, Texture oriented classification (sequential maximum a posteriori classification), Shape detection, Supervised classification (training areas, maximum likelihood classification), Unsupervised classification (minimum distance clustering, maximum likelihood classification).

- DTM-Analysis: Contour generation, Cost / path analysis, Slope / aspect analysis, Surface generation from spot heigths or contours

- 3D Visualization system: 3D surfaces with 3D query (NVIZ), Color assignments, Histogram presentation, Map overlay, Point data maps, RASTER maps, Vector maps, Zoom / unzoom –function.

- DBMS integrated (SQL) with dbf, PostgreSQL, MySQL and Sqlite drivers (GRASS GIS, 2012)

4.2. GRASS structure

The organization of GRASS data is based on a hierarchical structure and has been built in this efficient way in order to manage the access of the multiples users to the stored data.

1. **DATABASE**: Contains all GRASS data.

Each GRASS project is organized in a "Location" directory with subsequent "Mapset" directories

    1.1. **LOCATION**: Defines a coordinate system and a rectangular boundary for a Project. Every location has a PERMANENT directory which stores some basic information about the whole location, and is a good place to park base files.

        MAPSET(s): Used to subdivide data by user names or subregions or access rights and each user has the possibility to edit data.

            1.1.1. **PERMANENT**: Is a standard mapset that contains the definitions of the location. May also contains general cartography since it is visible to all the other mapsets. (Lucca & Luana, GRASS GIS intro, 2012)

The following figures sketch the GRASS structure for this study

Figure 27. GRASS structure
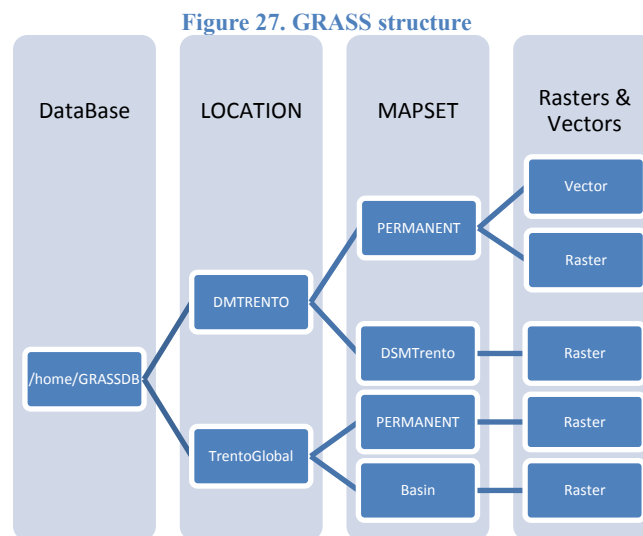
In the first chapter two kinds of data formats have been mentioned that evidently can be read by GRASS: vectors and rasters; each format has its own way to be stored in GRASS through different folders.

4.2.1. Vector data

Each vector map is stored in its own directory: the directory has the same name of the vector map. Each vector map contains the following files:

- cidx: Contains the topological index.
- coor: Binary file that contains the features coordinates and categories.
- dbln: Contains information about database connection (driver, database, table, column, etc...).
- head: Contains the vector header.
- hist: Contains information about the vector history: author, date, executed commands, etc...
- topo: Binary file containing topology.

4.2.2. RASTER data

RASTER maps are recorded in a matrix where each element represents a pixel with an **integer** or floating **point value**. RASTER maps are distributed in different files inside thematic subdirectories:

- cat: Contains category information
- cell – fcell: Is the file with the numerical matrix
- cellhd: Contains maps headers
- colr: Contains maps color informationon
- hist: Contains information about raster map history: author, date, executed commands, etc...

Chapter 5

5. Study Case: Trentino Region

Autonomous province of Trento is an Italian province generally known as Trentino. It is located in the North-East country side of Italy, in the Trentino Alto Adige Region. Trentino is divided in 217 municipalities and its capital is Trento. The province has an extension of 6.213 Km^2 with a total population of about 530.000 inhabitants (ISTAT-2010).

The main characteristic of this area is the presence of multiple reliefs namely, valleys and mountains belonging to the Alps ("Tridentine Alps") and the Dolomites ("Pale Mountains"). The most important valley is the well known Adige Valley that crosses Trento Province from North to South, the principal towns (including the capital) lay in this valley as it is the largest one. Aside from Adige Valley there are other valleys such as "**Val di Sole**", "Val di Non", "Val di Giudicarie", "Val di Fiemmea", "Val di Fassa", "Val di Lagarina", "Val di Mocheni", "Val di Sugana" which compose part of the Trentino reliefs.

Among other important characteristics of the province "La Marmolada" takes place. The highest mountain in the Dolomites is 3.343 meter above sea level. Other high mountains include the Monte Baldo, Carè Alto, Cermis, Crozzon di Brenta, Hintere Eggenspitze, Latemar, Paganella, Piz Boè, Presanella, Punta San Matteo and Vezzana. Due to this topography variability not only in climate but in environmental characteristics exists. (Wikipedia, 2013)

Figure 28. Location of Trentino Province
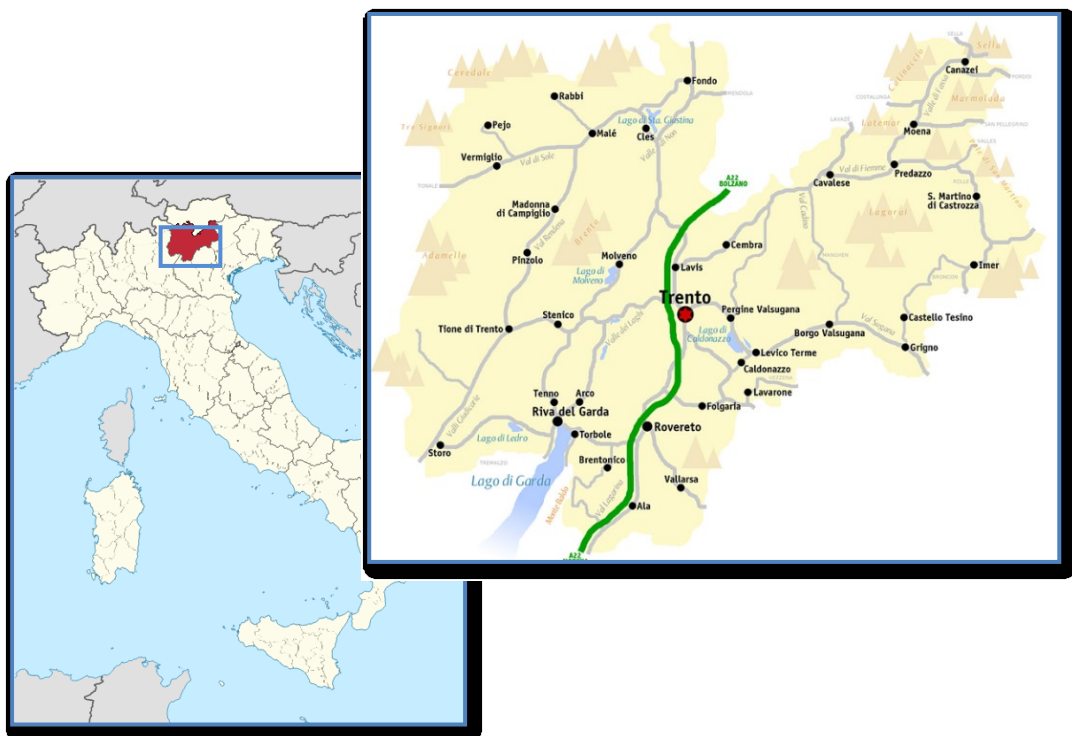
Morphology

The mountain ranges belong to different rock formations or origins. For instance, the western side of the province (bounding with Lombardy Province) is dominated by perennial snow, a massive structure and the "Adamello" and "Presanella" rock group which is characterized by the present of diorite tonalite. Surrounding the "Val di Sole" larger glacier areas and a new group of rocks "Ortles-Cevedale" composed by crystalline schist are found.

Contrary to the west part of the Adige River, the east part is composed by a dolomites complex such as "La Marmolada" and the massive "Castello" that fits in "Sella" group. Most of the dolomites group has the presence of dolomite and magnesium. Mainly, this area has variety of structural shapes and landscapes.(Wikipedia, 2013)

Climate and vegetation

Trentino Province experiments harsh climate due to the glacier and the abundant water bodies, subalpine valleys and small plains from the sub-Mediterranean climate which allows the olive crops.

The climate of this area is considered a transition between the semicontinental and the alpine climate. The temperature ranges from -10°C during winter to 25°C or more in summer.

Regarding to the plain relief there 4 distinguishable climate areas:
- Sub-Mediterranean area - the Lake Garda area and low Sarca Valley. It is the mildest part of the region, with mild winters. The vegetation is composed of olive trees, oaks and cypresses.
- Subcontinental area - climate transition that characterizes the valley bottom, with colder winters and snowy. The vegetation consists primarily of chestnut, beech and silver fir.
- Continental area - in the alpine valleys (such as the "Val di Fassa" or "Val di Sole") with cold winters and short summers, rainy and vegetation composed mainly of conifers.
- Alpine area - At the timberline (1800/900 m.a.s.l), with snow that lingers along the year.(Wikipedia, 2013)

The surface is covered by 50% of forest approximately 300.000 hectares. Some characteristics of this vegetation are explained bellow:

- Scots pine: Is a species of pine native to Europe and Asia, It is readily identified by its combination of fairly short, blue-green leaves and orange-red bark. Pinus sylvestris is an evergreen coniferous tree growing up to 35 m in height and 1 m trunk diameter when mature.(Wikipedia, 2013)

-Larch: Larches are conifers in the genus Larix, in the family Pinaceae. Growing from 20 to 45 m tall, they are native to much of the cooler temperate northern hemisphere, on lowlands in the north and high on mountains further south.(Wikipedia, 2013)

- Abies alba: Is a large evergreen coniferous tree growing to 40–50 m (exceptionally 60 m) tall and with a trunk diameter of up to 1.5 m. The largest measured tree was 68 m tall and had a trunk diameter of 3.8 m. It occurs at altitudes of 300-1,700 m (mainly over 500 m), on mountains with a rainfall of over 1,000 mm.(Wikipedia, 2013)

- Swiss pine: The Swiss pine, Swiss stone pine or Arolla pine, Pinus cembra, is a species of pine tree that occurs in the Alps. It typically grows at 1,200 meters to 2,300 meters altitude. It often reaches the alpine tree line in this area. The mature size is typically between 25 meters and 35 meters in height, and the trunk diameter can be up to 1.5 meters. (Wikipedia, 2013)

5.1.   Val di Sole area

In this project the "Val di Sole" has been selected because here is possible to evaluate the mountain area and the valley area in terms of data acquisition (DSM-DTM), taking into account the information supplied by the Trentino Province as OpenData.

Figure 29. Location of the "Val di Sole"
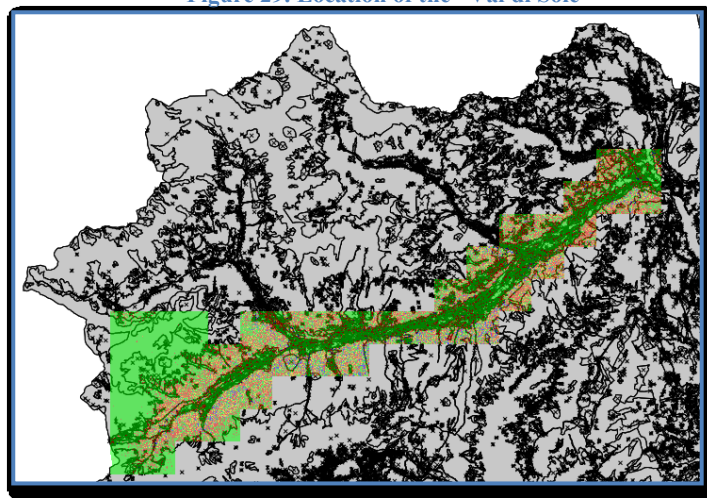
The "Val di Sole" is crossed by the Noce River. As mentioned previously, the valley belongs to different rock formations such as "Ortles-Cevedales" and "Brenta" group. The surface of the valley is of approximately 610 Km^2 and its population of about 15.000 inhabitants.

The main urban center in the region is Male, considered as an economic and activity center. Along this valley there are others 13 urban centers.

In order to have an overview of the slope's region a simplex profile has been produced (Figure 30). The graph presents an approximated valley length of about 40 Km with the

maximum and minimum height at 1888 m.a.s.l and 534 m.a.s.l respectively and the mean slope equal to 3.8%.
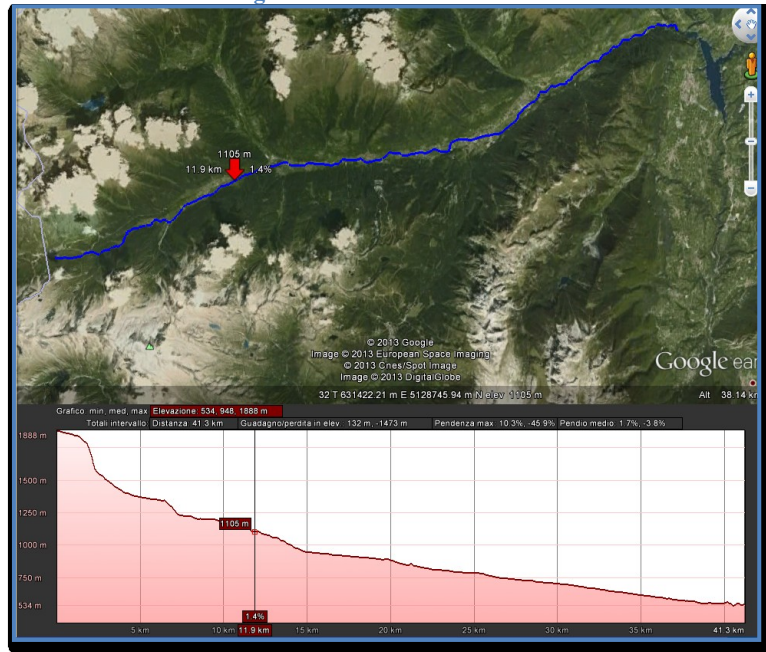
Chapter 6

6. Comparisons between DSM/DTM in Val di Sole area (Trentino subregion-LIDAR)

After characterized the study area, four processes are followed to compare Digital Surface Model with the respective Digital Terrain Model of the Val di Sole area. First of all, the download process and some general characteristics of the data set are explained then algebraic operations between DSM and DTM have been done with the idea of analyze statistically the obtained differences which in this case represent the different features (trees, buildings, bridge…) present in the study area and finally a categorization process is carried out based on a official land use map.

For this chapter two specific zones have been selected inside the whole Val di Sole area, namely, a small *mountain* sector and the main *valley*. This is done because one idea of the chapter is to compare the results between the Digital Models when the acquisition is performed in a mountain area and the plain.

6.1. Data source

The basic step is to get the Digital Surface Model and Digital Terrain Model regarding to this area, which are provided by the GeoCartographic Trentino Province Web Site ("Portale Geocartografico Trentino- WebGIS Pubblico"(Figure 31)).

Five steps exist to download the data:

1. Access to the WebGIS

Figure 31. WebGIS Pubblico
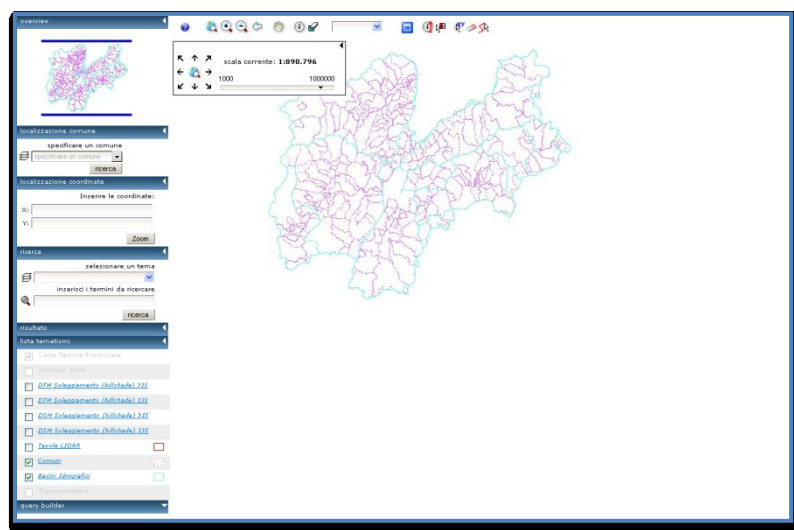
2. Activate LiDAR grid and select the needed cells. Each cell has a number which will identify the Digital Models either Surface or Terrain.
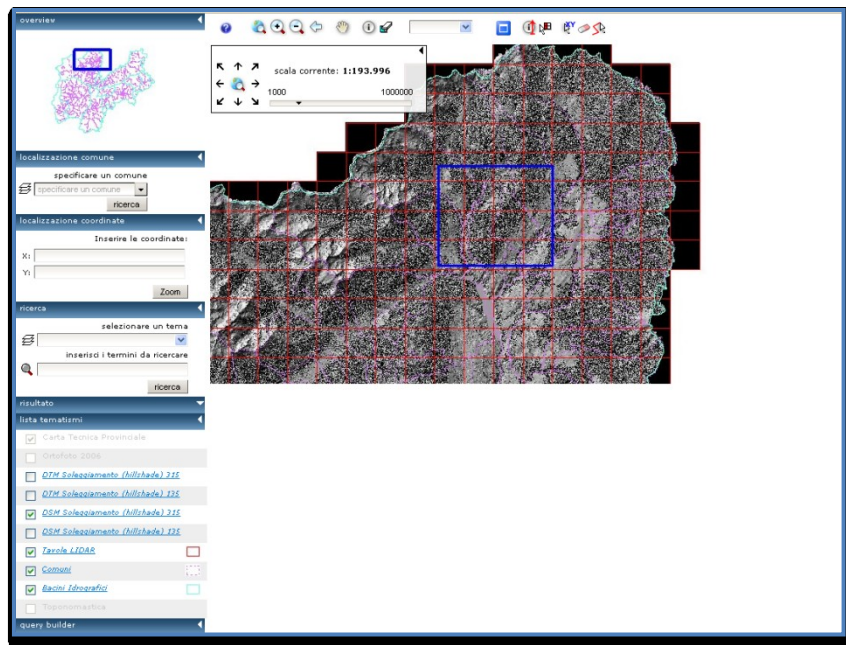
3. Once the cells are selected, tick Digital Surface Model ASCII and Digital Terrain Model ASCII and click "Stima". This step just calculates the size of the file and the download time.

4. Finally, click "Scarica". The data set is downloaded in a zip file which contains the DSM & DTM.

The outputs of this operation are:

- Mountain area: 11 files for each model (DSM – DTM, Table 3) have been selected to cover an area of about 44 Km^2 in the mountainous part of the "Val di Sole".

- Valley area: 36 files for each model (DSM – DTM, Table 3) have been selected to cover an area of about 144 Km^2 in the valleous area of "Val di Sole".

In general, each raster file has a size equal to 2020m x 2020m and a resolution at 1m or 2m.

The cells used for this chapter are present in the following table:

Table 3. DSM & DTM files

Mountain	Valley					
320	178	250	331	372	414	505
321	179	251	332	373	415	506
322	213	289	368	374	416	507
364	214	290	369	375	458	508
365	215	291	370	412	459	556
366	249	330	371	413	460	557
409						
410						
411						
456						
457						

6.2. Difference between DSM and DTM

Using the GRASS 7.0 Geographic Information System platform, the quantitative analysis that, as a first step, consists in calculate algebraic differences between the DSM and the DTM in such a way that the features present in the region can be highlighted like, for example trees, buildings, bridges, routes, etc., will be developed.

GRASS works in a particular platform where is necessary to define important characteristics such as:

1. Location (DM TRENTO): It refers to the geographic materialization of the "Val di Sole" through the assignment of a region taking into account the projection and a specific work area. In this case,

Projection: EPSG: 32632 WGS 84 / UTM zone 32N

Bounds: North 512082 m South 5084744 m East 710004 m West 613457 m

Note: EPSG, European Petroleum Survey Group maintains and publishes a dataset of parameters for co-ordinate reference system and co-ordinate transformation description. The EPSG Geodetic Parameter Dataset has been included as reference data in the heritage Oil & Gas UK and SEG positioning data exchange formats, the GeoTIFF interchange format for geo-referenced raster imagery, the IHS Energy Iris21, PPDM and POSC Epicentre data models. The dataset is distributed through a web-based delivery platform, or in an MS Access relational database and SQL script files

2. Mapset → PERMANENT

Once these parameters have been fixed the 47 files (11 Mountain & 36 Valley) can be imported to GRASS, using the command r.in.gdal (Figure 33), and loaded as raster files (Figure 34).

Figure 33. Import data – window
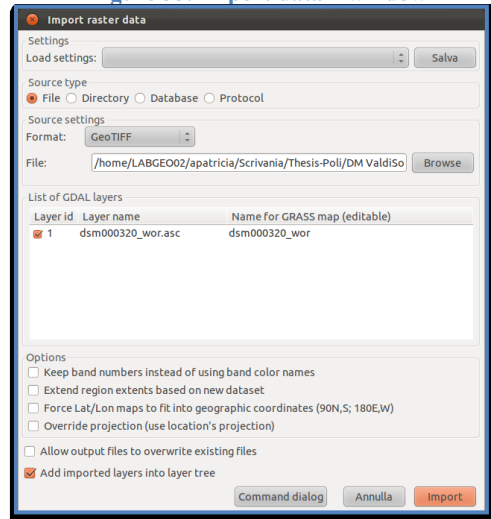
Remark: To perform **any** operation in GRASS is required to select a temporary sub-region, on the contrary the results won't be successful. Thus a new region is set as follows:

r. region rast= dsm000320@PERMANENT

The command r.region rast takes the area of this raster and the operations are done only inside the region.

Then the command r.mapcalc is used to calculate the difference between each DSM file and the corresponding DTM file (Figure 36 – Figure 37)

Example of the differential operation:

r.mapcalc expression= "diff000320 = dsm000320@PERMANENT – dtm000320@PERMANENT"

This command works with some *inputs* as is the case of dsm000320@PERMANENT and dtm000320@PERMANENT and an *output* which is the algebraic difference between the inputs; in the example it corresponds to diff000320@PERMANENT.

After these operations, 47 separated rasters containing the differences are produced. As an additional procedure, the r.patch command is applied to join the previous files allowing to visualize a general behavior of the regions.

Example of patch command

r.patch input=diff000178@PERMANENT, diff000179@PERMANENT, diff000213@PERMANENT, diff000214@PERMANENT, diff000215@PERMANENT, diff000249@PERMANENT output=diff1

The previous command joins the inputs files, i.e. the differences that have been obtained with r.mapcal, in an output file, for instance diff1@PERMANENT.

Figure 34. Raster files – Mountain Area
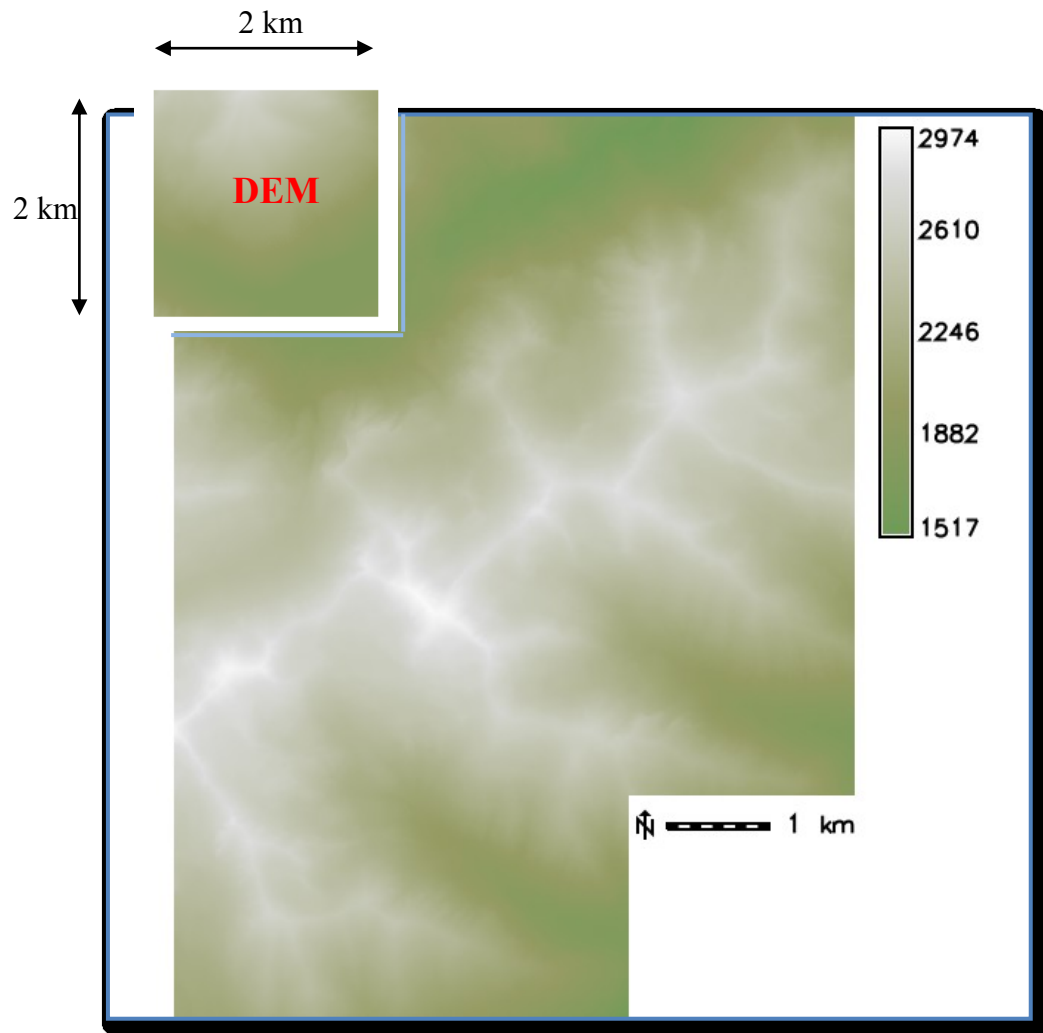
RASTER file (DEM) characteristcs

Size (m): 2020 x 2020

Resolution: 1 – 2 metros

Figure 35. RASTER files – Valley Area

RASTER file (DEM)
characteristcs

Size: 2020 x 2020 m
Resolution: 1 – 2 m
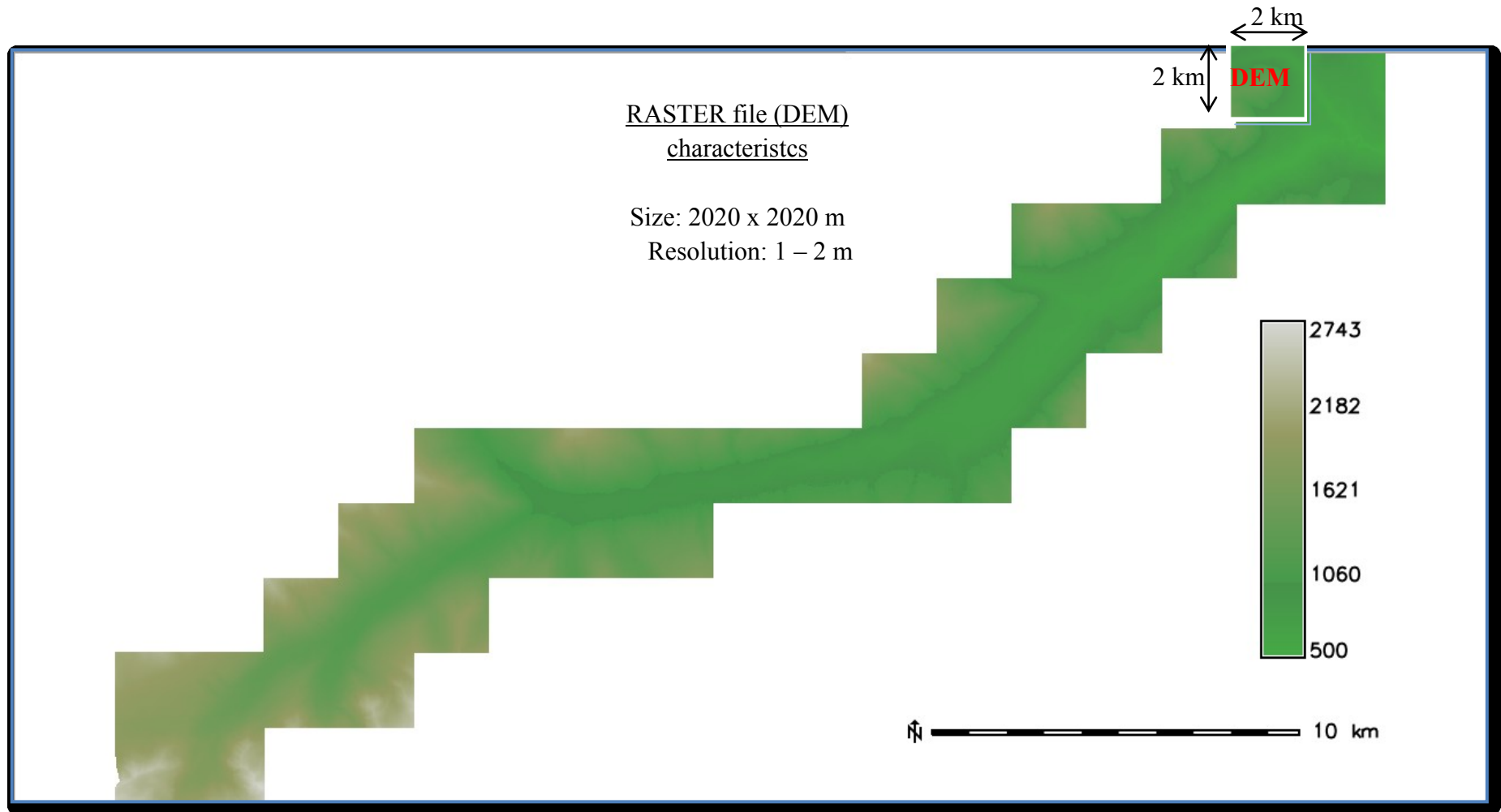
Figure 36. Differences in mountain area
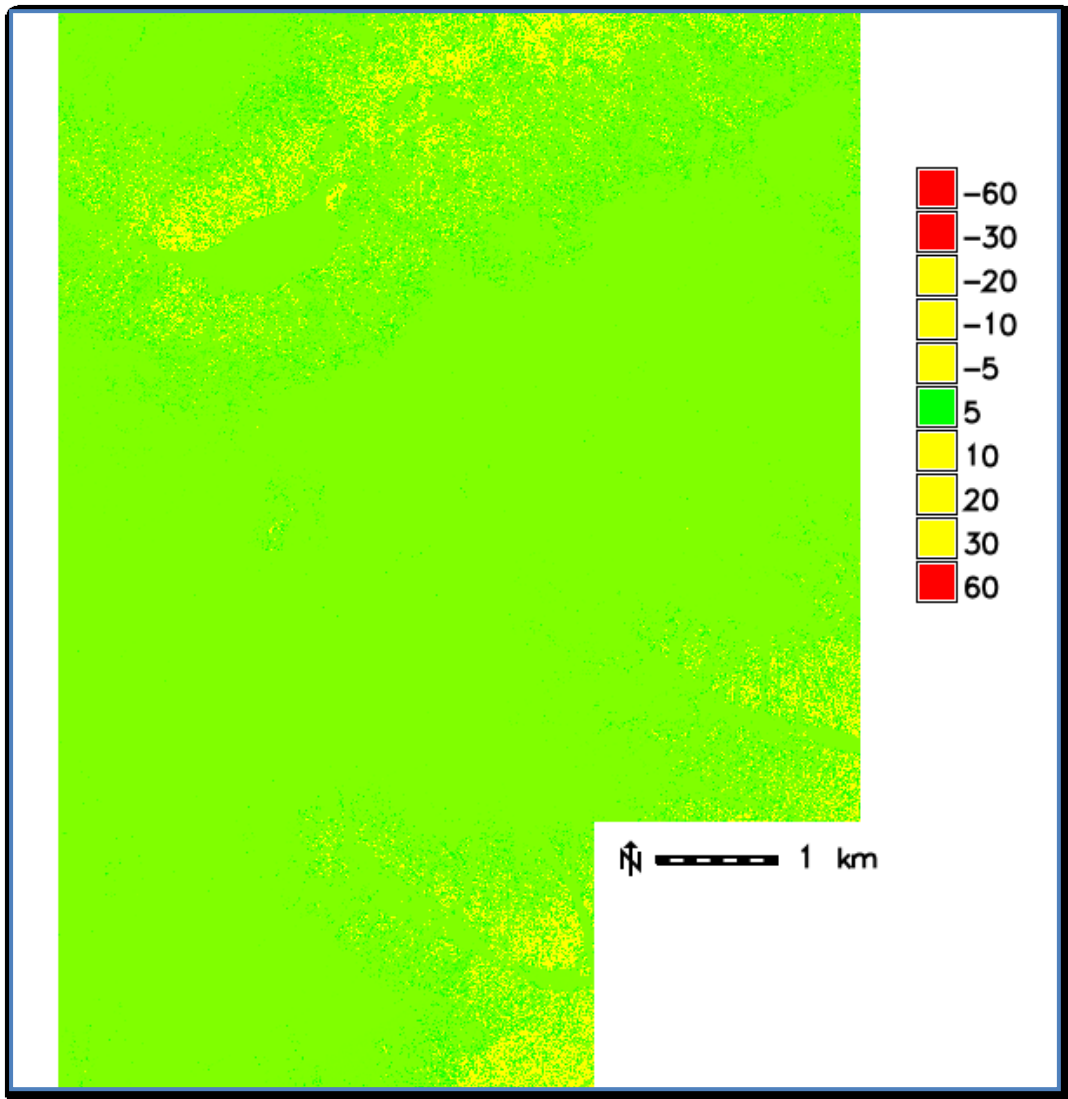
Figure 37. Differences in valley area
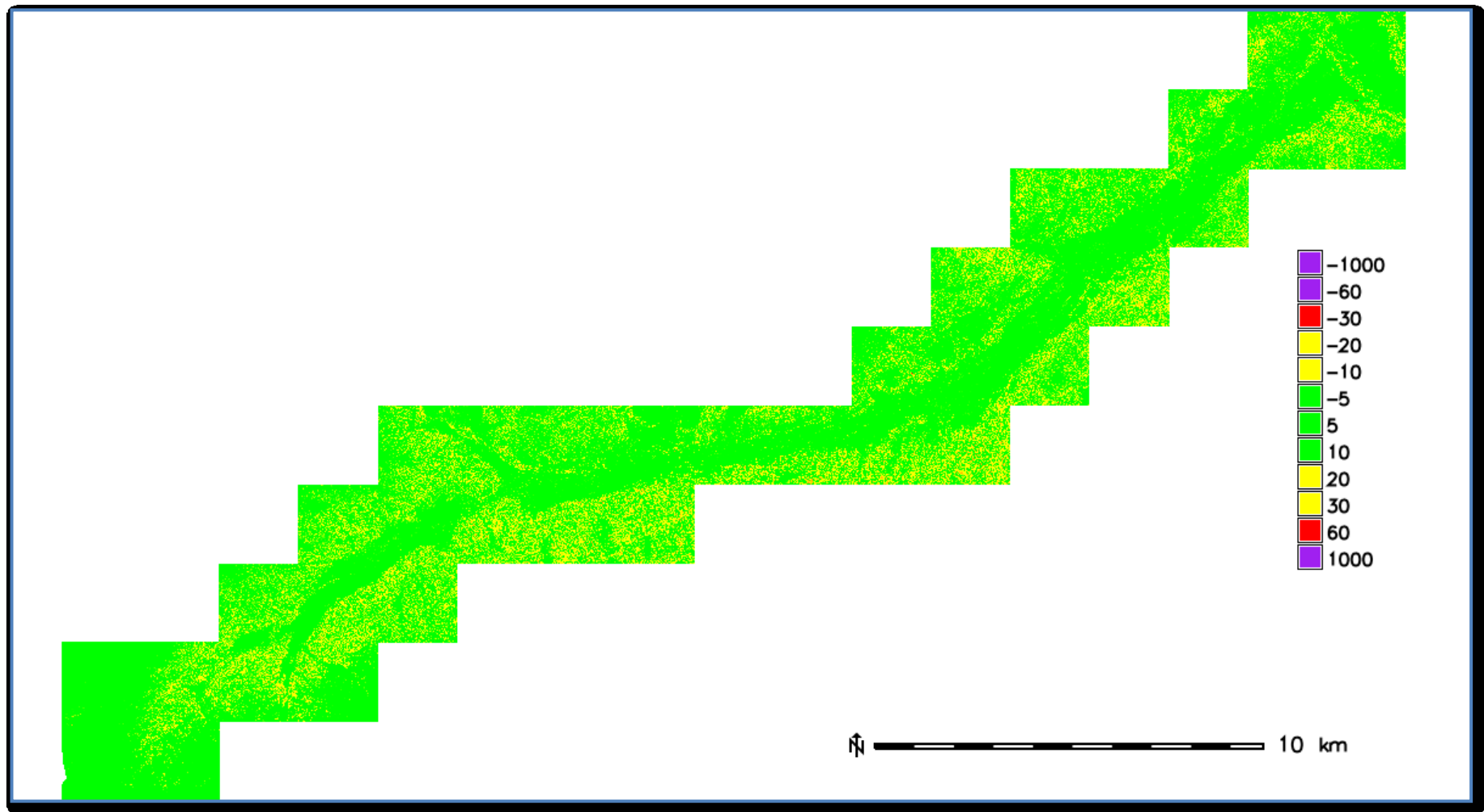
6.3. Analysis of differences

This new products, i.e. maps, one for Mountain and the other for Valley (Figure 36 – Figure 37), can give a specific idea of the probable components that could be present in the area as is mentioned in previous lines.

Nevertheless, before characterize the zone it is important to verify the outcomes in order to work with accurate data and without outliers; to do that some statistical tools will be calculated such as, minimum, maximum, mean, standard deviation (Table 4 and Table 5). To obtain these statistics the GRASS command r.univar is applied.

Example of statistics calculus

r.univar map=Valdisole_m@PERMANENT

Table 4. Differences Mountain Statistics

Differences Mountain Statistics – General	
No. pixels	44280227
Minimum	-23 m
Maximum	56 m
Range	78 m
Mean	1 m
Standard Deviation	4 m

Table 5. Differences Valley Statistics

Differences Valley Statistics – General	
No. pixels	145021187
Minimum	-39 m
Maximum	974 m
Range	1013 m
Mean	6 m
Standard Deviation	9 m

According with the Mountain results (Table 4), it is noticeable the presence of a negative difference -22.65 m, which shows likely a rare behavior on one of the models either DSM or DTM. The meaning of this minus (-) is that, in some random places, the surface seems to be below the terrain, which does not have a physical meaning thus this phenomenon will be considered in the following.

Besides this irregularity, another remark can be done with respect to the mean which gives a rough characterization of the area where the presence or the absence of small trees can be figured out from its value namely, 1.10m. Moreover, as the study area "Mountain" has a height bigger than 1500 meters above sea level it does not able to grow too much vegetation and therefore most of the surface is covered just by grass and bare rock.

In order to complement this evaluation, is also appreciable a maximum value equal to 56 m which generates certain doubts considering either the absence of trees and buildings or in the opposite case an atypical element belonged to this altitude. As well as the negative differences, this anomaly will be discussed.

By analyzing the Valley results (Table 5) can be appreciated also negative differences which are bigger than the Mountain case, probably subject to the plain area. On the order hand, the positive differences have a maximum value of 973.79 m that is clearly one anomalous value. These two behaviors will be analyzed in the following pages.

Another important aspect connected with Valley characterization is the mean value. It can be considered a key feature due the nature of its meaning which denotes the area cover with different elements such as vegetation, buildings, bridges, factories, roads characterized by their own heights.

6.3.1. Negative differences

The results obtained by Mountain's map differences and Valley's map differences have produced unexpected negative values, taking into account that the approach was analyzed just positives differences though in the presence of these anomalies a work has to be done, thus this section is developed. The mentioned negatives results are going to be studied in order to find out the cause or causes to the problem.

The first approximation develop a separation process of the outcomes, making use of r.mapcalc GRASS command, either for Mountain or Valley, this filter can deal only with the interested features as follows,

r.mapcalc expression "Negat_v = if (Valdisole_v@PERMANENT < 0, Valdisole_v@PERMENT, null())"

This syntax just store in an output raster, for example Negat_v@PERMANENT, the difference values, for this case Valdisole_v@PERMANENT, that are smaller than zero and the values that do not fulfill the condition are filled with a null value.

The statistics of the negative pixels are given in Table 6 and Table 7 applying the command r.univar.

Example

r.univar map=Negat_m@PERMANENT

Table 6. Mountain Negative Data

Mountain Statistics – Negatives	
N. pixels	86301
Minimum	-23 m
Maximum	0 m
Range	23 m
Mean	-0.1 m
Standard Deviation	0.4 m

Table 7. Valley Negative Data

Valley Statistics – Negatives	
N. pixels	968041
Minimum	-39 m
Maximum	0 m
Range	39 m
Mean	-0.1 m
Variance	0.2 m

The results are stated below

Case a. *Mountain*

Negative pixels/ Total pixels (‰) = 86301 / 44280227 * 1000 = 0.001949*1000 = 1.94‰

This value, does not seem significant, in any case, it is necessary to deeply analyze the data to ensure the data confidence and remove the possible outliers.

But first to have an overview of the studied phenomenon a histogram is presented in the Figure 38.

Figure 38. Histogram of negative values in Mountain area
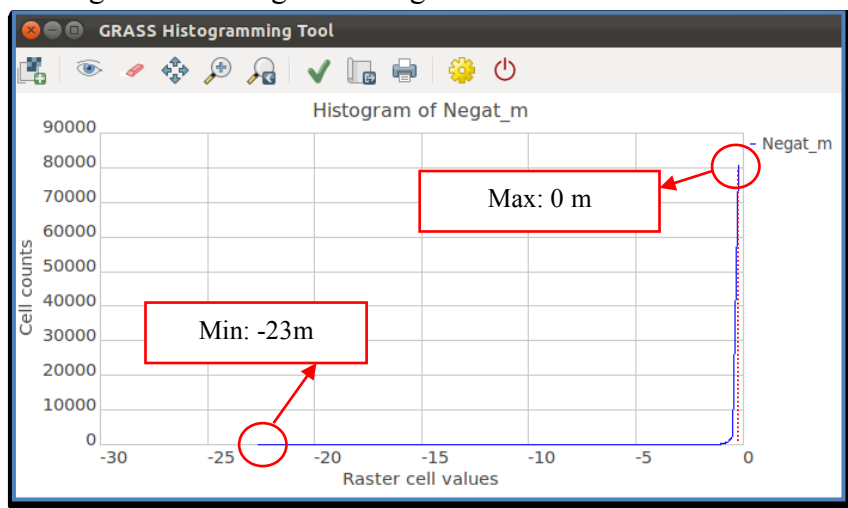
The histogram is hardly readable. With the purpose of describing better the histogram three more filter patterns have been applied, namely, pixels with a difference value smaller than -0.25 m, -0.50 m and -1.00 m which allow to highlight the atypical points comparing them with the total negatives pixels (% with respect to negative) and also with the total number of pixel (% with respect to total), respectively.

Before present the results of the filters is important clarified that the criterion for selecting those three filter values (0.25 m, 0.50 m and 1.00 m) is connected on the LIDAR accuracy (Table 1).

1. -0.25 m < # Pixels difference <= 0 m= 81346

- Relative value = 81346/86301 * 100 = 94.3%
- Absolute value = 0.942584675 * 0.001949 * 100 = 0.2%

2. -0.50 m < # Pixels difference <= -0.25 m = 2289

- Relative value = 2289/86301 * 100 = 2.7%
- Absolute value = 0.026523447* 0.001949 * 100 = 0.01%

3. -1.00 m < # Pixels difference <= -0.50 m = 1400

- Relative value = 1400/86301 * 100 = 1.6%
- Absolute value = 0.016222292* 0.001949 * 100 = 0.003%

4. # Pixels difference <= -1.00 m = 1266

- Relative value = 1266/86301 * 100 = 1.5%
- Absolute value = 0.014669587* 0.001949 * 100 = 0.003%

These outcomes, especially the ones which are compared with the overall number of pixels, show that the amount of significantly irregular data has a low incidence in the studied area: 3 out of 100.000 pixels can be possible errors just caused by the method that has been used to smooth the surface. A discussion of negative outliers is given in the section 6.4.

The same procedure is applied to the larger zone, to be precise the Valley area

 Case b. *Valley*

Negative pixels/ Total pixels (‰) = 968041 / 145021187 *1000 = 0.006675*1000 = 6.67‰

In this case the negative values are more. In this case, c.a. 7 out of 1000 pixels evidence this phenomenon. Moreover, to understand the previous event a histogram is shown in the Figure 39.

Figure 39. Histogram of negative values in Valley area
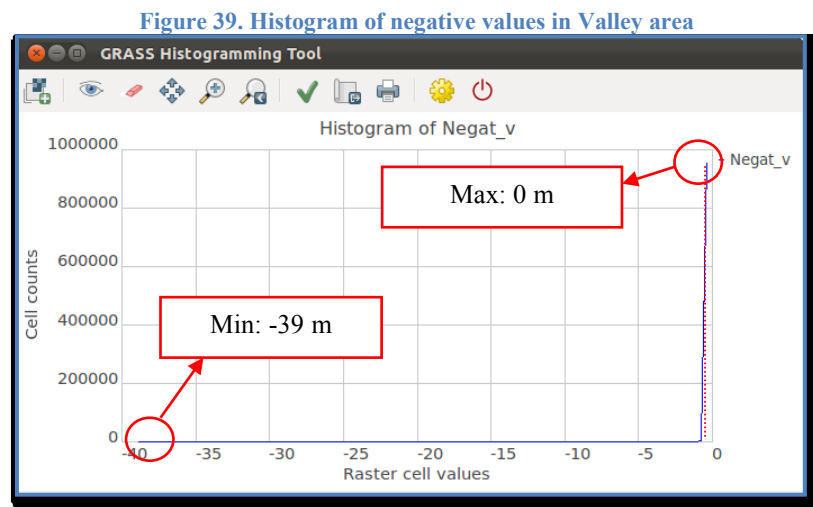
Likewise this histogram is hardly readable as the Mountain case. Also three filters have been applied, namely, pixels with a difference value smaller than -0.25 m, -0.50 m and -1.00 m which allow to highlight the atypical points comparing them with the total negatives pixels (% with respect to negatives) and also with the total number of pixel (% with respect to total), respectively.

1. -0.25 m < # Pixels difference <= 0 m= 950484

- Relative value = 950484/968041 * 100 = 98.2%
- Absolute value = 0.981863371 * 0.006675 * 100 = 0.7%

2. -0.50 m < # Pixels difference <= -0.25 m = 9650

- Relative value = 9650/968041 * 100 = 1.0%
- Absolute value = 0.009968586 * 0.006675 * 100 = 0.01%

3. -1.00 m < # Pixels difference <= -0.50 m = 4282

- Relative value = 4282/968041 * 100 = 0.4%
- Absolute value = 0.004423366 * 0.006675 * 100 = 0.003%

4. # Pixels difference <= -1.00 m = 3625

- Relative value = 3625/968041 * 100 = 0.4%
  Absolute value = 0.003744676 * 0.006675 * 100 = 0.002%

54

From the previous process can be concluded that the amount of significantly irregular data have a low incidence in the studied area, for instance 2 out of 100.000 pixels can be possible errors just caused by the method that has been used to smooth the surface.

In the valley area more negatives values, in percentage, are present. This regards the values comprised between - 0.25 m and 0 m, which are small errors due to filtering of buildings.

A discussion of negative outliers is given in section 6.4.

6.3.2. Positive differences

These differences are expected because represent the elements or objects that are between the terrain (DTM) and the surface (DSM).

To analyze these data, as in the negative differences, a separation process is developed on the outcomes either for Mountain or Valley: this filter deals only with the interested features as follows,

r.mapcalc expression = "Posit_m = if (Valdisole_m@PERMANENT >= 0, Valdisole_m@PERMANENT, null())"

This syntax just store in an output raster, for example Posit_m@PERMANENT, the difference values, for this case Valdisole_m@PERMANENT, that are greater than zero and the values that do not fulfill the condition are filled with a null value.

In order to visualized the behavior of these filtered data is applied the r.univar command which, as is mentioned before, gives some basic statistics.

Example

r.univar map=Posit_v@PERMANENT

Table 8. Mountain Positive Data

Mountain Statistics – Positives	
N. pixels	44193926
Minimum	0 m
Maximum	56 m
Range	56 m
Mean	1 m
Standard Deviation	4 m

Table 9. Valley Positive Data

Valley Statistics – Positives	
N. pixels	144053146
Minimum	0 m
Maximum	974 m
Range	974 m
Mean	7 m
Standard Deviation	9 m

The next part presents the analysis of each region and the respective conclusions.

Case a. *Mountain*

Positive pixels/ Total pixels (%) = 44193926/44280227 * 100 = 0.998051*100 = 99.8%

Once is obtained this value, which ensures that most of the data (99.8%) behave as was expected meaning a DSM location above DTM, is necessary to reprocess the information because in Table 8 appears a maximum difference value equals to 56 m that produces some doubts, inasmuch as in this mountain area, the vegetation, that is the only obstacle for the acquisition, cannot reach more than 40 m which is an average height of the trees present in this zone; in chapter 5 are listed and explained different types of vegetation. Even though, isolated values up to 60 or 70 meters are possible when speak about structures built by humans, such as high tension pylons.

But first to have an overview of the studied phenomenon a histogram is presented in the Figure 40.
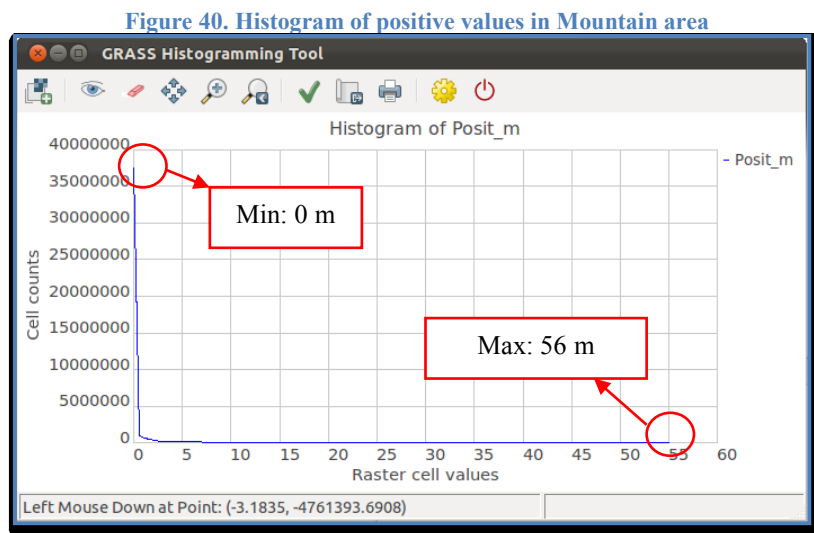
Figure 40. Histogram of positive values in Mountain area

With the purpose of describing better this histogram and also to verify if those values that are greater than 40 meters correspond or not to outliers, seven filters have been applied, namely, pixels with a difference value greater than 0 m, 1m, 5m, 10m 20m, 40m and 60m which

56

allow to highlight the atypical points. This is executed comparing the filtered data with the total positive pixels (% with respect to positives) and also with the total number of pixel (% with respect to total), respectively.

1. 0 m <= # Pixels difference < 1 m= 38320007

- Relative value = 38320007/44193926 * 100 = 86.7%
- Absolute value = 0.86708764* 0.998051026 * 100 = 86.5%

2. 1 m <= # Pixels difference < 5 m = 2610340

- Relative value = 2610340/44193926 * 100 = 5.9%
- Absolute value = 0.059065583 * 0.998051026 * 100 = 5.9%

3. 5 m <= # Pixels difference < 10 m = 1324254

- Relative value = 1324254/44193926 * 100 = 3.0%
- Absolute value = 0.029964615 * 0.998051026 * 100 = 3.0%

4. 10 m <= # Pixels difference < 20 m = 1532754

- Relative value = 1532754/44193926 * 100 = 3.5%
- Absolute value = 0.034682458 * 0.998051026 * 100 = 3.5%

5. 20 m <= # Pixels difference < 40 m = 406450

- Relative value = 406450/44193926 * 100 = 0.9%
- Absolute value = 0.009196965 * 0.998051026 * 100 = 0.9%

6. 40 m <= # Pixels difference < 60 m = 121

- Relative value = 121/44193926 * 100 = 0%
- Absolute value = 0 * 0.998051026 * 100 = 0 %

7. # Pixels difference >= 60 m = 0

- Relative value = 0/44193926 * 100 = 0%
- Absolute value = 0 * 0.998051026 * 100 = 0%

The last two filters show that the amount of irregular data has a low incidence in the studied area. A discussion of positive outliers is given in section 6.4.

The same procedure is applied to the larger zone, to be precise the Valley area.

Case b. *Valley*

\# Positive pixels/ Total pixels (%) = 144053146/145021187 * 100 = 0.993325*100 = 99.3%

Once is obtained this value, is necessary to reprocess the information because in Table 9 appears a maximum difference value equals to 974 m that produces some doubts about the data obtained by LIDAR method, inasmuch as in this valley area, the objects (trees, buildings, etc) that are majority of the obstacles for the acquisition, cannot reach more than 65 m else there are other structures such as roads or bridges that can reach heights, with respect to the terrain, over 65 m but typically no more than 100m. Thus it is needed to verify which the case is.

In valley area bigger differences can be expected because the presence of structures as bridges (roads and railways), chimneys and high tension pylons.

But first, to have an overview of the studied phenomenon a histogram is presented in the Figure 41.
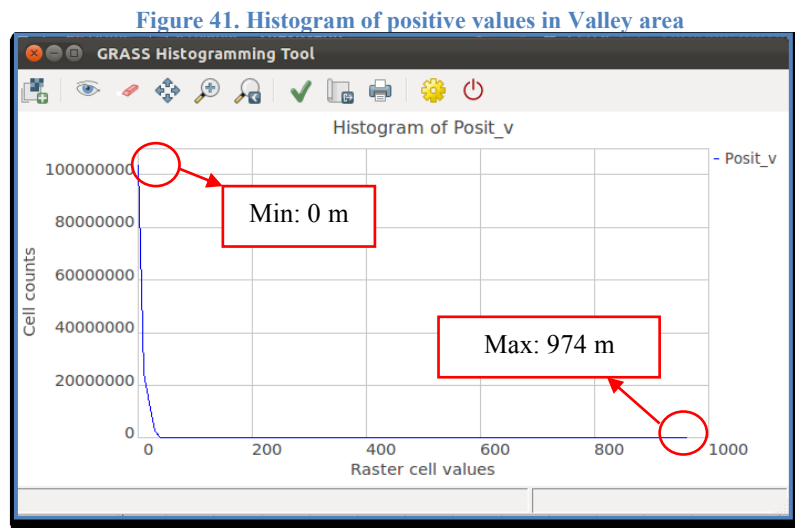
Figure 41. Histogram of positive values in Valley area

With the purpose of describing better this histogram and also to verify if those values that are greater than 65m or 100 meters correspond or not to outliers, five filters have been applied, namely, pixels with a difference value greater than 0m, 5m, 20m, 60m and 100 m which allow to highlight the atypical points. This is executed comparing the filtered data with the total positive pixels (% with respect to positives) and also with the total number of pixel (% with respect to total), respectively.

1. 0 m <= \# Pixels difference < 5 m= 87243823

- Relative value = 87243823/144053146 * 100 = 60.6%
- Absolute value = 0.60563636 * 0.993324831* 100 = 60.2%

58

2. 5 m <= # Pixels difference < 20 m = 41818926

- Relative value = 41818926/144053146 * 100 = 29.0%
- Absolute value = 0.290302067 * 0.993324831* 100 = 28.8%

3. 20 m <= # Pixels difference < 60 m = 14989895

- Relative value = 14989895/144053146 * 100 = 10.4%
- Absolute value = 0.104058088 * 0.993324831* 100 = 10.3%

4. 60 m <= # Pixels difference < 100 m = 495

- Relative value = 495/144053146 * 100 = 0%
- Absolute value = 0 * 0.993324831* 100 = 0%

5. # Pixels difference >= 100 m = 7

- Relative value = 7/144053146 * 100 = 0%
- Absolute value = 0 * 0.993324831* 100 = 0%

The last two filters show that the amount of irregular data has a low incidence in the studied area. A discussion of positive outliers is given in section 6.4.

6.4. Categorization and verification

Nevertheless, to be sure that the hypotheses of outliers presence are true, an additional step is carried out where these doubtful points are placed in an ortophoto, so that can be recognized and moreover that the resulting objects (vegetation, buildings, roads, bridges, crops, etc) match with an official clasification map (Real land use map - Trentino Region), in such a way that these data also remain categorized.

6.4.1. Official clasification map

In order to produce the well mentioned 'categorization' a base map is needed that contains the Real Land Use ("USO DEL SUOLO REALE URBANISTICA (ED. 08/2003)"). This has been downloaded for free from GeoCartographic Trentino Province Web Site ("Portale Geocartografico Trentino) in a vector format.

Download link:

   http://www.territorio.provincia.tn.it/portal/server.pt/community/sgc_-_geocatalogo/

The vector format of the ("USO DEL SUOLO REALE URBANISTICA (ED. 08/2003)" presents these files:

- usr_urb.dbf
- usr_urb.prj
- usr_urb.shp
- usr_urb.shx

6.4.2. Methodology

Here the methodology that is used to obtain the aforementioned map will be presented.

Firstly, it is necessary to convert the vector map in a raster map because all the data are in this format. The land use and the differences maps must be in the same format to be manipulated and matched. This operation is called Rasterization, the command line is:

'v.to.rastinput=usr_urb@PERMANENT output=usr_urb_tot use=attr attrcolumn=SYM_SHD labelcolumn=UR_NO'

To work with this command an input file, certainly a vector file, is required, for this case the usr_urb@PERMANENT represents the vector file to be transformed. Then an output name can be chosen. Finally, the "use" parameter reads values from attribute table to perform the raterization through a specific column of the attribute table, for instance SYM_SHD (numberic representation of the land use) and also adding a label which is the name of column used as raster category labels.

Once the data are in raster format the categorization process begins. To facilitate this procedure a Script named 'category_caltulation' is built up then it must run in the GRASS command window in this way:

GRASS 7.0. svn (DMTRENTO):~/Scrivania/Thesis-Poli/Script > ./category_calculation

Then the following code handles for group objects through 5 simple steps

<div align="center">Script 'category_caltulation'</div>

```
#!/bin/bash

g.region
rast=11_DSM@PERMANENT,22_DSM@PERMANENT,1_DSM@PERMANENT,2_DSM@PERMANENT,3_DSM@PE
RMANENT,4_DSM@PERMANENT
```
1

```
for i in {1..88}

do
```

```
r.mapcalc expression="Cat_${i}_v = if( usr_urb_tot@PERMANENT==$i , Valdisole_v@PERMANENT , null()  )"

r.mapcalc expression="Cat_${i}_m = if( usr_urb_tot@PERMANENT==$i , Valdisole_m@PERMANENT , null())"
```
2

```
r.patch input=Cat_${i}_v@PERMANENT,Cat_${i}_m@PERMANENT output=Cat_$i
```
3

```
r.mapcalc expression="Cat_${i}_n = if( Cat_${i}@PERMANENT<0  , Cat_${i}@PERMANENT  , null()  )"

r.mapcalc expression="Cat_${i}_p = if( Cat_${i}@PERMANENT>=0  , Cat_${i}@PERMANENT  , null()  )"
```
4

```
r.univar map=Cat_${i}_n@PERMANENT > Cat_${i}_n.txt

r.univar map=Cat_${i}_p@PERMANENT > Cat_${i}_p.txt
```
5

done

1 Defined the geographic area in which GRASS will work.

2 Depending on the 88 categories that have been extracted from the vector map, the respective zonal map (Mountain/Valley) is generated. This syntax just store in an output raster, for example Cat_61_v@PERMANENT, the difference values, in this case Valdisole_v@PERMANENT, that belongs to land use category (usr_urb_tot@PERMANENT) that is been considered and the values that do not fulfill the condition are filled with a null value.

3 Both the mountain region and the valley are merged to form a single categorical raster map Cat_61@PERMANENT. The command just merges these input files in an output file.

4 Each category is filtered into negative and positive values, as done in the previous sub-charter 6.3. This syntax just store in an output raster, for example Cat_61_n@PERMANENT, the difference values, for this case Cat_61@PERMANENT, that are smaller than zero and the values that do not fulfill the condition are filled with a null value.

5 Statistical tools have been used to analyze the different behavior of each category and create a single .txt file to analyze each category.

The results are shown in two tables, the first one which represents the positive pixels and the second the negative pixels both along the 32 categories that in fact belongs to this selected area 'Val di Sole'.

Table 10. Positive values of categories

Cathegory	Minimum	Maximum	Mean	Standard Dev.	Description
3	0	53,5	2,9	4,3	Discontinuos urban fabric/Tessuto urbano discontinuo
4	0	31,9	2,2	3,3	Individual houses/Case singole
7	0	34,8	2,1	3,6	Industrial and Craft production areas/Aree produttive industriali ed artigianali
8	0	30,0	3,5	4,2	Commercial areas/Aree commerciali
10	0	973,8	1,3	4,1	Road networks/Reti stradali
11	0	91,5	0,5	3,5	Rail networks/Reti ferroviarie
19	0	8,6	0,1	0,3	Surface parking/Parcheggi di superficie
22	0	21,7	1,3	2,7	Areas of railway station/Aree di stazione ferroviaria
26	0	14,9	2,0	3,2	Stations/services for cableway/Stazioni/servizi per impianto a fune
28	0	21,0	2,5	2,1	Waste water treatment plant/Impianti di depurazione
30	0	17,7	1,0	2,5	Hydro electric power plants/Centrali idroelettriche
33	0	13,5	3,8	4,2	Hospital complexes/Complessi ospedalieri
35	0	26,6	1,0	2,6	Cemetery complexes/Complessi cimiteriali
41	0	29,1	0,5	1,8	Stone quarries/Cave di pietra
43	0	39,2	1,5	3,5	Unclassificable building site and artificial covering areas/ Cantieri e aree a copertura artificiale non classificabile
45	0	12,4	1,3	2,5	Urban green areas/Aree verdi urbane
46	0	24,0	0,6	2,1	Areas for sporting and recreational activities/Aree per attivita' sportiva e ricreativa
47	0	29,0	2,6	4,4	Camping areas/resort/Aree per campeggio/villaggio turistico
52	0	58,4	0,5	1,7	Orchards and small fruit/Frutteti e frutti minori
55	0	56,9	0,5	2,1	Meadows/Prati stabili
56	0	10,9	0,7	1,1	Heterogeneous agricultural crops/Colture agricole eterogenee
60	0	802,7	5,1	6,6	Hardwood forests/Boschi di latifoglie
61	0	117,2	8,7	9,1	Coniferous forests/Boschi di conifere
62	0	87,2	11,0	9,2	Mixed forests/Boschi misti
64	0	36,3	0,1	1,0	Areas of natural pasture and mountain meadows/Aree a pascolo naturale e praterie di alta quota
66	0	51,3	2,0	4,0	Shrubs and lilies of the valley/Arbusteti e mugheti
68	0	41,5	1,2	3,0	Grass and trees/Prato alberato
72	0	44,9	0,2	1,0	Bare rocks/Rocce nude
78	0	33,9	0,2	1,2	Bogs/Torbiere
83	0	85,5	1,8	4,3	Natural water bodies/Corsi di acqua naturale
86	0	20,7	0,2	1,3	Natural lakes/Laghi naturali
87	0	42,8	0,4	2,1	Artificial lakes/Laghi artificiali

Table 11. Negative values of categories

Cathegory	Minimum	Maximum	Mean	Standard Dev.	Description
3	-3,1	0,00	-0,04	0,11	Discontinuos urban fabric/Tessuto urbano discontinuo
4	-0,7	0,00	-0,03	0,04	Individual houses/Case singole
7	-0,9	-0,01	-0,03	0,06	Industrial and Craft production areas/Aree produttive industriali ed artigianali
8	-0,5	-0,01	-0,03	0,04	Commercial areas/Aree commerciali
10	-1,9	0,00	-0,04	0,07	Road networks/Reti stradali
11	-1,9	-0,01	-0,04	0,07	Rail networks/Reti ferroviarie
19	-0,1	-0,01	-0,03	0,03	Surface parking/Parcheggi di superficie
22	-0,2	-0,01	-0,02	0,02	Areas of railway station/Aree di stazione ferroviaria
26	-0,5	-0,01	-0,04	0,08	Stations/services for cableway/Stazioni/servizi per impianto a fune
28	-0,1	-0,01	-0,02	0,02	Waste water treatment plant/Impianti di depurazione
30	-0,7	-0,01	-0,10	0,16	Hydro electric power plants/Centrali idroelettriche
33	-0,2	0,00	-0,03	0,03	Hospital complexes/Complessi ospedalieri
35	-0,4	-0,01	-0,03	0,04	Cemetery complexes/Complessi cimiteriali
41	-0,7	-0,01	-0,04	0,06	Stone quarries/Cave di pietra
43	-0,4	0,00	-0,03	0,04	Unclassificable building site and artificial covering areas/ Cantieri e aree a copertura artificiale non classificabile
45	-0,1	0,00	-0,01	0,01	Urban green areas/Aree verdi urbane
46	-0,6	0,00	-0,02	0,03	Areas for sporting and recreational activities/Aree per attivita' sportiva e ricreativa
47	-1,7	-0,01	-0,03	0,09	Camping areas/resort/Aree per campeggio/villaggio turistico
52	-1,2	-0,01	-0,03	0,04	Orchards and small fruit/Frutteti e frutti minori
55	-3,1	0,00	-0,03	0,04	Meadows/Prati stabili
56	-0,1	-0,01	-0,02	0,01	Heterogeneous agricultural crops/Colture agricole eterogenee
60	-10,3	0,00	-0,04	0,08	Hardwood forests/Boschi di latifoglie
61	-22,6	0,00	-0,05	0,25	Coniferous forests/Boschi di conifere
62	-38,9	0,00	-0,07	0,45	Mixed forests/Boschi misti
64	-14,6	0,00	-0,03	0,18	Areas of natural pasture and mountain meadows/Aree a pascolo naturale e praterie di alta quota
66	-6,4	0,00	-0,06	0,21	Shrubs and lilies of the valley/Arbusteti e mugheti
68	-0,7	-0,01	-0,03	0,04	Grass and trees/Prato alberato
72	-16,0	0,00	-0,16	0,51	Bare rocks/Rocce nude
78	-0,3	-0,01	-0,02	0,02	Bogs/Torbiere
83	-19,5	0,00	-0,04	0,27	Natural water bodies/Corsi di acqua naturale
86	-0,2	-0,01	-0,04	0,04	Natural lakes/Laghi naturali
87	-18,6	-0,01	-0,06	0,29	Artificial lakes/Laghi artificiali

6.4.3. Product 1 - Categorization based on Land Use

After dealing with verification process, the categorization map looks like that:

Figure 42. Categorization map
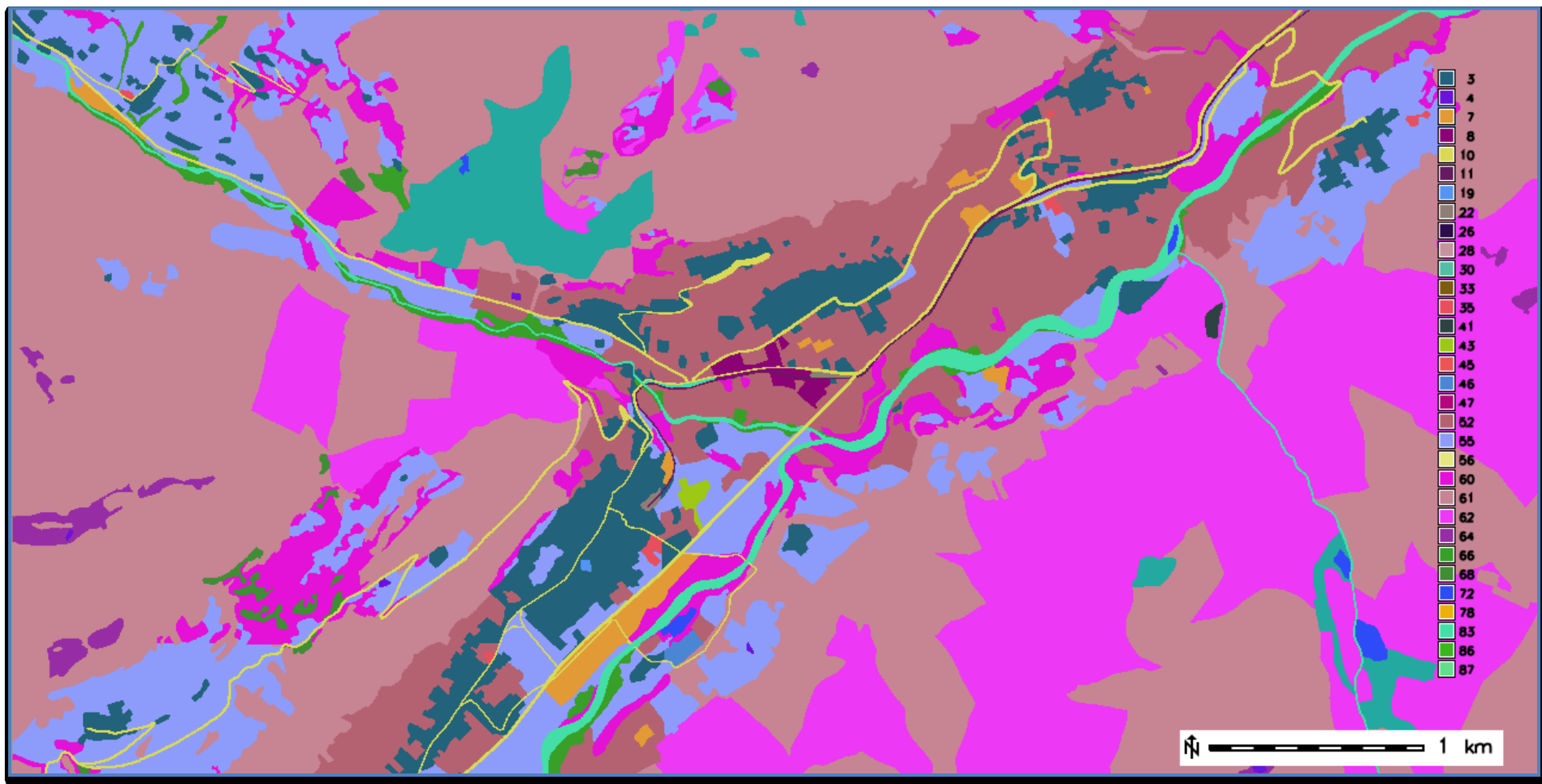
Figure 43. Extracted figure of Category 61 – Coniferous forests
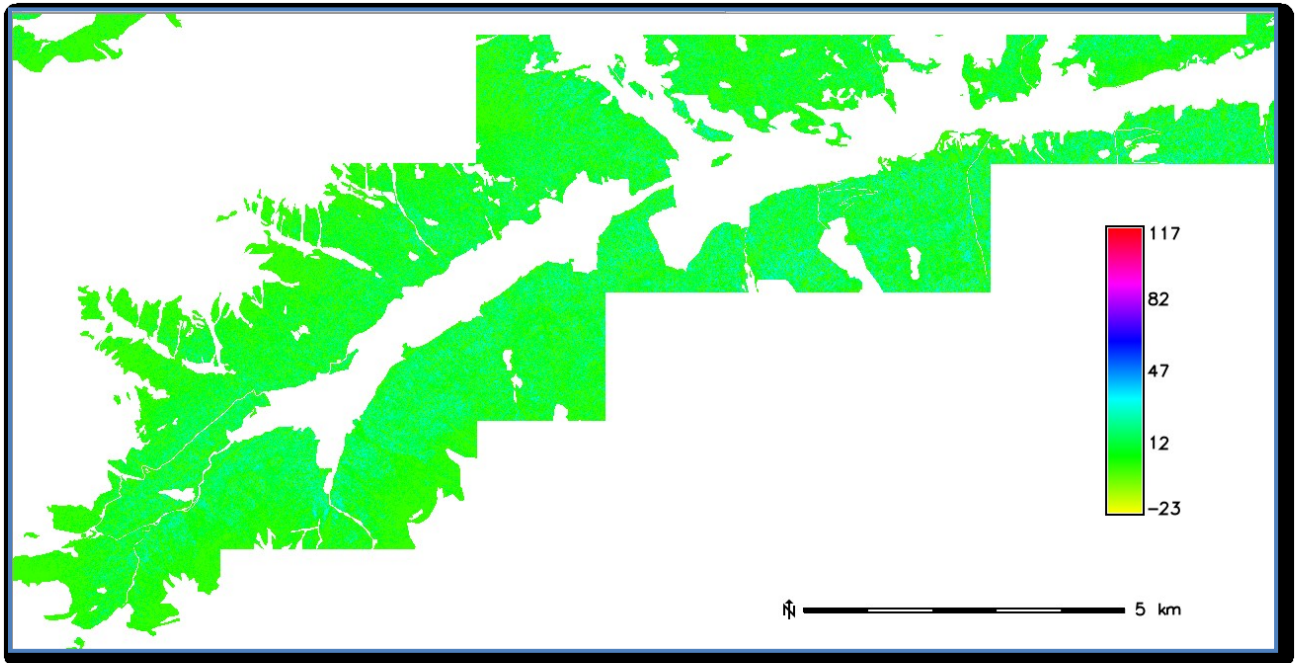
Figure 43 shows the Coniferous forest which is the main category along the Val di Sole.

6.4.4. Product 2 - Categorization based on difference values

Simply to present the results obtained by the positives differences and also to visualize the different features present in the study area, another classification is carried out:

Category 1. No difference between DSM and DTM, which means no vegetation or buildings.
Category 2. Small trees and buildings with a height between zero and 5 meters. (Urban/Not urban)
Category 3. Trees and buildings with a height between 5 and 10 meters. (Urban/Not urban)
Category 4. Big trees with height between 10 and 20 meters. (Not Urban area)
Category 5. Big trees with height between 20 and 30 meters. (Not Urban area)

Table 12. Categorization "Val di Sole"

Cat	Difference (m)		Color	
1	-1	1	Green	
2	1.1	5	Red	
3	5.1	10	Red	
4	10.1	20	Yellow	
5	20.1	30	Blue	

Figure 44. Differences map
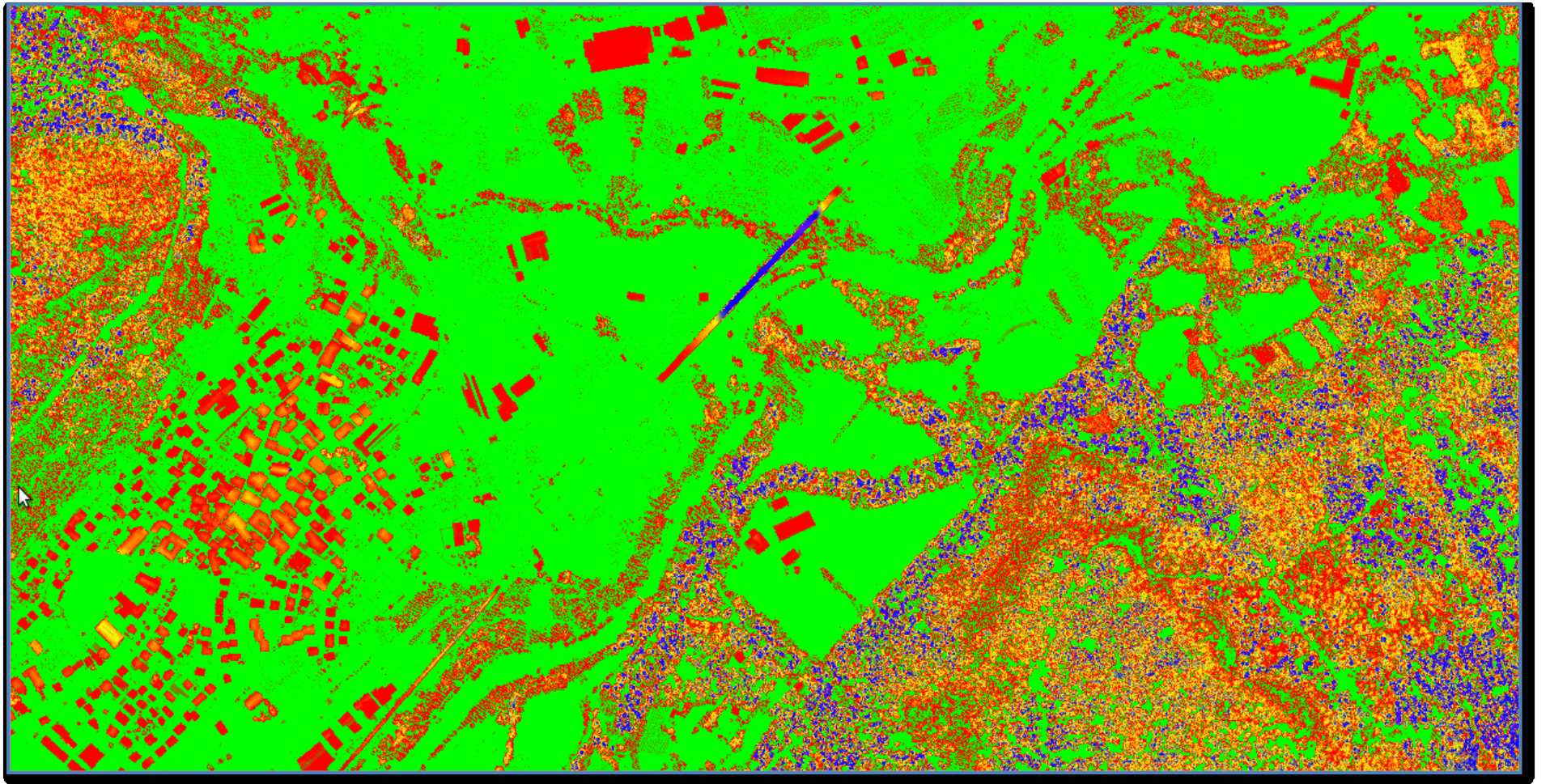
From this image is possible to identify an area than in its majority belongs to the first category. In the valley can be detected all the categories allowing better distinction of the features such as trees, buildings (factories, routes, brigdes).

6.4.5. Verification

The main idea is to check if the points of each category correspond in reality to what is present in the ground or to prove that the highlighted data of Table 10 and Table 11 are or not outliers. In fact, it is necessary the aforementioned ortophoto that is downloaded for free through the Geoportale Nazionale by a WMS.

URL:

http://www.pcn.minambiente.it/PCNDYN/catalogowms.jsp

When is obtained the URL, the next step is to load in a GIS platform, in this case ArcGIS, following this path

Add data → GIS Servers → Add WMS Server → Paste the URL link → Get layers → OK

Next step is to load the points that are required to be checked, in this point, some category data are export in a XYZ format. It can be done applying the GRASS command line below

r.out.xyz input=Cat61gr40@PERMANENT output= Cat61gr40 fs=_

This command requires the raster file that will be transformed in an array of data and the parameter "the symbol space (fs)" that separate data columns.

The outcome of this operation is shown in the next figure.

Figure 45. Exported format XYZ


```
Cat61_gr40.txt
 1  X     Y      DH
 2  655482,5   5142009,5   40,1699829102
 3  655696,5   5142004,5   40,25
 4  655695,5   5142003,5   42,8699951172
 5  655696,5   5142003,5   42,5100097656
 6  655697,5   5142003,5   40,0899658203
 7  655695,5   5142002,5   41,3900146484
 8  655696,5   5142002,5   41,1599731445
 9  655693,5   5142001,5   40,4299926758
10  655691,5   5142000,5   40,2100219727
11  655694,5   5142000,5   40,1699829102
12  655692,5   5141999,5   42,549987793
13  655690,5   5141995,5   40,5700073242
14  655697,5   5141989,5   40,0100097656
15  655469,5   5141802,5   40,8499755859
16  655468,5   5141801,5   41,1799926758
17  655469,5   5141801,5   40,9099731445
18  655467,5   5141800,5   41,8200073242
19  655468,5   5141800,5   43,7699584961
20  655470,5   5141800,5   42,4599609375
21  655468,5   5141799,5   41,6599731445
22  655526,5   5141791,5   42,0800170898
23  655527,5   5141791,5   40,7200317383
24  655526,5   5141790,5   41,549987793
25  655527,5   5141790,5   41,9799804688
26  655540,5   5141782,5   41,4899902344
27  655539,5   5141781,5   40,2799682617
28  655655,5   5141759,5   42,950012207
29  655652,5   5141758,5   45,1399536133
30  655653,5   5141758,5   45,6799926758
31  655655,5   5141758,5   47,8299560547
32  655653,5   5141757,5   46,5900268555
33  655654,5   5141757,5   46,0599975586
34  655655,5   5141757,5   43,8099975586
```

Finally, this set of data is upload in ArcGIS as follows:

File → Add Data → Add data XY Data → Look for the txt file (Cat61_gr40.txt)

The table is automatically load is the layer list and then to facilitate the verification process this table exported, through ArcGIS, into vector format.

Right click on the table (Cat61_gr40.txt) → Data → Export Data → Select the directory and the output name

With these two tools, the verification process can be done. To show how it works some examples will be exposed.

- Example 1
  This pixel belongs to the category 62 and it has a negative value of -38.9 meters.

Figure 46. Example 1 – Category 62
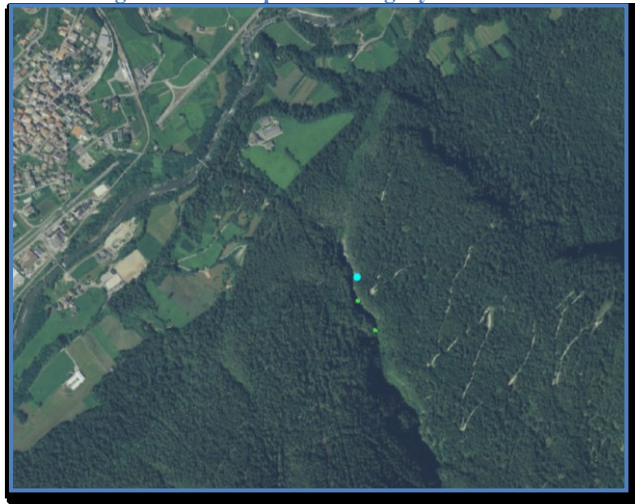
- Example 2
  This pixel belongs to the category 61 and it has a negative value of -22.64 meters.

Figure 47. Example 2 – Cat 61

The pixel belongs to a vegetation category as seen is the ortophoto. This negative difference occurs due to the smoothing methods, when the surface is no longer flat and becomes mountainous; there is not another reason that produces this behavior.

- Example 3

This pixel belongs to the category 10 and it has a positive value of 973.79 meters.

Figure 48. Example 3 – Category 10
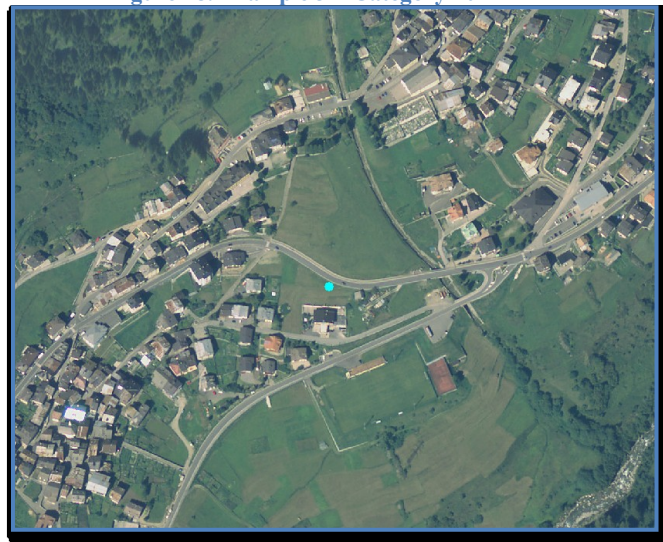
The first comment is that, indeed, the pixel almost belongs to a roadway infrastructure category as seen is the ortophoto. The last but not the least, this huge positive differences effectively occur due to an error in the acquisition or in the processing of the data, considering them as outliers.

- Example 4

These pixels belong to the categories 10 and 11 and they have positive values of 80.32 and 91.47 meters, respectively.

Figure 49. Example 4 – Categories 10 and 11
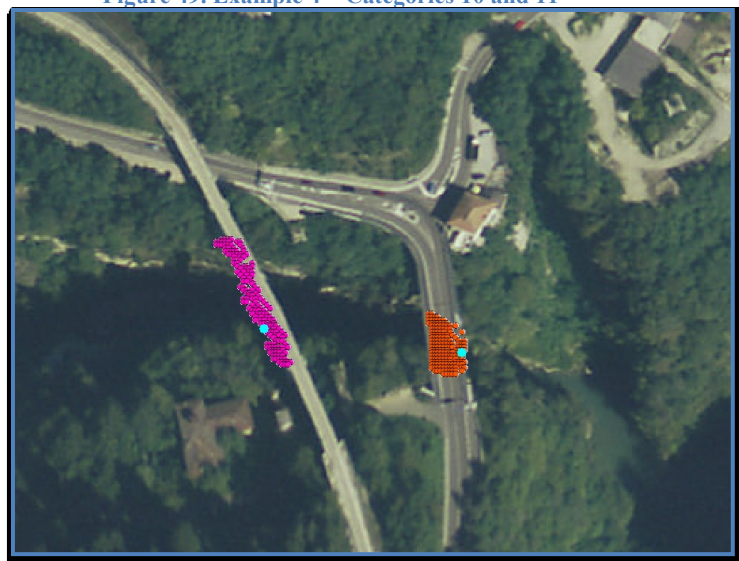
The last example shows that the pixels belong to a roadway infrastructure category as seen is the ortophoto. The presence of these two elements produce greater positive differences with respect to the rest of the pixels, excluding them from the outliers list.

Chapter 7

7. Download process of global DSMs

This chapter explains step by step the process of download SRTM, ASTER and GMTED2010 Digital Elevation Models from its respective websites and also the way to import them into the GRASS platform.

7.1 Data sources

SRTM and ASTER data

Another set of data needed is the global DEM produced by ASTER as well as SRTM. They can be downloaded for FREE from the link mentioned in Chapter 2. Here is explained step by step the procedure.

1. Access to the webpage and log in

    http://gdex.cr.usgs.gov/gdex/

Figure 50. Website – SRTM and ASTER
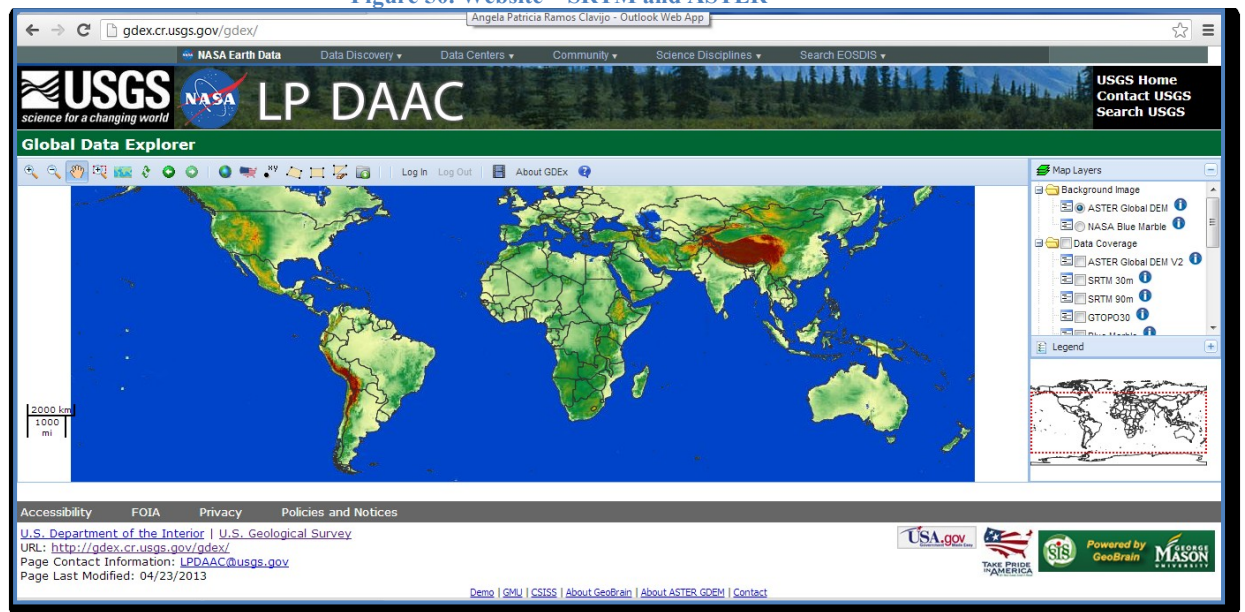
2. Define area by entering the boundaries in Lat/Lon. It is important to define very well these limits so that, the resolution that it is expected to obtain coincides with the one gives by the webGIS.

    The consideration on setting the correct boundaries must be done because, according with the download statements of this webGIS, the resolution of the downloaded

rasters is generated respecting exactly defined boundaries. Thus with a correct boundary set it will be obtain a square gridding, for these cases 1 x 1 arc-second (ASTER) and 3 x 3 arc-second(SRTM).

In this case, the bounds are:

North Latitude = 46.6
South Latitude = 45.6
East Longitude = 12.0
West Longitude = 10.4

3. Select the type of DEM either ASTER or SRTM that is need to download, also the type of format to be store (ArcASCII) and last but not least the project (Lat/Lon).

Figure 51. Download format
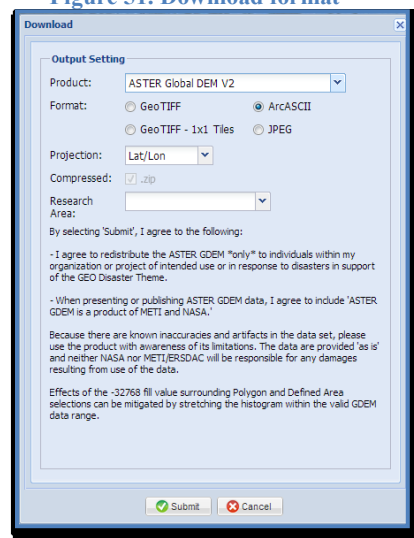
Note: ArcASCII format. The grid defines geographic space as an array of equally sized square grid points arranged in rows and columns. Each grid point stores a numeric value that represents a geographic attribute (such as elevation or surface slope) for that unit of space. Each grid cell is referenced by its x,y coordinate location. the first six lines indicate the reference of the grid, followed by the values listed in "English reading order" (left-right and top-down). For example, consider a grid, shown to the left. This could be encoded into an ASCII grid file, that would look like:

Figure 52. ArcASCII format

```
ncols        2109
nrows        2169
xllcorner    480025
yllcorner    4900025
cellsize     50
NODATA_value -9999
-9999 -9999 -9999 0 0 0 1 2 3 4 .....etc
```

4. Finally the DEMs are downloaded and present the following shape.

Figure 53. ASTER (Left) and SRTM (Right) outputs
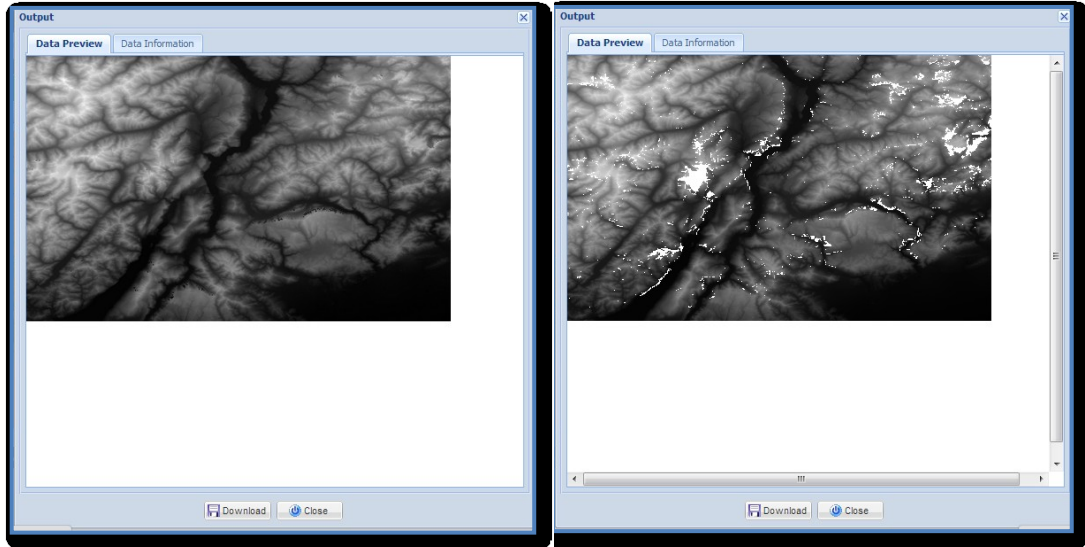
From this to DEMs it is possible to recognize the higher resolution and the amount of data of the ASTER model with respect to the SRTM model.

GMTED2010 data

The last DEM needed for the comparison is the GMTED2010 which will be downloaded for free almost in the same way as the previous part.

1. Access to the website

http://earthexplorer.usgs.gov/

Figure 54. Website - GMTED2010
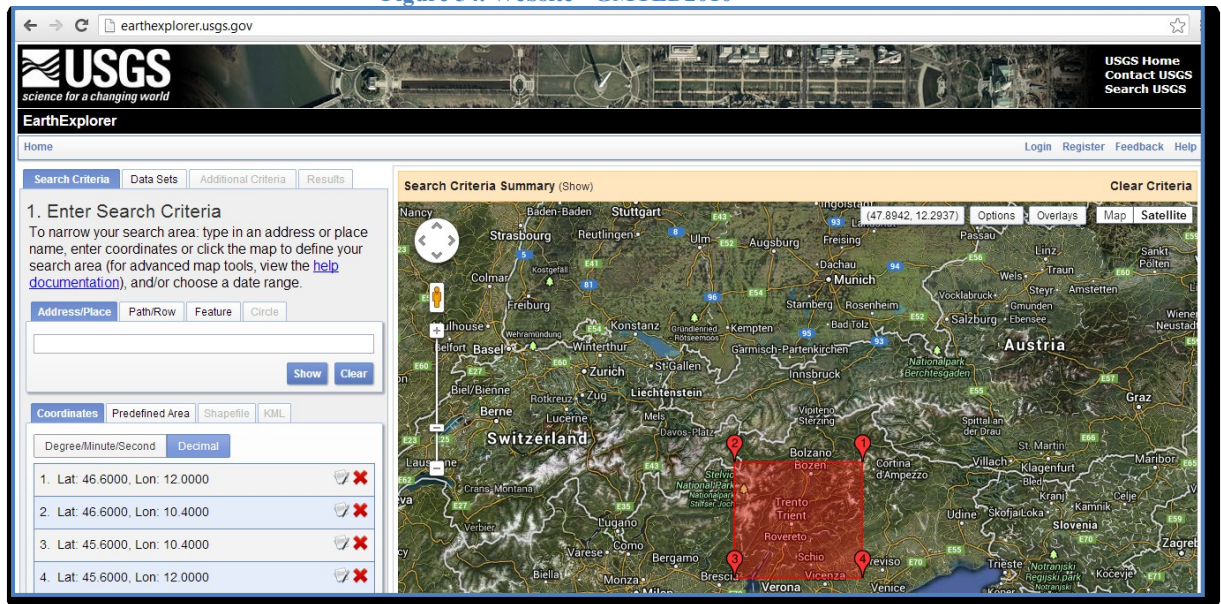
2. Define area by entering the boundaries in Lat/Lon which are the same as for ASTER and SRTM.

3. Select the type of DEM in this case GMTED2010

4. Select the resolution: As is mentioned in the Chapter 2, GMTED2010 has 3 different resolutions, for this study is used the higher resolution which means 7.5-seconds.

5. The downloaded DEM shows a region greater than the imposed, even though if another region is chosen the output raster will be the same. The format of this DEM is GeoTIFF.

Figure 55. GMTED2010 – Data set
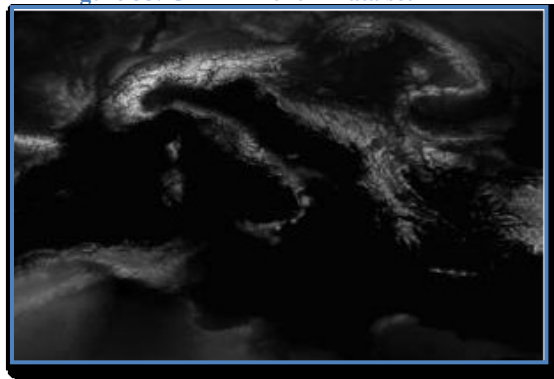
Note: GeoTIFF, refers to TIFF files which have geographic (or cartographic) data embedded as tags within the TIFF file. The geographic data can then be used to position the image in the correct location and geometry on the screen of a geographic information display. GeoTIFF is a metadata format, which provides geographic information to associate with the image data. But the TIFF file structure allows both the metadata and the image data to be encoded into the same file.

GeoTIFF makes use of a public tag structure which is platform interoperable between any and all GeoTIFF-savvy readers. Any GIS (GRASS GIS), CAD, Image Processing, Desktop Mapping and any other types of systems using geographic images can read any GeoTIFF files created on any system to the GeoTIFF specification.

7.2 Importing of DEMs

Once all the data are downloaded, they are imported in the GRASS platform. With respect to the previous chapters, the Location is different since the projection of the global DEMS are in Lat/Lon. Then the first operation is to create the new Location.

- Location (TrentoGlobal): It refers to the geographic materialization of the "Regione di Trentino" through the assignment of a region taking into account the projection and a specific work area. In this case,

Projection: WGS 84 / Lat/Lon
Bounds: North 46.60 South 45.60 East 12.00 West 10.40

- Mapset → PERMANENT

7.2.1. Import ASTER DEM & SRTM DEM

The command used to realize this operation is r.in.gal. When the command is executed, a new window appears.

In this window through the bottom browser it is possible to select the raster location file, in this case ASTER/SRTM. By default, the type of format of the browser windows is GeoTiff, taking into account that ASTER/SRTM are not is this format ArcASCII then in this parameter is selected All files.

Figure 56. Browser window
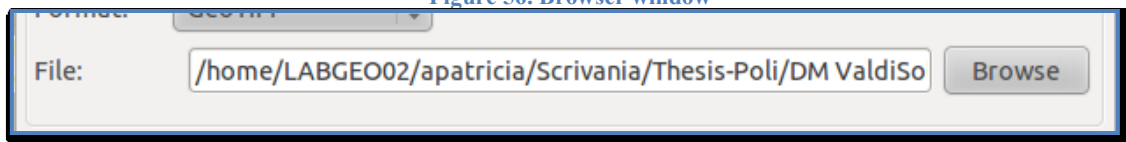
The last but not the least step, in the main import window tick "Overrite projection" and click "Import".

Note: Overrite projection uses locations projection.

Figure 57. Import window – r.in.gdal
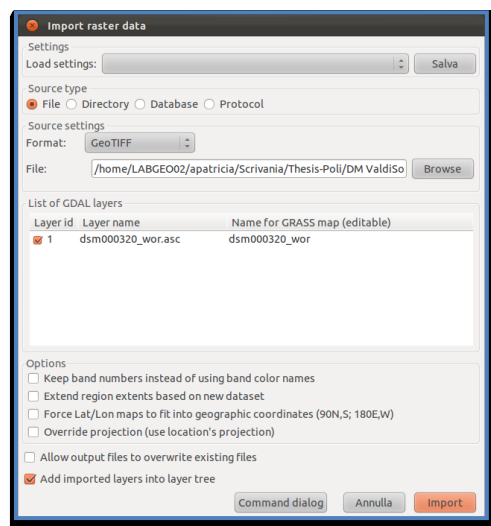
Resuming

r.in.gdal → browser → type of format = All files → ASTER → Overwrite projection → Import

7.2.2. Import GMTED2010

The same steps are followed, just with a small variation in the type of format. In addition, it is worth noting that the data that have been downloaded contains 7 files inside the DEM folder, just recalling something that was mentioned in the Chapter 2. Thus, from these 7 files only the raster map corresponding to the mean is going to be used.

Resuming

r.in.gdal → browser → **type of format = GeoTIFF** → 30n000e_20101117_gmted_mea075.tif → Overwrite projection → Import

Finally, the tree global Digital Surface Models are shown:

Figure 58. ASTER model
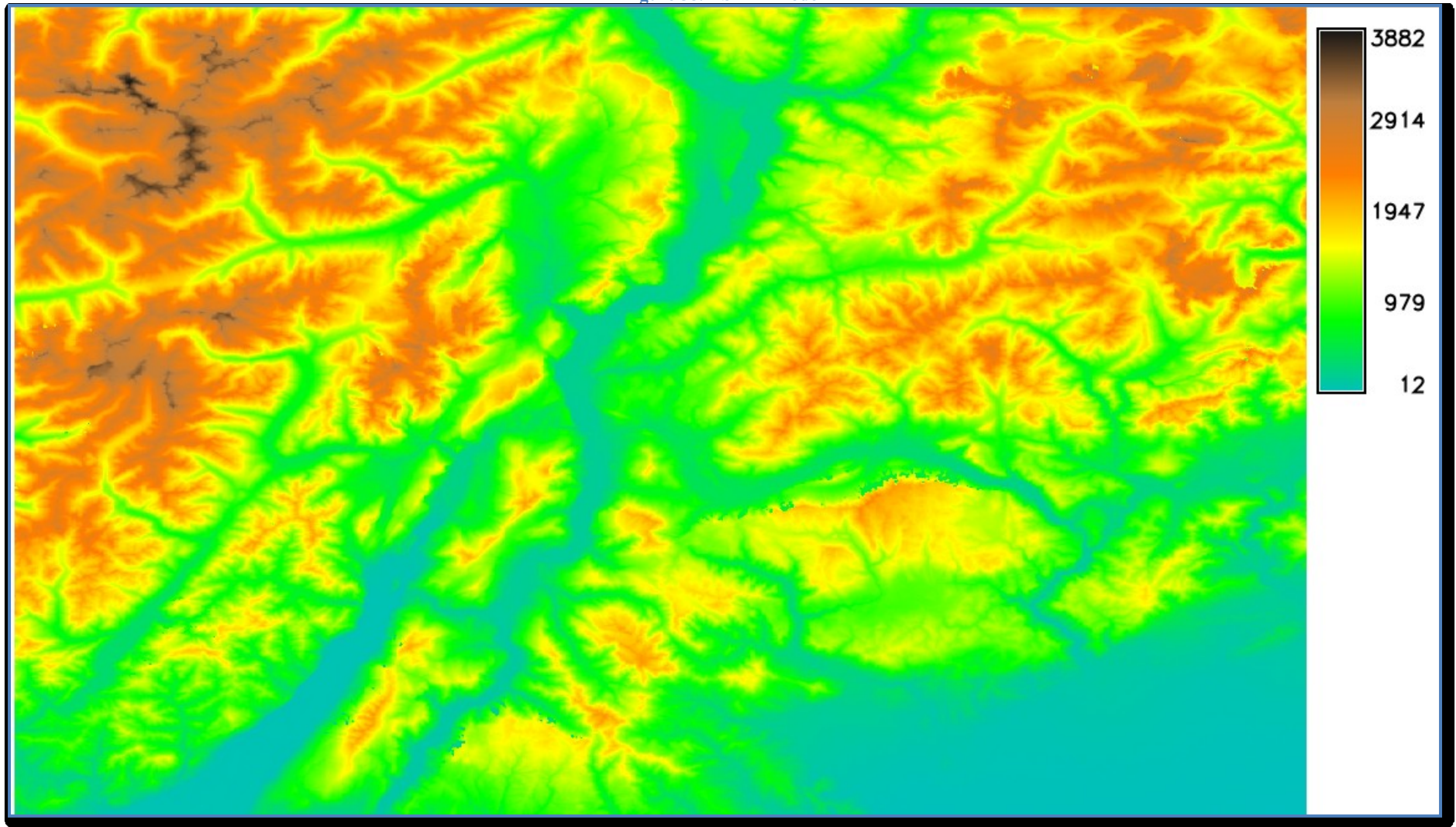
Figure 59. SRTM 90 model
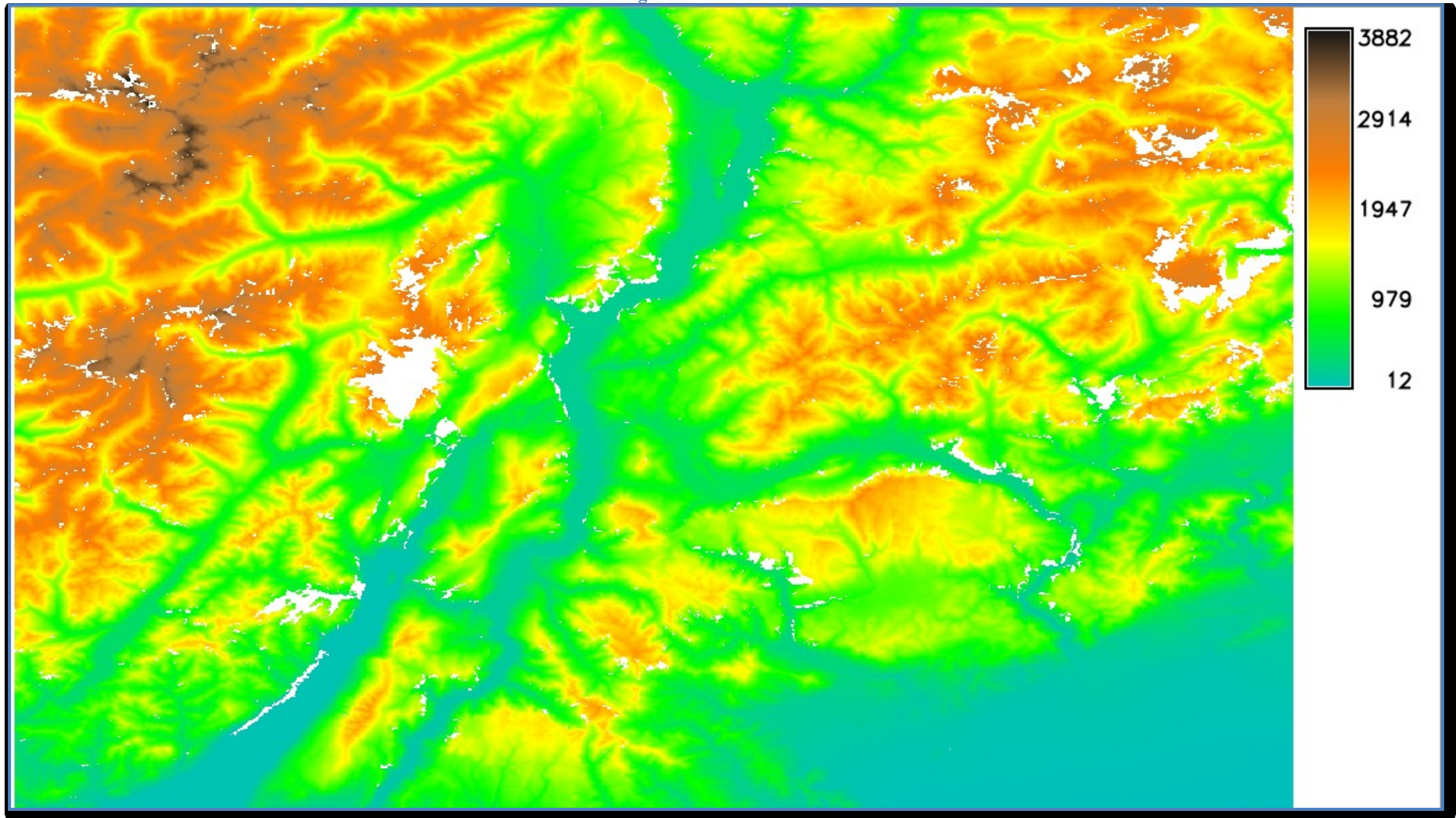
Figure 60. GMTED2010 model
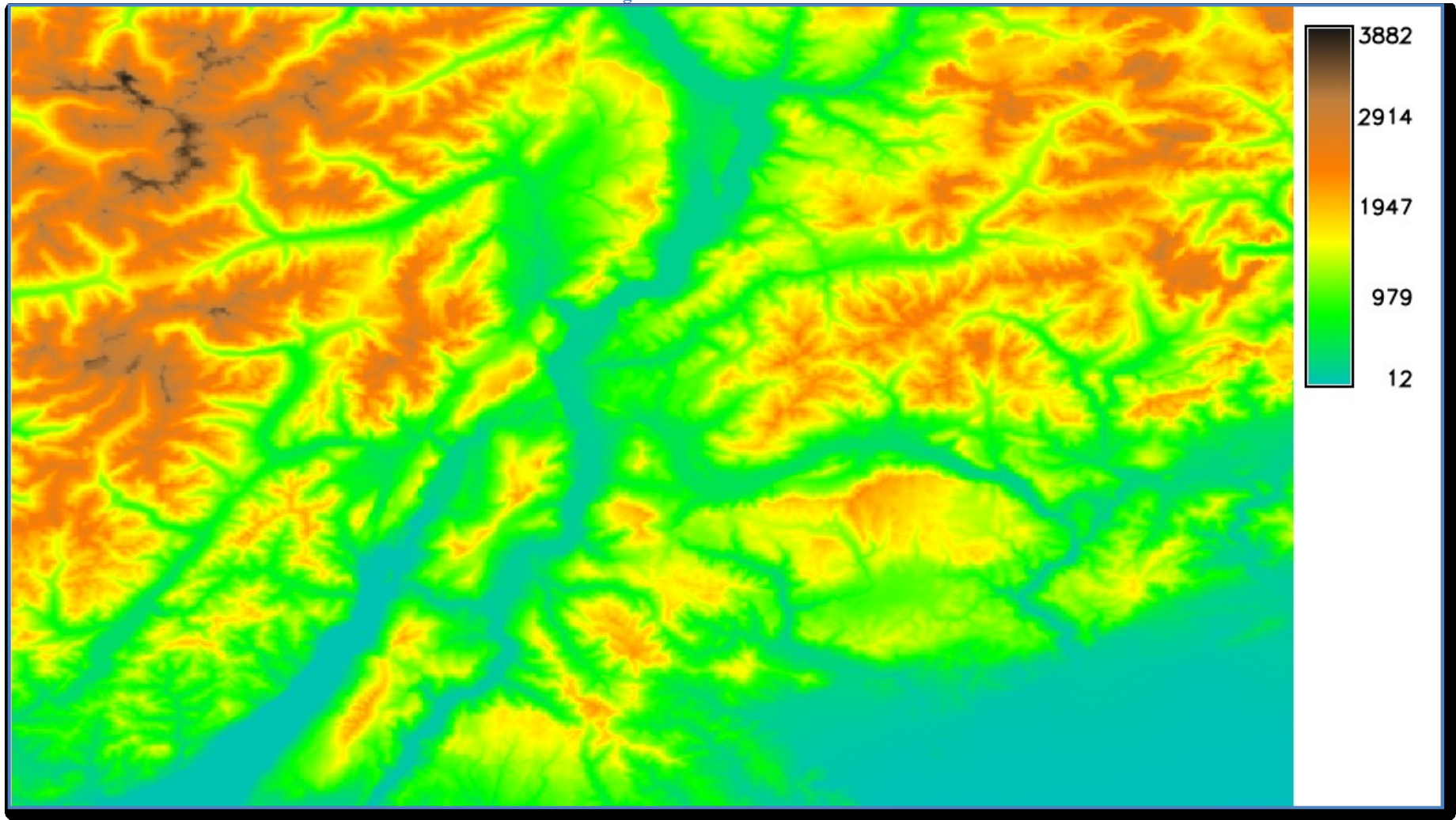
7.3. Statistics of Data

In order to have an overview of the data, some statistics are calculated. Table 13 presents the results that have been obtained by using the command r.univar.

<div align="center">

Table 13. Statistics of global DEMs

	ASTER	SRTM	GMTED
Resolution	**30 m**	**90 m**	**250 m**
No. pixels	20736000	2232124	368640
Minimum	12 m	18 m	21 m
Maximum	3882 m	3865 m	3820 m
Range	3870 m	3847 m	3799 m
Mean	1271 m	1255 m	1271 m
Standard Deviation	774 m	767 m	772 m

</div>

The first comment is about the number of pixel, clearly, the number of pixels of the ASTER model is bigger than the rest of the models due to the higher resolution that ASTER has with respect to SRTM and GMTED resolution.

The minimum and maximum values of each model illustrate the presence of plain areas that are represented by these low values as well as mountain areas with heights that can reach 3882 m.

Good agreement in general statistics.

Chapter 8

8. Comparison DSM local / DSM global SRTM

These comparisons take place in the "Val di Sole - Basin" region for two main reasons: the first is that, as was mentioned in the Chapter 5, this area comprises both an area of mountain and plain which allow identifying better the behavior of future results.

A final map is produced with the set of data of Val di Sole, downloaded from GeoCartographic Trentino Province Web Site, and the SRTM model. This map let study and analyze the difference between a local model, as is the case of Trento and the global model which correspond to SRTM. Furthermore, calculate a correlation index between the height (h) and the variation of height between the two models (Δh).

8.1. Data source

Before starting to work in the GRASS interface, is important to assign a new Mapset just with the purpose of separate the tasks. This means that the Location remains the same as in the Chapter 6 with a small modification in the boundary.

8.1.1. Import Basin data (LiDAR) to DMTRENTO Location

- Location (DMTRENTO): It refers to the geographic materialization of the "Regione di Trentino" through the assignment of a region taking into account the projection and a specific work area. In this case,

    Projection: EPSG: 32632 WGS 84 / UTM zone 32N

    Bounds: North 5157086 South 5059539 East 730000 West 612488

- Mapset → DSMTrento

Now, taking into account the quantity of data, c.a. 400 rasters, it has been developed a useful function that can import in GRASS these folders; the only required input is the path where the data are located.

Command → r.in.hdem

Figure 61. r.in.hdem - window
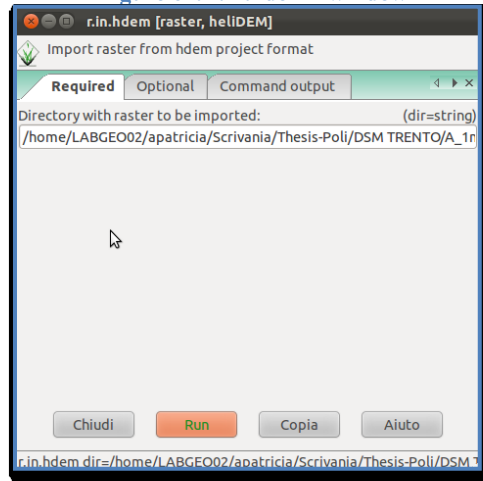
8.1.2. Import Basin data (LiDAR) to TrentoGlobal Location

Before develop this section, the question is important to be formulate:

What happens with the Trentino region data that are in another projection? The answer to the question formulated is the following: it is needed an additional transformation step.

Since the LIDAR data are imported so as referenced in the UTM projection (Location=DMTRENTO), these are suitable for processing to a Latitud/Longitud projection through the GRASS commands.

The way how GRASS does this procedure is using two different Locations. At this point the DMTRENTO location is back and from now on is worked on the GlobalTrento. It is worth noting that as two models are going to be confronted, meaning a local DSM with a global one, the local model (LiDAR) has to be resample or better has to decrease the resolution until reach the once of the global model. In this vein, three cases are going to be presented.

In addition, another mapset (Basin) is created to manage the information in an optimal way.

To develop this transformation task an useful the script is developed, this script is explained in a general way so it can be applied to SRTM model and also to ASTER and GMTED , this script not only transforms the UTM projections in Lat/Lon but also is able to decrease the resolution level from the LiDAR resolution (1m) to the global model resolutions (30m – 1 arcsecond, 90m – 3 arcsecond or 250m – 7.5 arcsecond) using different interpolation methods, namely, nearest, cubic and cubic_f.

```
#$1 min
#$2 max
#$3 res decimal degrees
#$4 method - nearest or cubic_f
```
1

```
g.region rast=Global_model_raster@PERMANENT
```
2

```
RAST_PATCH="
MIN=$1
MAX=$2
for (( i=MIN; i<=MAX; i++ ))
do
        if [ $i -lt 10 ]
        then
                NAME='00000'
        else
   if  [ $i -lt 100 ]
   then
    NAME='0000'
   else
    if  [ $i -lt 1000 ]
    then
       NAME='000'
    else
       NAME='00'
    fi
   fi
  fi
```
3

```
RAST_NAME="dsm${NAME}${i}"
 OUT_RAST="${NAME}${i}_${4}"
 echo $RAST_NAME
 echo $OUT_RAST
 r.proj  input=$RAST_NAME  location=DMTRENTO  output=$OUT_RAST  mapset=DSMTrento  method=$4
resolution=$3
 RAST_PATCH="${RAST_PATCH}${OUT_RAST},"
done
```
4

```
RAST_PATCH="${RAST_PATCH}${OUT_RAST}"
g.region rast=$RAST_PATCH
```
5

```
r.patch input=$RAST_PATCH output=trentino_$1_$2_$3_$4
```
6

```
g.remove rast=$RAST_PATCH
```
7

To understand better the script, it is explained step by step as follows

These are the four parameters that the Script needs to be executed; they coincide with the last four numbers present in the command line. The **first** and the **second** are the minimum and the maximum values of a certain range in which the 400 raster files are divided.

The division has been done because the chain of names that is built for selecting the region and/or joining the raster files, for instance, "000045_nearest, 000064_nearest, 000047_nearest, 000048_nearest, … , 000847_nearest" is so long that does not allow the execution of the script.

Remark: In the between the minimum and the maximum values exist some raster files (DSM) that have an additional feature, i.e. dsm000004-1; as these cannot be reading by the script then it is required to process it by hand.

The **third** parameter is the future resolution; it is worth noting that the value must be entered in decimal degrees. Thus, depending on the studied case this parameter changes.

- ✓ Case a = 0.000277777 (1 second)
- ✓ Case b = 0.000833333 (3 seconds)
- ✓ Case c = 0.002083333 (7.5 seconds)

The **fourth** parameter corresponds to the method use to resample the projected data. GRASS gives many options in the selection of the method but for this study are chosen the *nearest* and the *cubic*.

The nearest method, which performs a nearest neighbor assignment, is the fASTER than the cubic method. The cubic method determines the new value of the cell based on a weighted distance average of the 16 surrounding cells in the input map, however, it is possible to find a situation where at least one of the surrounding cells used to interpolate the new cell value are null, and then the resulting cell becomes also null, even if the nearest cell is not null. This will cause some thinning along null borders. The *cubic_f* interpolation method can be used if thinning along null edges is not desired. This method "falls back" to simpler interpolation methods along null borders.

This part assigns a general region which is given by a raster map and also the embedded characteristics of it.

- ✓ Case a = ASTER@PERMANENT
- ✓ Case b = SRTM_90@PERMANENT
- ✓ Case c = Gmted_mea7.5sec@PERMANENT

This fragment selects the number of ceros that are presents in the raster name file.

4 This is the body of the script where the projection process takes place. The command r.proj reads a map from a different location, here is used DMTRENTO and also the mapset DSMTrento, projects it and writes it output to the current location (GlobalTrento).

Aside from the projection, each time that is created a new projection, the output file is saved in the aforementioned chain of names because it will be useful to join or merge the new raster files.

5 Here is set the new region for joining the output files.

6 The command r.patch allows merging the raster files and creates a new map that is named with four parameters mentioned in the first part.

7 Last but not least the single raster outputs are removed.

An example for the SRTM model,

Located in the command window of GRASS 7.0, the folder where the script is written is necessary to be opened, for instance the script for this model is called trentino_dem3SRTM and it is located in the folder Script. Then the parameters specified in the box number 1 are placed next to the script name. One example is shown below:

```
GRASS 7.0. svn (TrentoGlobal):~/Scrivania/Thesis-Poli/Script > ./ trentino_dem3SRTM 362 387 0.000833333 nearest
```

For this example, the files than are going to be transform range from the 362 to 387, the resolution that is going to be reach is 0.000833333 degrees which means 3 seconds and finally the nearest method is the interpolation method used.

As is mentioned in the step 1, the 400 raster are transformed in subsets therefore when are obtained all the subsets, a merge function is needed trough the function r.patch.

At the end, the generated rasters are:

<div align="center">

Trentinor_3sec_nearest@Basin
Trentinor_3sec_cubicf@Basin

</div>

The next figures present the mentioned maps; one is generated using the nearest method and the second the cubic_f method.

Figure 62. Trentino basin – 3 sec resolution – nearest method
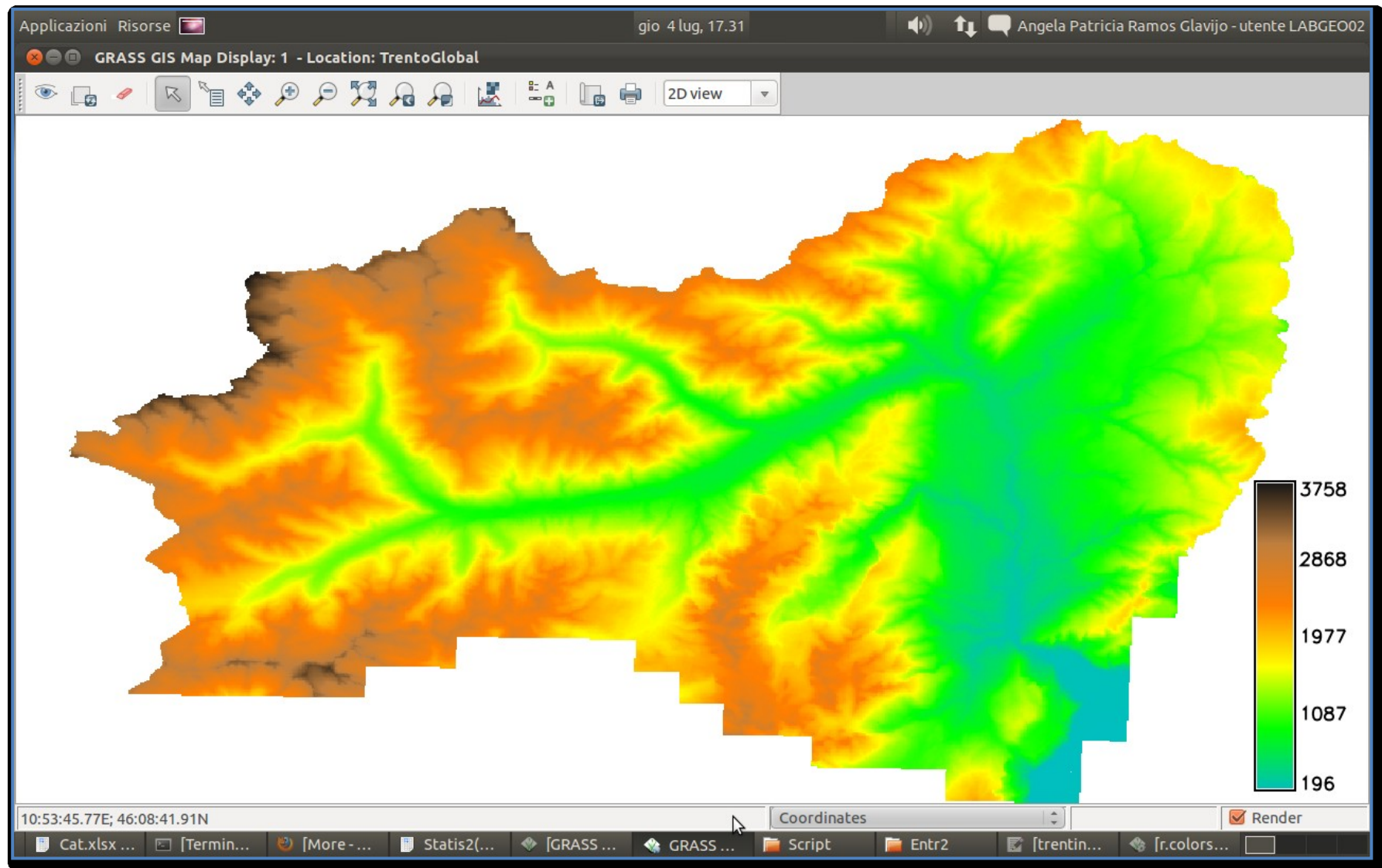
Figure 63. Trentino basin – 3 sec resolution – cubic_f method
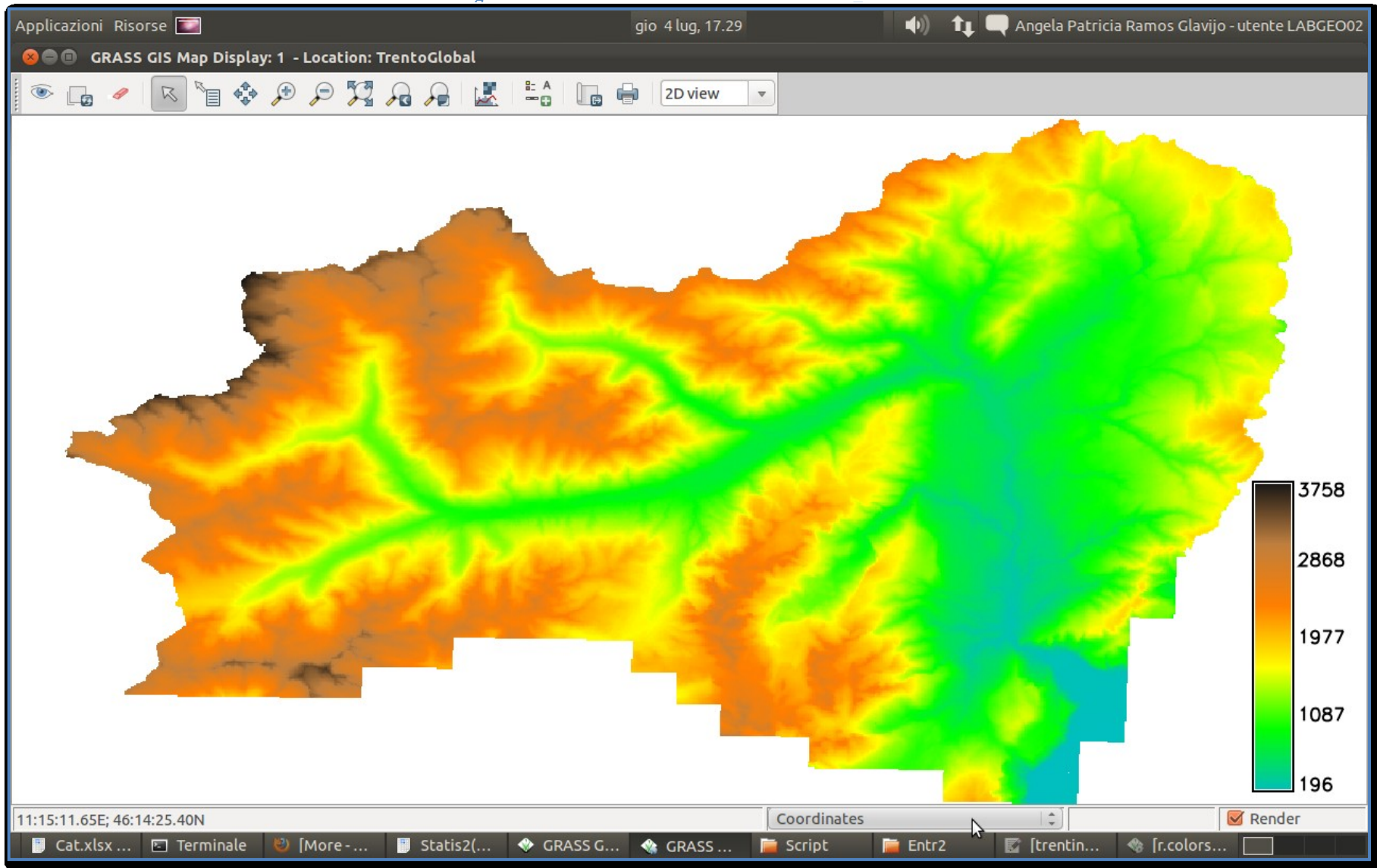
At first glance there is not much difference between the two methods; the variations can be detected, at the end of the process, just when the difference maps are done.

To interpret better this statement, an extra calculus is done, this corresponds to the difference between the raster produced by cubic_f method and the nearest method. The operation is performed through the command r.mapcal.

Example:

r.mapcalc expression= "Diff_SRTM_near_cubf= Trentinor_3sec_nearest@Basin - Trentinor_3sec_cubicf@Basin"

Once the differences are is obtained, the statistics tools are used to have a clear behavior of the results. This is done using the command r.univar.

r.univar map= Diff_SRTM_near_cubf@Basin

Then the statistics are presented in the following tables:

Table 14. Statistics of the differences between nearest and cubic_f

SRTM_cubf_near	
N. pixels	**241151**
Minimum (m)	-16 m
Maximum (m)	25 m
Range (m)	41 m
Mean (m)	0 m
Standard Deviation (m)	1 m

Also for this statistics, the respective histograms are shown

Figure 64. Histrogram of the differences between nearest and cubic_f
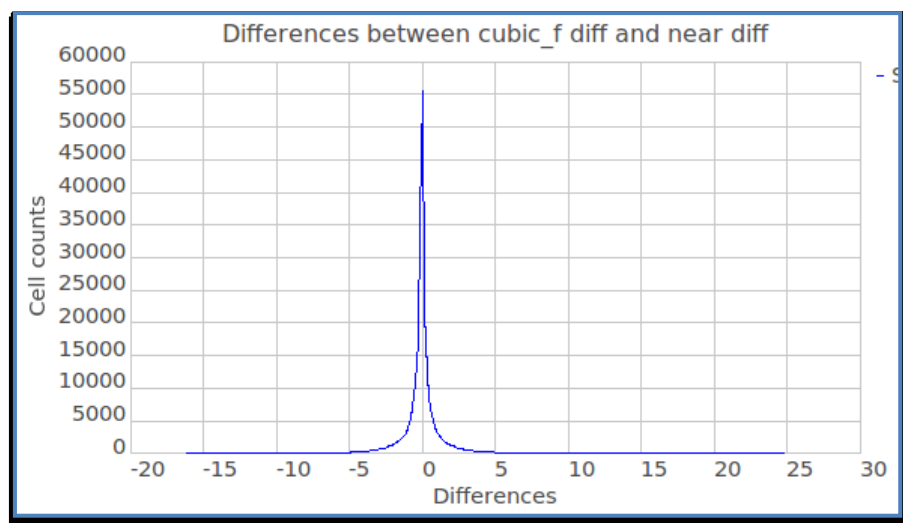
87

From the statistics and the histogram is observed that the methods either nearest or cubic_c are not so divergent.

8.2.   Data processing, analysis and final results

The main step is to run the subsequent command with its instructions:

r.mapcalc expression 'Diffr_SRTM_Trenear = SRTM_90@PERMANENT - Trentinor_3sec_nearest@Basin'

r.mapcalc expression 'Diffr_SRTM_Trecubf = SRTM_90@PERMANENT - Trentinor_3sec_cubicf@Basin'

The respective maps obtained by these operations are shown below, but before print them, it is going to be presented the basic statistics of these products:

Table 15. Differences SRTM – Trentino-Basin_nearest

SRTM -Trentinor_3sec_nearest	
N. pixels	**241151**
Minimum (m)	-263.75
Maximum (m)	351.43
Range (m)	615.18
Mean (m)	0.51
Standard Deviation (m)	25.40

Table 16. Differences SRTM – Trentino-Basin_cubic_f

SRTM -Trentinor_3sec_cubic_f	
N. pixels	**241151**
Minimum (m)	-263.82
Maximum (m)	352.00
Range (m)	615.82
Mean (m)	0.51
Standard Deviation (m)	25.37

The next histograms allow to visualize the data in a better way.

Figure 65. Histogram Differences SRTM – Trentino-Basin_nearest
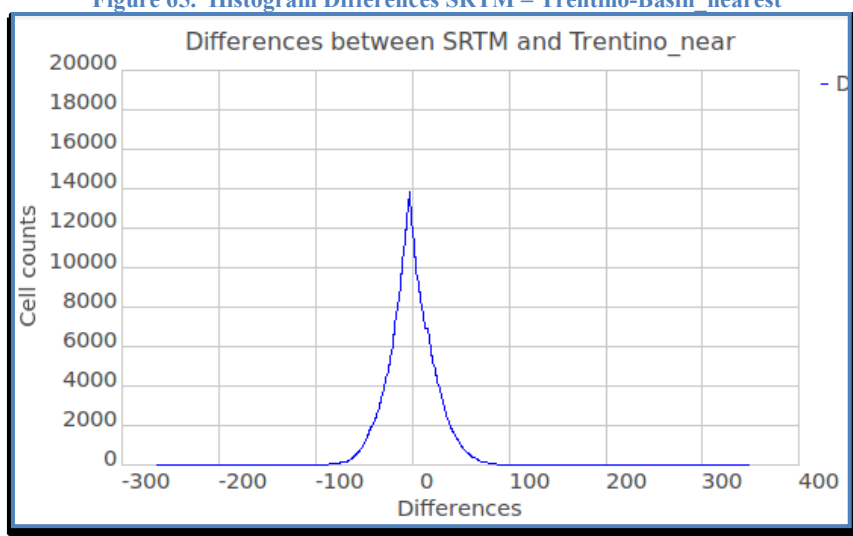
Figure 66. Histogram Differences SRTM – Trentino-Basin_cubic_f
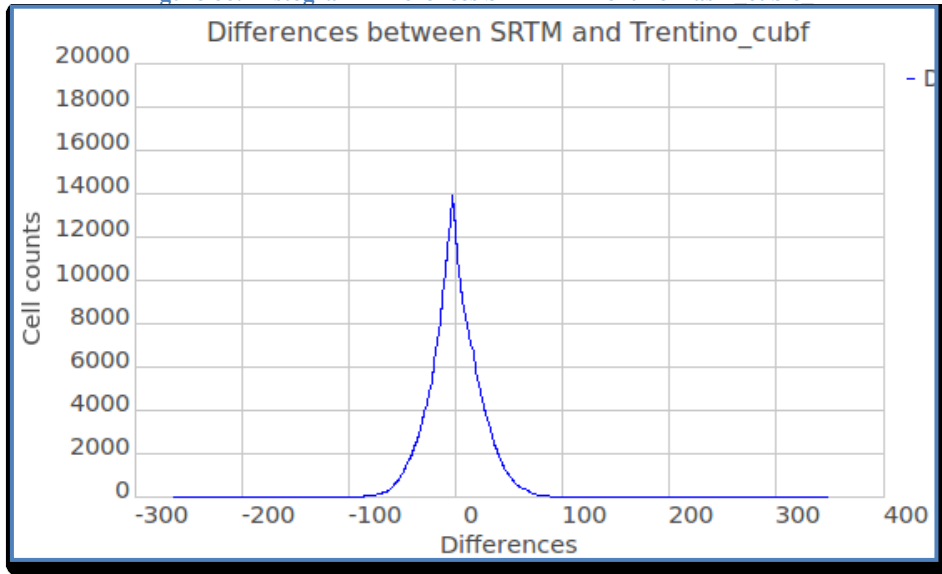
By analyzing these results, it can be said that the mean values are close to zero which means that the majority of the global DEMs are not so far from the local DEM, even though exist some outliers evidenced in the minimum and the maximum values that can be detected when built the standard deviation filter. With respect to the standard deviation, it is taken as a parameter to construct a classification; this filter is explained bellow:

General Standard Deviation classification

As the name implies this classification is based on the standard deviation (σ) of each map, then values that are between the absolute value of standard deviation ($\pm\sigma$) are considered right data, while the ones that go far from this range can become outliers that depend on acquisition errors or the type of terrain (mountain or valley), so as to understand better the behavior of data and so to picture them, they are filter in the subsequent manner:

Figure 67 Filter for map of differences

Color	Filter
	$-\sigma <$ Differences map $< \sigma$
	$-3\sigma < Dm < -\sigma$ and $\sigma < Dm < 3\sigma$
	$-5\sigma < Dm < -3\sigma$ and $3\sigma < Dm < 5\sigma$
	$-5\sigma >$ Differences map $> 5\sigma$

Figure 68. General filter model - Histogram
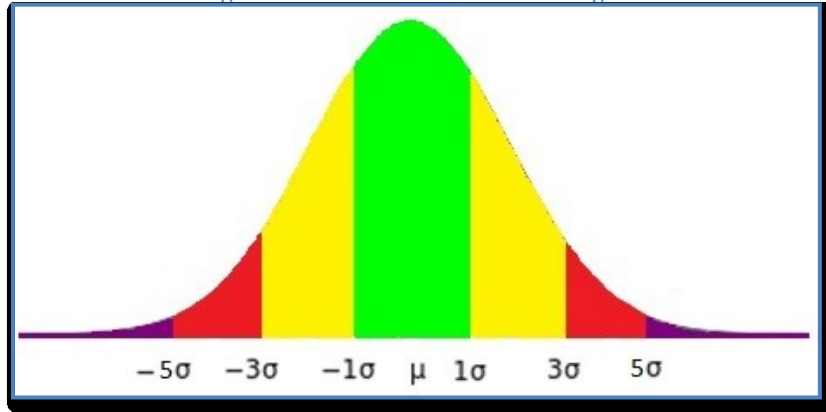
The results are the maps illustrate in Figure 69 and Figure 70.

Figure 69.Differences of Trentino-Basin (Nearest method)

Color	Filter
	-25.40 < Differences map < 25.40
	-76.20 < Dm < -25.40 and 25.40 < Dm < 76.20
	-127.00 < Dm < -76.20 and 76.20 < Dm < 127.00
	-127.00 > Differences map > 127.00
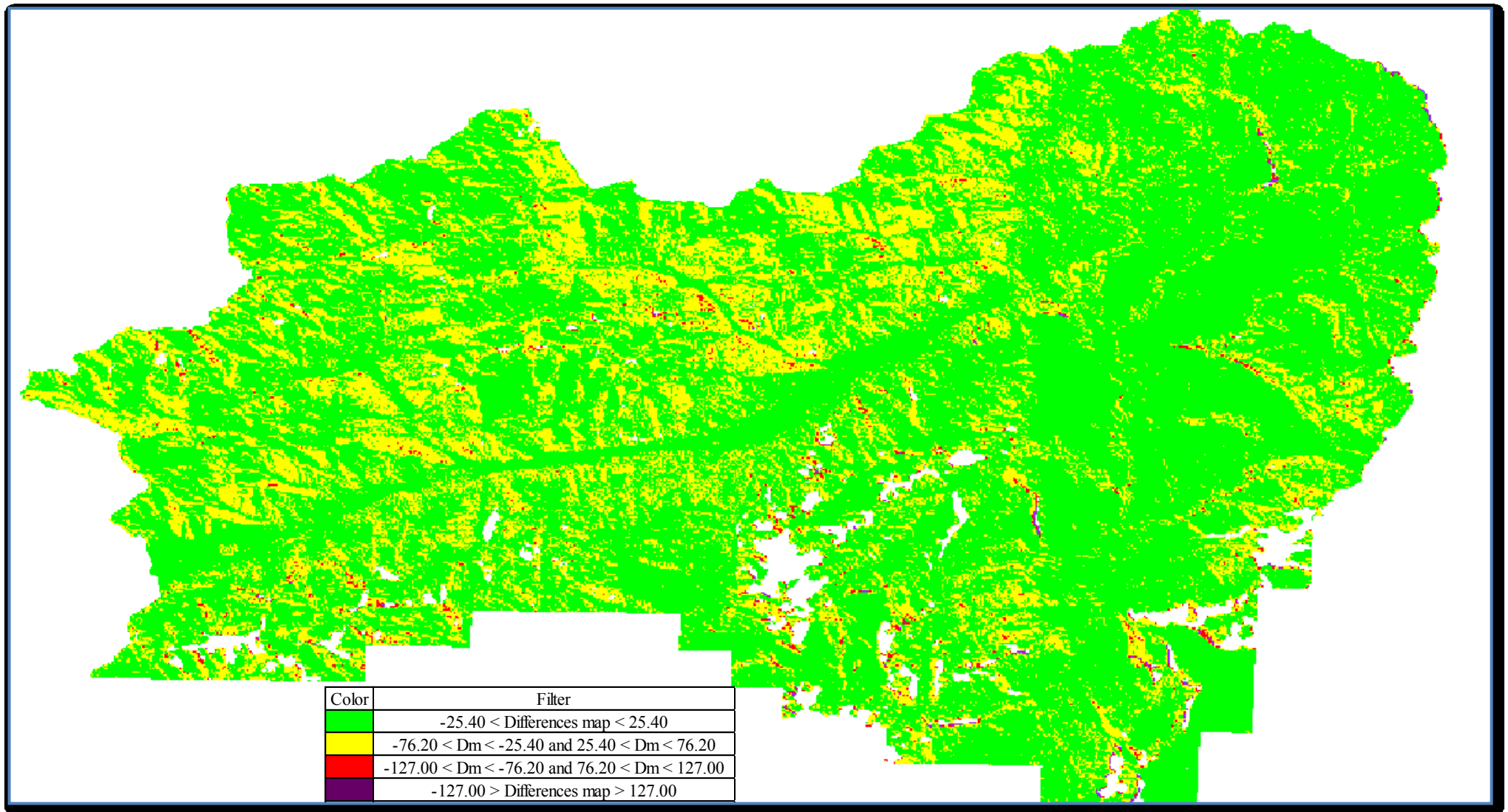
Figure 70. Differences of Trentino-Basin (Cubic_f method)

Color	Filter
	-25.37 < Differences map < 25.37
	-76.10 < Dm < -25.37 and 25.37 < Dm < 76.10
	-126.83 < Dm < -76.10 and 76.10 < Dm < 126.83
	-126.83 > Differences map > 126.83
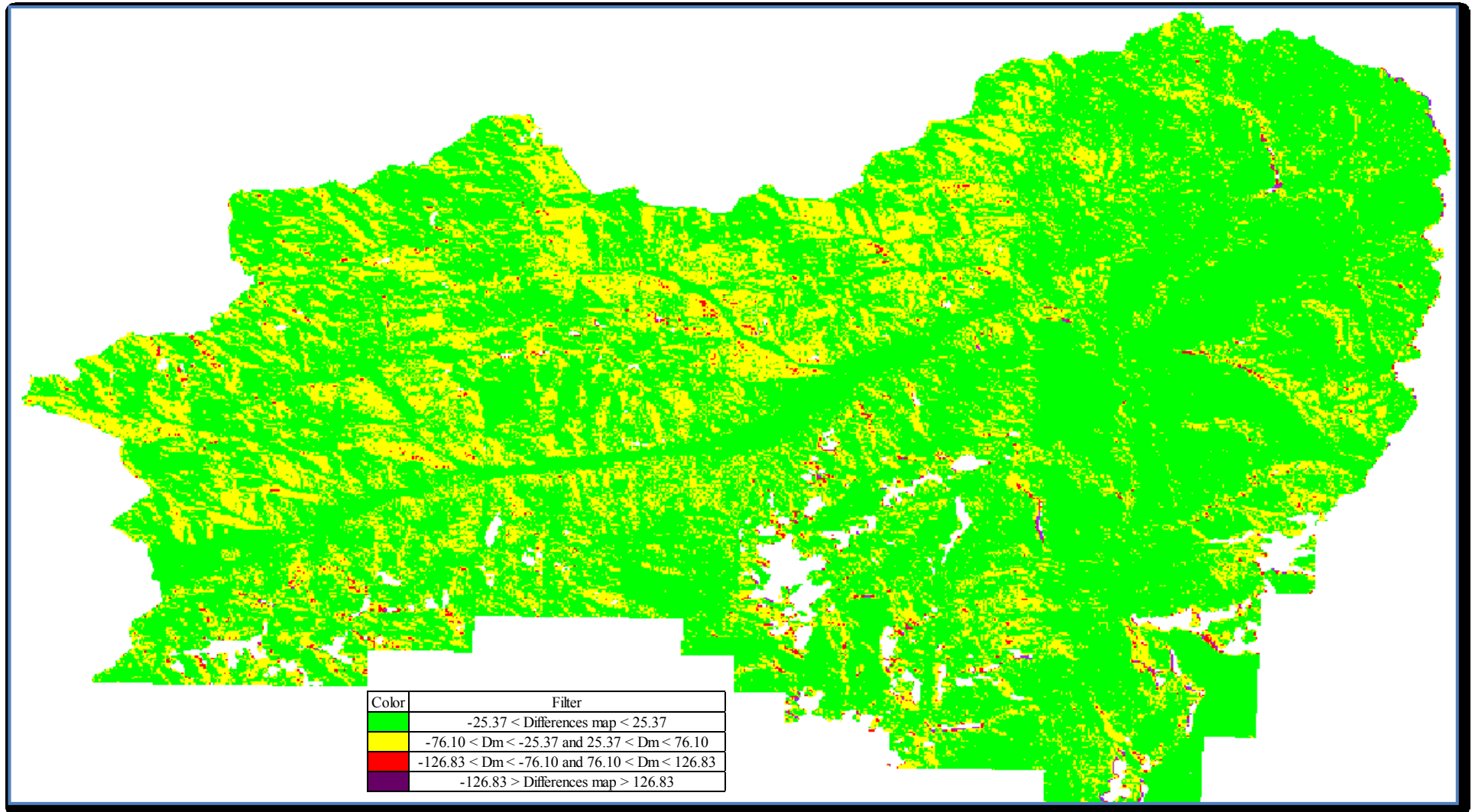
Figure 69 and Figure 70 evidence outliers that could be generated at the moment of the flight or in the data interpolation. Moreover, most of them which are represented by the yellow, red and purple color are located in the mountain area where acquiring data process is not an easy task; while in plains the results are positively.

To complete this general analysis, two new parameters are calculated:

- Correlation coefficient between Trentino elevations (h) and the differences SRTM/LiDAR (Δh): The idea of this coefficient is to verified if the differences between SRTM and LiDAR model depend on the terrain height. Otherwise another coefficient must be calculated.

    GRASS has two useful tools that allow visualizing the mentioned value; the first is Bivariate Scatterplot Tool which draws a point cloud of two variables, in this case h and Δh.

Figure 71. Bivariavate Scatterplot Trentino elevations (h) vs differences SRTM/LiDAR (Δh)
Near method (left) & Cubic_f method (right)
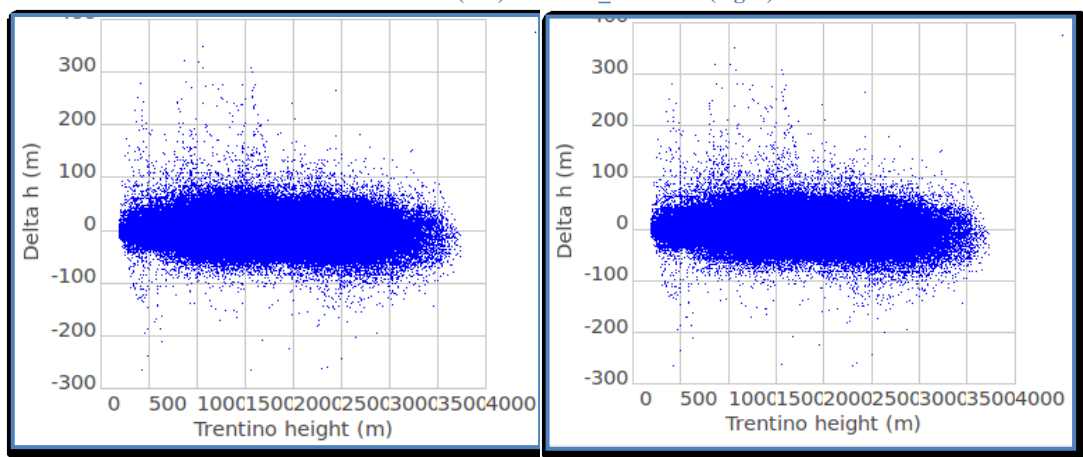
These plots indicate that for all the points no trend exists and also few points are spread along the graphs. In order to give a number to this interpretation, the second tool of GRASS is used, namely, r. covar which returns the covariance matrix and implicitly the **correlation value**.

r.covar -r map=trentino_r2_3sec_cubf@Basin,Diff2_SRTM_trenrcubf@Basin

Finally, it is obtained the matrix

$$\begin{vmatrix} 1 & -0.08 \\ -0.08 & 1 \end{vmatrix}$$

The value -0.08 indicates a no significant correlation, meaning that the differences between the two models do not depend on the height.

- Correlation coefficient between the slope (degrees) and the differences SRTM/LiDAR (Δh): With the calculation of this coefficient it is pretended to find a relationship inasmuch as, in some cases, the errors of the low accuracy DEMs do depend on the topography rather than the absolute heights. The first step is to produce the slope map through the "r.slope.aspect" GRASS command.

  r.slope.aspect elevation=trentino_r2_3_cubf@Basin slope=Slope23_3sec_cubf format=degree

The following images are the outputs of the near and cubic_f method, respectively:

Figure 72. Slope (degrees) – Trentino_3sec_near
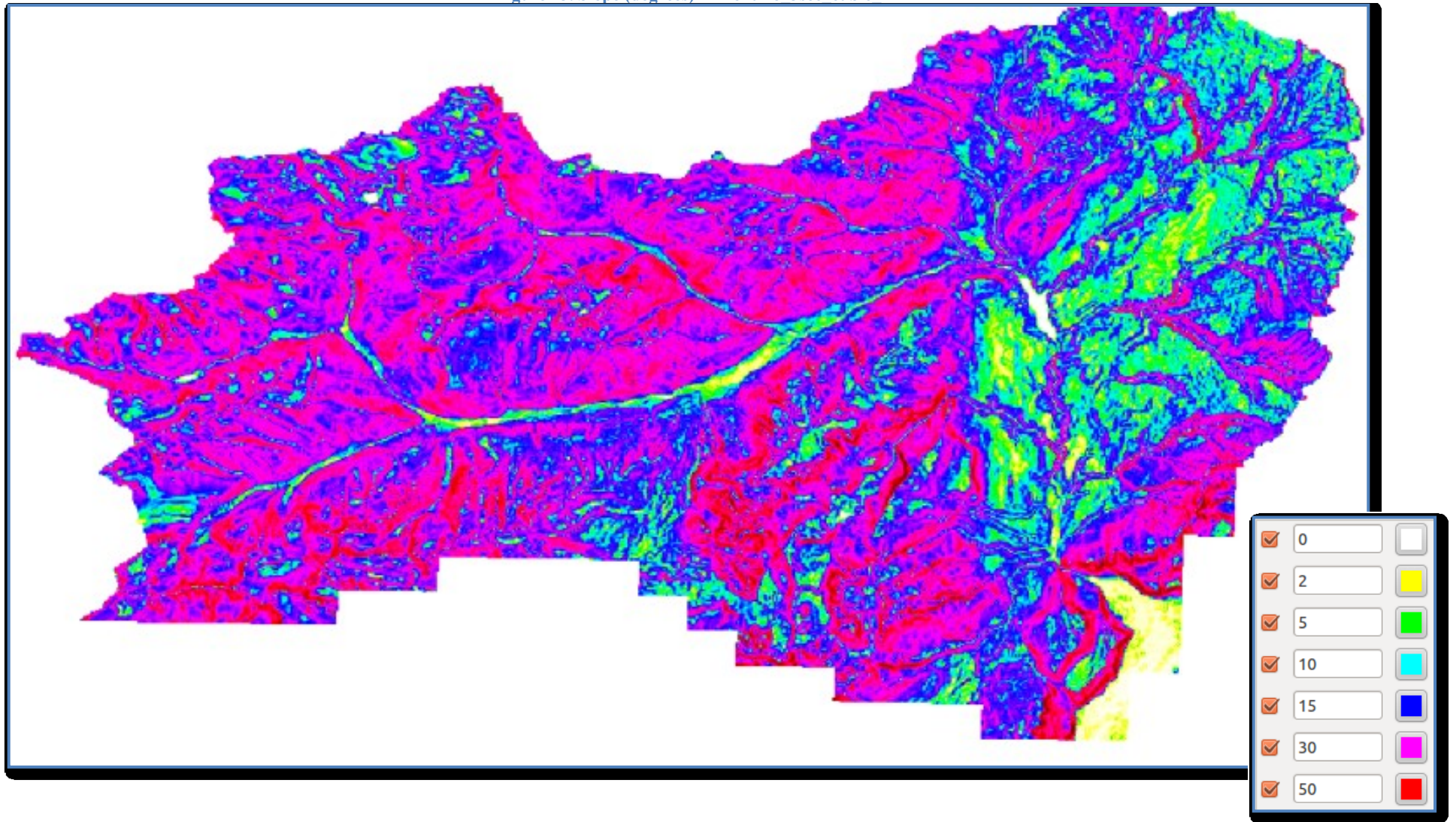
0	⬜
2	🟨
5	🟩
10	🟦
15	🟦
30	🟪
50	🟥

Figure 73. Slope (degrees) – Trentino_3sec_cubic_f

Figure 73. Slope (degrees) – Trentino_3sec_cubic_f

The pink-purple color presents a terrain with a slope between 20 – 30 degrees which in fact belongs to a mountain region while yellow-green color marks flat areas between 0-10 degrees. Last but not least blue color shows a slope from 10-20 degrees approximately.

Once again Bivariate Scatterplot and r.covar are used and they produce the following results:

Figure 74. Bivariate Scatterplot Slope_3sec (degrees) vs differences SRTM/LiDAR (Δh)
Near method (left) & Cubic_f method (right)
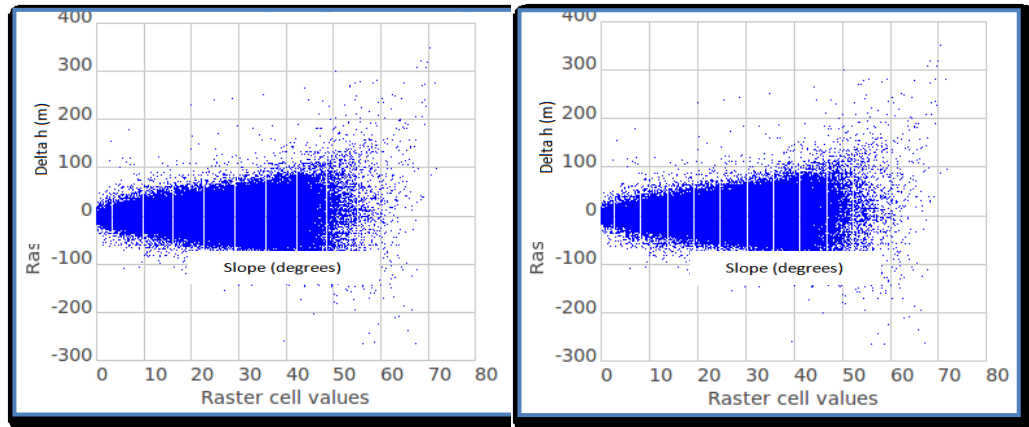
The figures illustrate a non linear dependency; however there is an increase of dispersion with the slope. The correlation coefficient can be useful to confirm it.

r.covar -r map=Slope23_3sec_cubf@Basin, Diff2_SRTM_trenrcubf@Basin

Getting,

$$\begin{vmatrix} 1 & 0.07 \\ 0.07 & 1 \end{vmatrix}$$

The results obtained by the covariance matrix produce a correlation coefficient equal to 0.07 which indicates a null relationship between these two parameters.

8.3. Classification of differences with respect to Land Uses Classes

This additional step is done to verify if the differences between the LiDAR DSMs and the SRTM DSMs depend or better can be affected due to the land uses, the land uses map used is the same that in the Chapter 6 was rasterized. Five classes are used for the classification, because they represent the bigger amount of pixels and also urban and non urban classes:

- Category 2 = Continuous urban fabric - Urban
- Category 3 = Discontinuous urban fabric - Urban
- Category 61 = Coniferous forests – Non urban
- Category 64 = Areas of natural pasture and mountain meadows – Non urban

- Category 72 = Bare rocks – Non urban

The main distinction that has to be done is that the raster usr_urb_tot@PERMANENT (Land Uses Map) is georeferenced in UTM zone 32N – WGS 84 while the raster differences Diff2_SRTM_trenrcubf@Basin is georeferenced in Lat/Lon – WGS 84. Thus a transformation process is required through the command r.proj as follows.

Example

r.proj input= usr_urb_tot@PERMANENT location=DMTRENTO output= usr_urb_tot mapset=PERMANENT method=cubic_f resolution=0.000833333

The syntax of the command requires an input which corresponds to usr_urb_tot@PERMANENT, location and mapset where the raster to be transformed is placed, an output name and the most important parameters: the interpolation method used to change the reference system and the resolution of the transformation.

Once the Land Uses map is rasterized and transform, the classification process begins.

Example for Cat 2

r.mapcalc expression="Cat_2= if(usr_urb_tot@Basin==2 , Diff2_SRTM_trenrcubf@Basin, null())"

This syntax just store in an output raster, for example Cat_2@Basin, the difference values, in this case _SRTM_trenrcubf@Basin, that belongs to land use category (usr_urb_tot@Basin) that is been considered and the values that do not fulfill the condition are filled with a null value.

This process is done for the five categories mentioned.

Once the differences are classified (Figure 75
Figure 76, Figure 77, Figure 78, Figure 79) some statistics tools are applied to the output maps Cat_2@Basin.

Table 17. Statistics of classification

Cat	No. Pixels	Min (m)	Max (m)	Mean (m)	Std Dev (m)	Description
2	214	-29	40	-2	9	Continuous urban fabric
3	3012	-59	51	-5	10	Discontinuous urban fabric
61	88816	-261	244	3	25	Coniferous forests
64	25673	-262	140	-7	21	Areas of natural pasture and mountain meadows
72	37070	-258	320	-1	31	Bare rocks

Figure 75. Cat 2_diff_SRTM_Trent
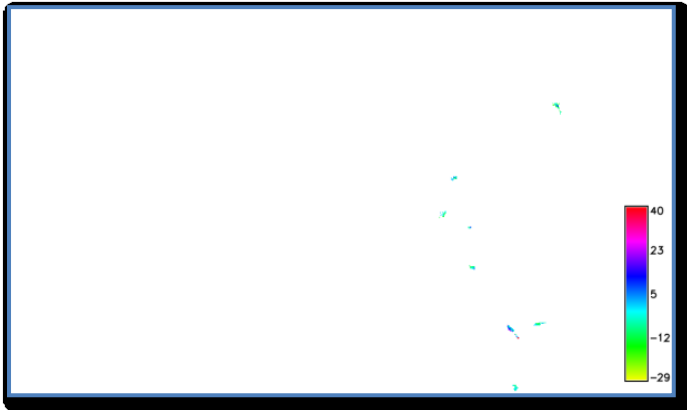
Figure 76. Cat 3_ diff_SRTM_Trent
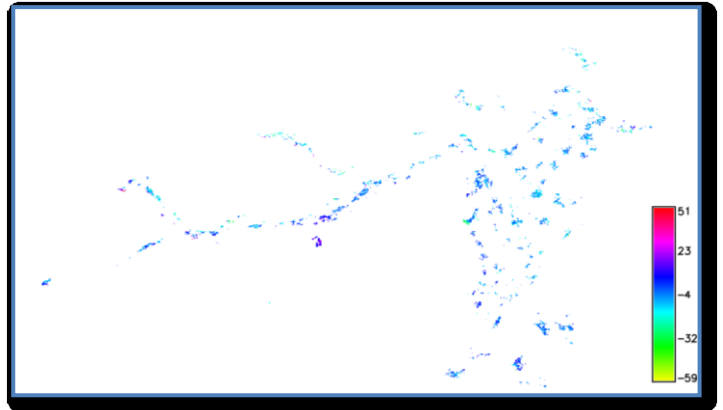
Figure 77. Cat 61_ diff_SRTM_Trent
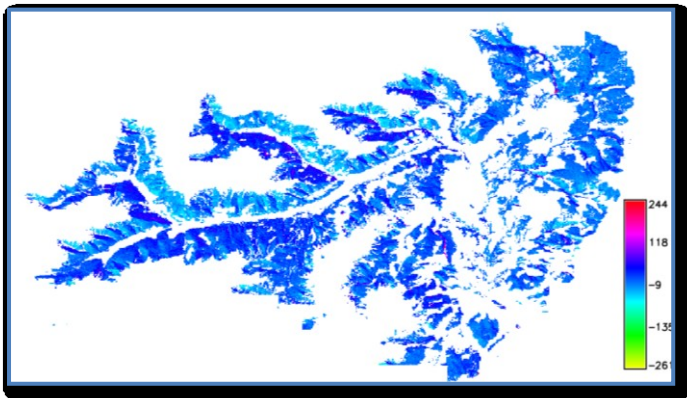
Figure 78. Cat 64_ diff_SRTM_Trent
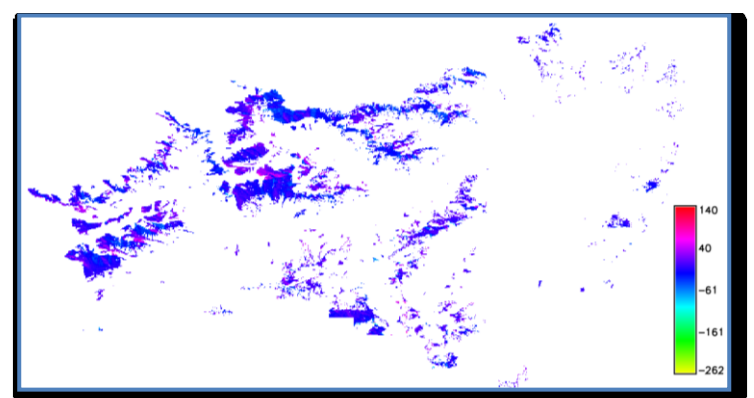
Figure 79. Cat 72_ diff_SRTM_Trent
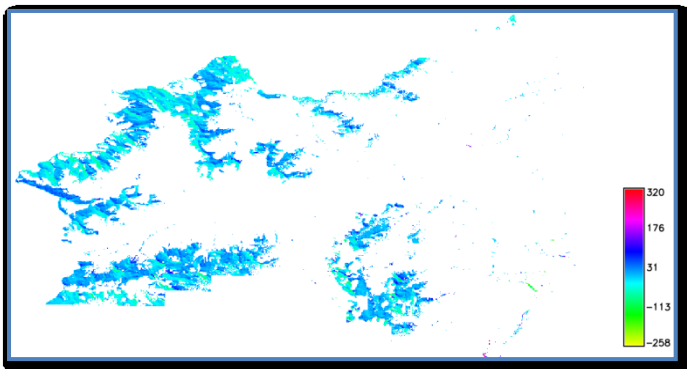
Regarding to the statistics, specifically the mean, is possible to conclude that the biggest errors are present in the Category 61 that represents areas of natural pasture and mountain meadows where the acquisition process has its weaknesses. On the other hand the smallest errors are present in Category 72 represented by bare rock, indeed as no obstruction exits between the acquisition and the terrain the expected values should be smaller.

Chapter 9

9. Comparisons DSM local / DSM global ASTER

As the previous later chapter, here the upgrated set of data of the Val di Sol – Basin region is going to be employed to produce a final map that displays differences between the ASTER DEM with respect to the local DEM. Furthermore, it is calculated a correlation index between the height (h) and the variation of height between the two models (Δh).

9.1.   Data source

The same process for importing data is required then, working from GlobalTrento-Location and the mapset Basin, the script mentioned in the section 8.1.2.is run with some specific characteristics using the following sintaxys.

Example for ASTER model:

 GRASS 7.0. svn (TrentoGlobal):~/Scrivania/Thesis-Poli/Script > ./ trentino_dem3 164 187 0.000277777 cubic_f

For this example, the files than are going to be transform range from the 164 to 187, the resolution that is going to be reach is 0.000277777 degrees which means 1 second and finally the cubic_f method is the interpolation method used.

As is mentioned in section 8.1.2 , the 400 raster are transformed in subsets therefore when are obtained all the subsets, a merge function is needed trough the function r.patch.

At the end, the generated rasters are:

<div align="center">
Trentinor_1sec_nearest@Basin
Trentinor_1sec_cubicf@Basin
</div>

The next figures present the mentioned maps; one is generated using the nearest method and the second the cubic_f method.

Figure 80. Trentino basin – 1 sec resolution – nearest method
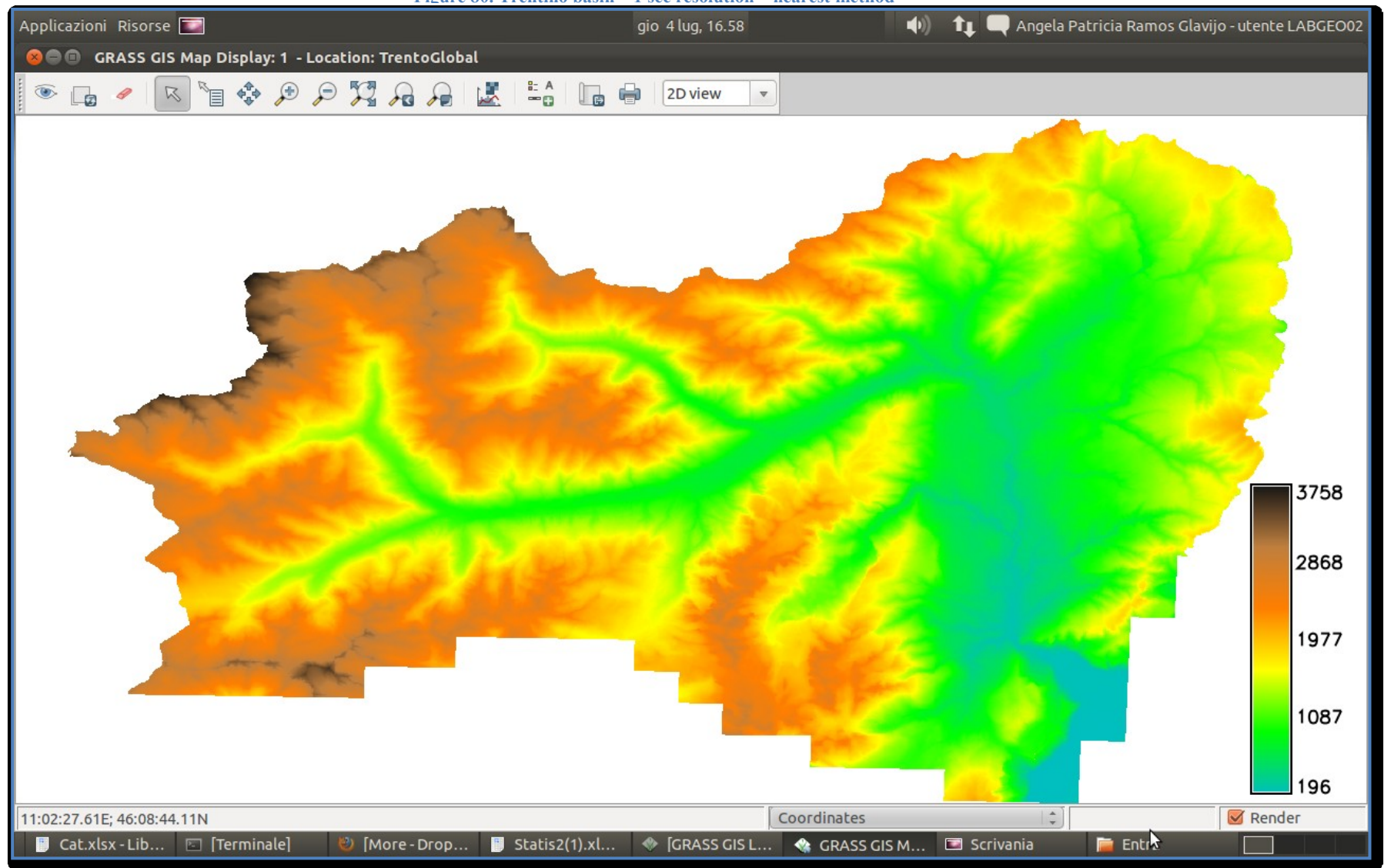
Figure 81. Trentino basin – 1 sec resolution – cubic_f method
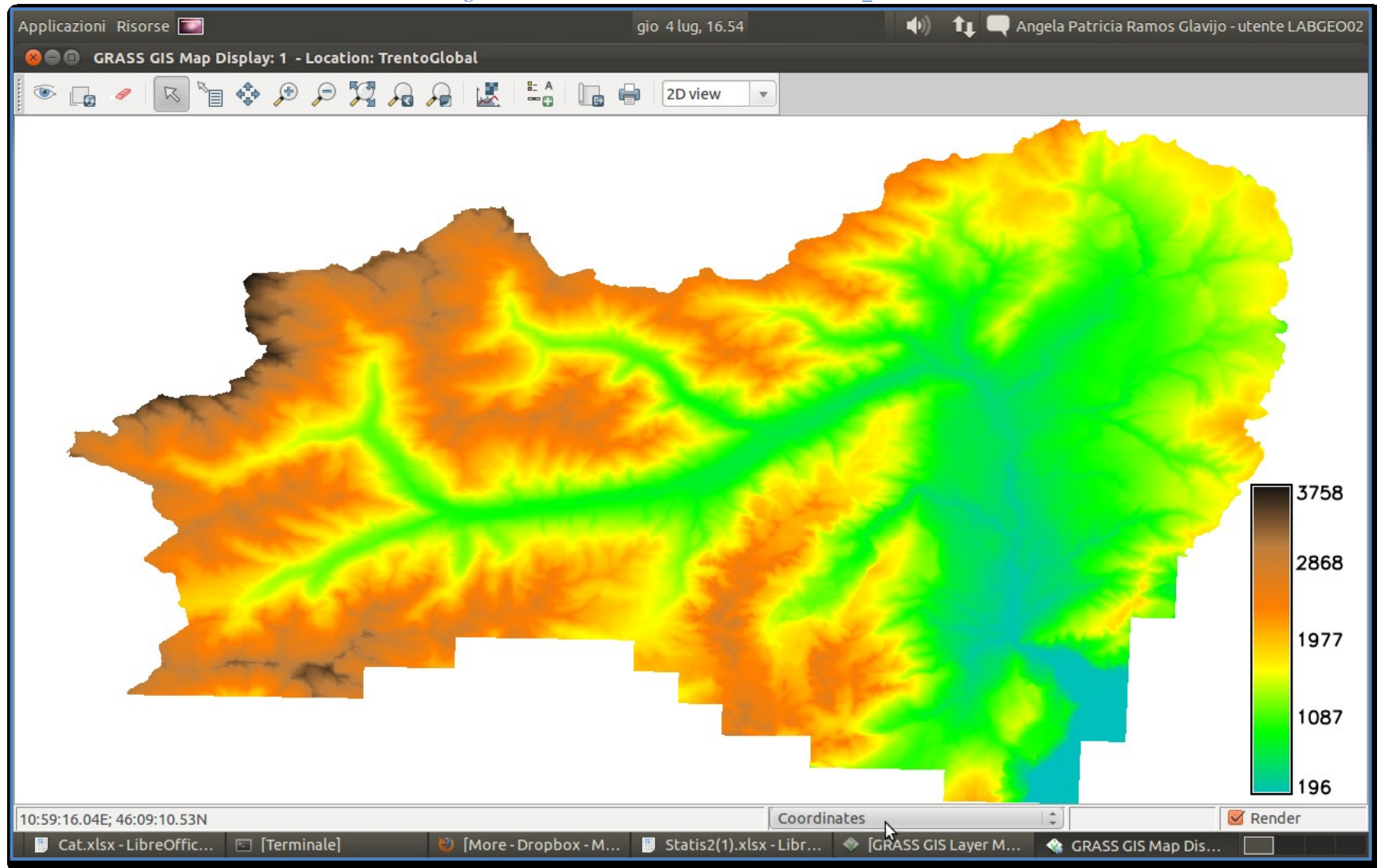
At first glance there is not much difference between the two methods; the variations can be detected, at the end of the process, just when the difference maps are done.

9.2. Data processing, analysis and final results

The main step is to run the subsequent command with its instructions:

r.mapcalc expression 'Diffr_ASTER_Trenear = ASTER - Trentinor_1sec_nearest'

r.mapcalc expression 'Diffr_ASTER_Trecubf = ASTER - Trentinor_1sec_cubicf
'

The respective maps obtained by these operations are shown below, but before print them, it is going to be presented the basic statistics of these products:

Table 18.Differences ASTER – Trentino-Basin_nearest

Table 19. Differences ASTER – Trentino-Basin_cubic_f

ASTER -Trentinor_1sec_nearest	
N. pixels	**2234673**
Minimum (m)	-1292.28
Maximum (m)	301.94
Range (m)	1594.22
Mean (m)	2.21
Standard Deviation (m)	20.43

ASTER -Trentinor_1sec_cubic_f	
N. pixels	**2234673**
Minimum (m)	-925.11
Maximum (m)	301.85
Range (m)	1226.96
Mean (m)	2.21
Standard Deviation (m)	20.37

Also for this statistics, the respective histograms are shown

Figure 82. Histogram Differences ASTER – Trentino-Basin_nearest
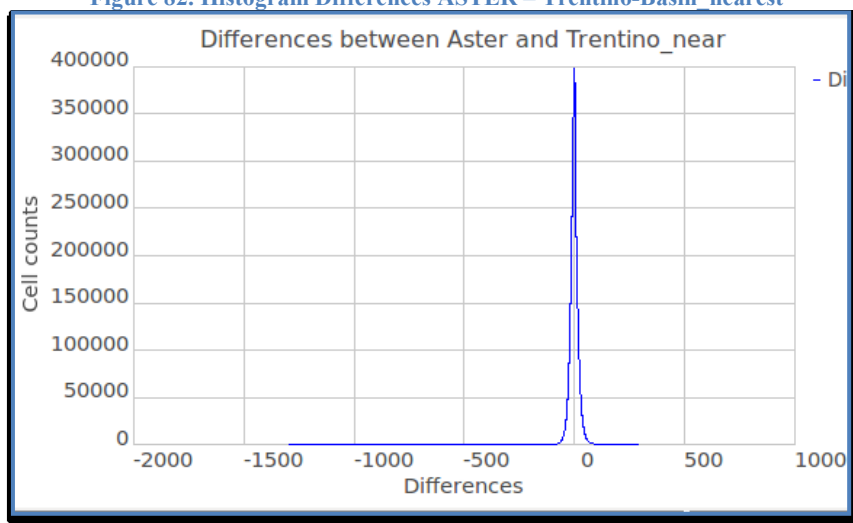
Figure 83. Histogram Differences ASTER – Trentino-Basin_cubic_f
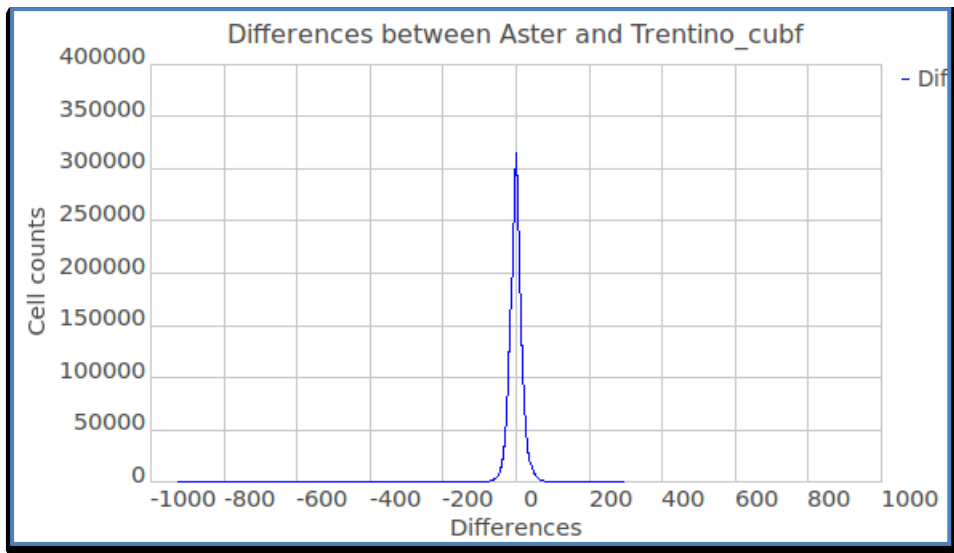
103

Differences between Aster and Trentino_cubf

By analyzing these results, it can be said that the mean values are close to zero which means that the majority of the global DEMs are not so far from the local DEM, even though exist some outliers evidenced in the minimum and the maximum values that can be detected when built the standard deviation filter. With respect to the standard deviation, it is taken as a parameter to construct the mentioned classification; this was explained in the chapter 8.

The results are the maps illustrate in Figure 85 and Figure 86

To complete this analysis an extra calculus is done, this corresponds to the difference between the differences produced by cubic_f method and the ones produced by the nearest method. The operation is performed through the command r.mapcal.

Example:

r.mapcalc expression= "Diff_ASTER_near_cubf= Diffr_ASTER_Trenear@Basin - Diffr_ASTER_Trecubf@Basin"

Once the differences are is obtained, the statistics tools are used to have a clear behavior of the results. This is done using the command r.univar.

r.univar map= Diff_ASTER_near_cubf@Basin

Then the statistics are presented in the following tables:

ASTER_cubf_near	
N. pixels	**2234673**
Minimum (m)	-475 m
Maximum (m)	598 m
Range (m)	10739 m
Mean (m)	0 m
Standard Deviation (m)	2 m

Also for this statistics, the respective histogram is shown

Figure 84. Histrogram of the differences between nearest diff and cubic_f diff
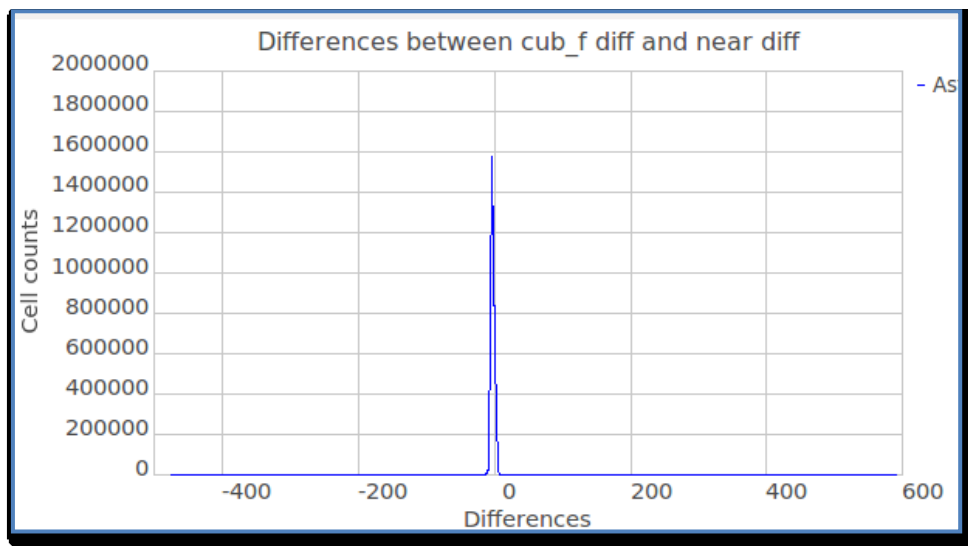
From the statistics and the histogram is observed that the methods either nearest or cubic_c are not so divergent because the mean is zero which means that most of the data do not have differences.

Figure 85.Differences of Trentino-Basin (Nearest method)

Color	Filter
	-20.43 < Differences map < 20.43
	-61.29 < Dm < -20.43 and 20.43 < Dm < 61.29
	-102.15 < Dm < -61.29 and 61.29 < Dm < 102.15
	-102.15 > Differences map > 102.15
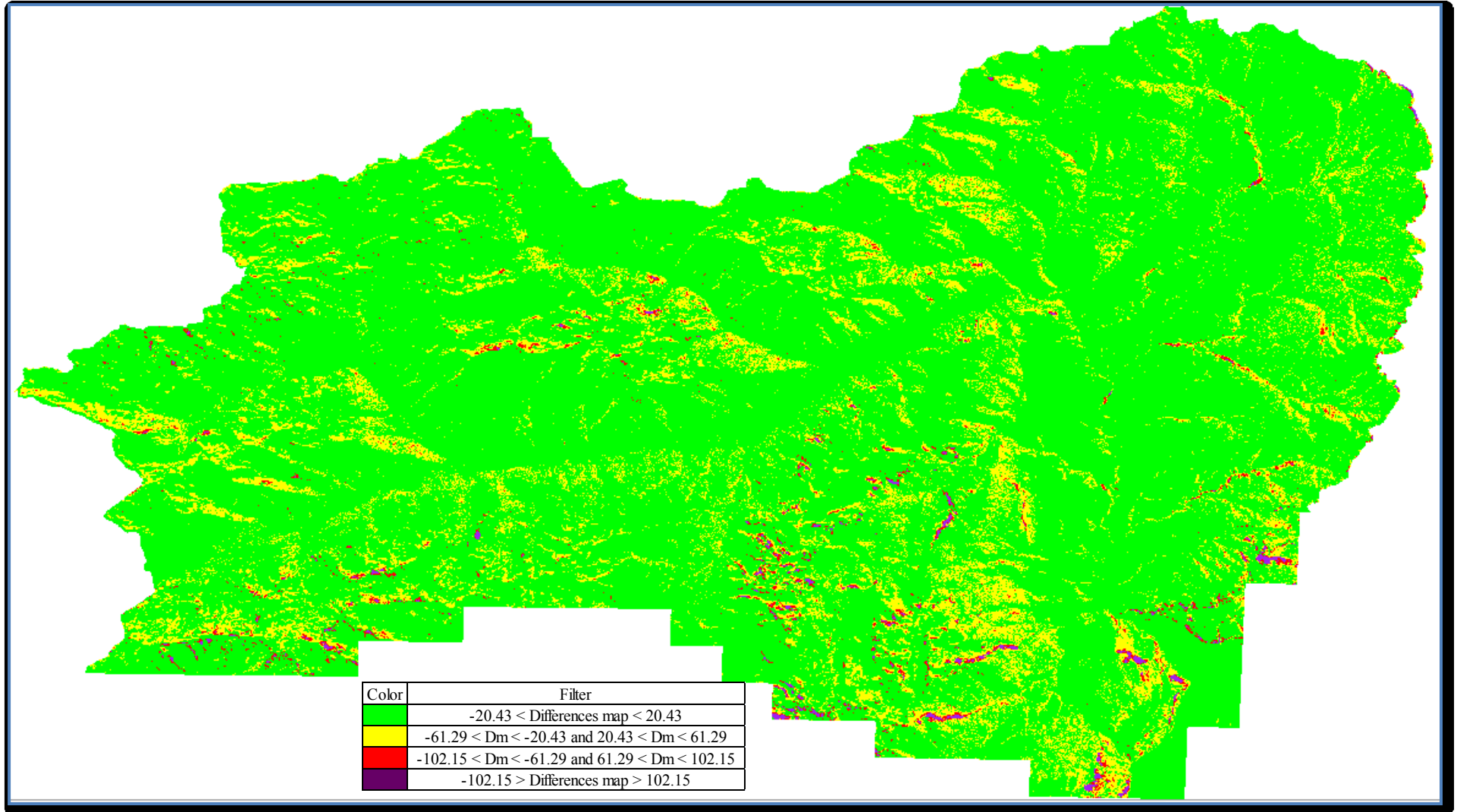
Figure 86. Differences of Trentino-Basin (Cubic_f method)

Color	Filter
	-20.37 < Differences map < 20.37
	-61.11 < Dm < -20.37 and 20.37 < Dm < 61.11
	-101.85 < Dm < -61.11 and 61.11 < Dm < 101.85
	-101.85 > Differences map > 101.85
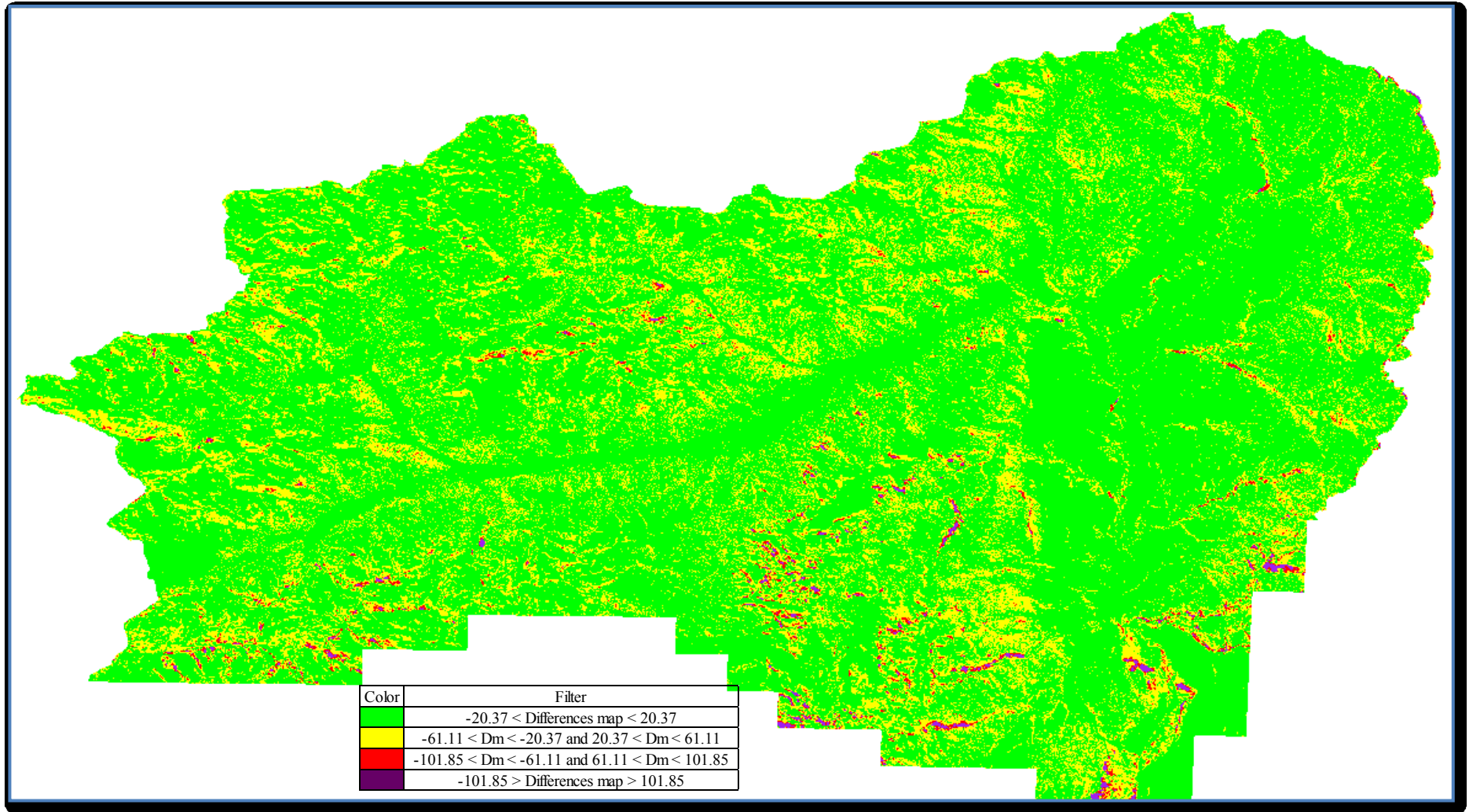
Figure 85 and Figure 86 evidence outliers that could be generated at the moment of the flight or in the data interpolation. Nevertheless, due to the accuracy of both models (LiDAR & ASTER) the amount of them (outliers) is less with respect to SRTM/LiDAR outliers. Here the atypical points are also drawn with yellow, red and purple colors. In a lower scale the anomalies are located in the mountain areas where acquiring data process is not an easy task; while in plains the results are positively accepted.

To complete this general analysis, two new parameters are calculated:

- Correlation coefficient between Trentino elevations (h) and the differences ASTER/LiDAR (Δh): The idea of this coefficient is to verified if the differences between SRTM and LiDAR model depend on the terrain height. Otherwise another coefficient must be calculated.

  GRASS has two useful tools that allow visualizing the mentioned value; the first is Bivariate Scatterplot Tool which draws a point cloud of two variables, in this case h and Δh.

Figure 87.Bivariavate Scatterplot Trentino elevations (h) vs differences ASTER/LiDAR (Δh) Near method (left) & Cubic_f method (right)

These plots indicate a certain trend for the majority of the points as well as few of them spread along the graphs. In order to give a number to this interpretation, the second tool of GRASS is used, namely, r. covar which returns the covariance matrix and implicitly the **correlation value**.

r.covar -r map=trentino_r2_1sec_cubf@Basin,Diff2_ASTER_trenrcubf@Basin

Finally, it is obtained the matrix

$$\begin{vmatrix} 1 & -0.08 \\ -0.08 & 1 \end{vmatrix}$$

The value -0.08 indicates an insignificant correlation, meaning that the differences between the two models do not depend on the height.

- Correlation coefficient between the slope (degrees) and the differences ASTER/LiDAR (Δh): With the calculation of this coefficient it is pretended to find a relationship inasmuch as, in some cases, the errors of the low accuracy DEMs do depend on the topography rather than the absolute heights. The first step is to produce the slope map through the "r.slope.aspect" GRASS command.

    r.slope.aspect elevation=trentino_r2_1_cubf@Basin slope=Slope23_1sec_cubf format=degree

The following images are the outputs of the near and cubic_f method, respectively:

Figure 88. Slope (degrees) – Trentino_1sec_near

0	⬜
2	🟨
5	🟩
10	🟦
15	🟦
30	🟪
50	🟥
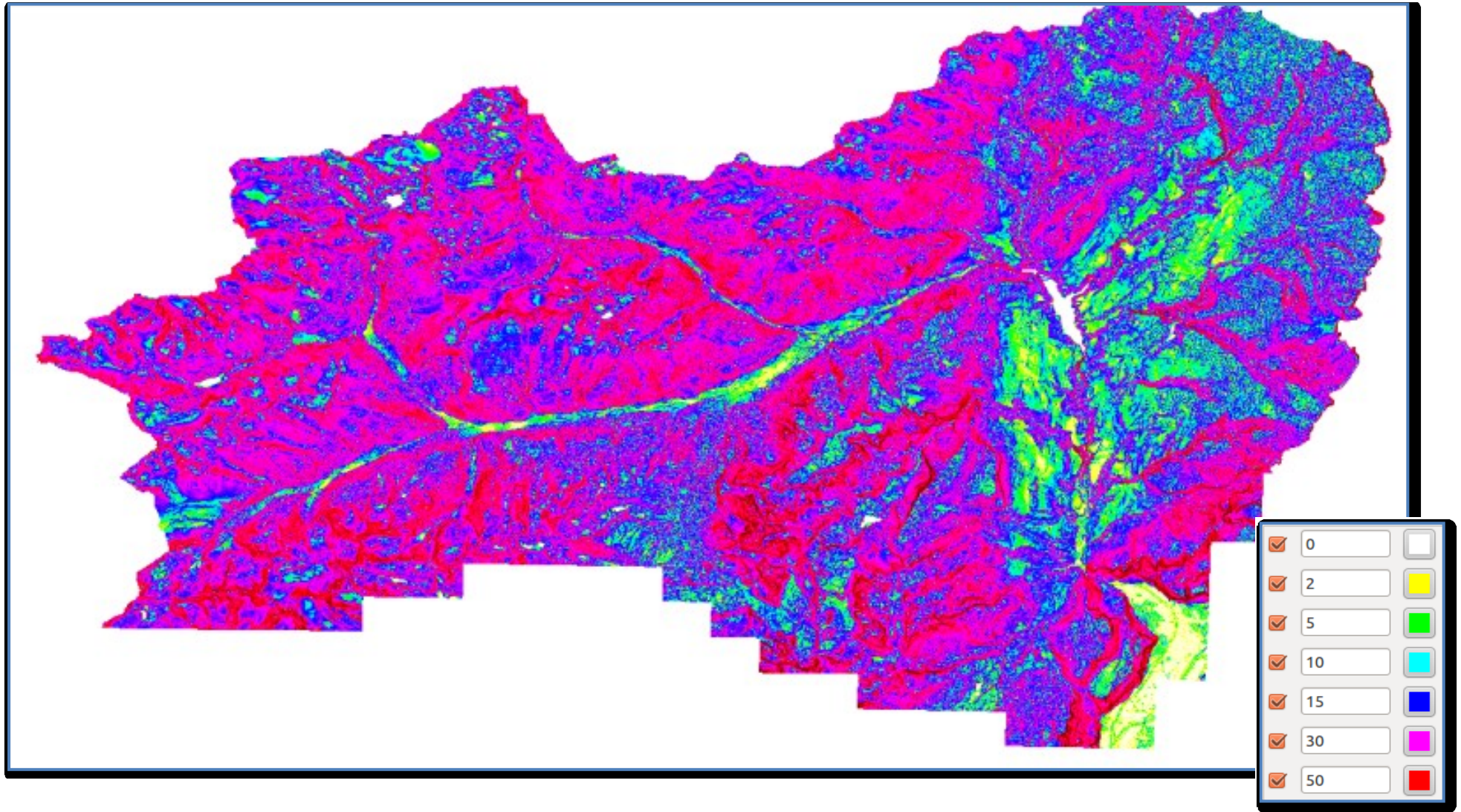
Figure 89. Slope (degrees). Trentino_1sec_cubic_f
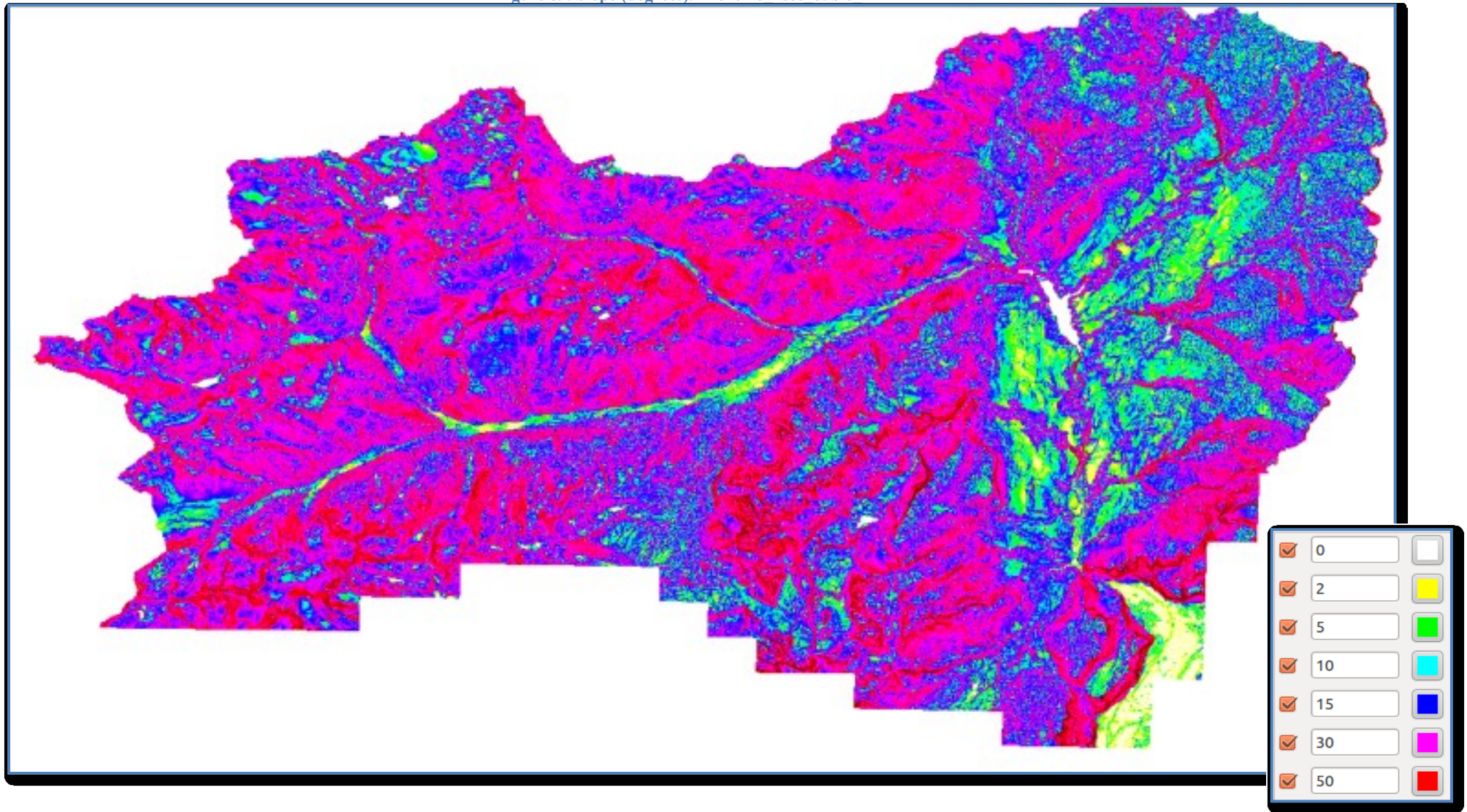
The pink-purple color presents a terrain with a slope between 20 – 30 degrees which in fact belongs to a mountain region while yellow-green color marks flat areas between 0-10 degrees. Last but not least blue color shows a slope from 10-20 degrees approximately.

Once again Bivariate Scatterplot and r.covar are used and they produce the following results:

Figure 90. Bivariavate Scatterplot Slope_1sec (degrees) vs differences ASTER/LiDAR (Δh)
Near method (left) & Cubic_f method (right)
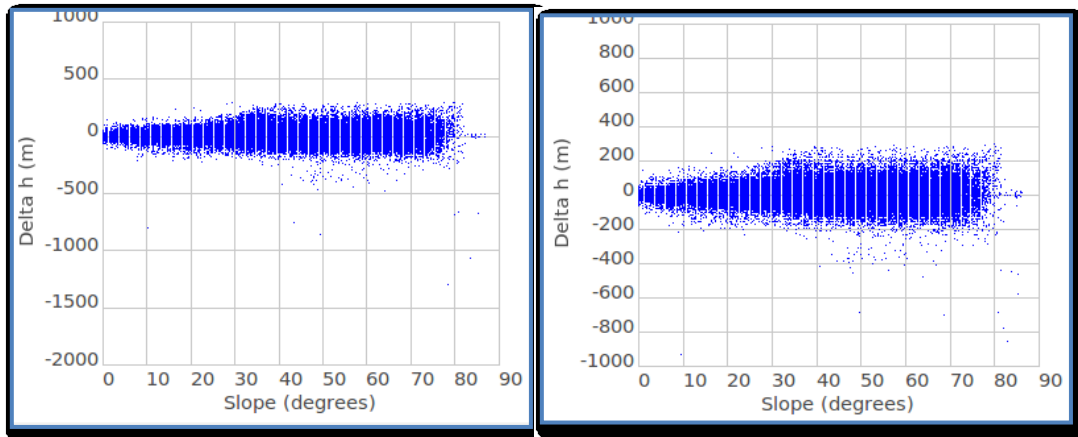
Figures do not illustrate a specific shape or tendency; however the correlation coefficient can be useful to confirm it.

r.covar -r map=Slope23_1sec_cubf@Basin, Diff2_ASTER_trenrcubf@Basin

Getting,

$$\begin{vmatrix} 1 & 0.03 \\ 0.03 & 1 \end{vmatrix}$$

The results obtained by the covariance matrix produce a correlation coefficient equal to 0.03 which indicates a null relationship between these two parameters.

9.3. Classification of differences with respect to Land Uses Classes

This additional step is done to verify if the differences between the LiDAR DSMs and the ASTER DSMs depend or better can be affected due to the land uses, the land uses map used is the same that in the Chapter 6 was rasterized and then transform in Chapter 8. The same five classes are used for the classification, which represent urban and non urban classes:

- Category 2 = Continuous urban fabric - Urban
- Category 3 = Discontinuous urban fabric - Urban

- Category 61 = Coniferous forests – Non urban
- Category 64 = Areas of natural pasture and mountain meadows – Non urban
- Category 72 = Bare rocks – Non urban

Once the Land Uses map is rasterized and transform, the classification process begins.

Example for Cat 61

```
r.mapcalc expression="Cat_61= if( usr_urb_tot@Basin==61 , Diff2_ASTER_trenrcubf@Basin, null() )"
```

This syntax just store in an output raster, for example Cat_61@Basin, the difference values, in this case_ASTER_trenrcubf@Basin, that belongs to land use category (usr_urb_tot@Basin) that is been considered and the values that do not fulfill the condition are filled with a null value.

This process is done for the five categories mentioned.

Once the differences are classified (Figure 91
Figure 92, Figure 93, Figure 94, Figure 95) some statistics tools are applied to the output maps Cat_61@Basin.

Table 21. Statistics of classification

Cat	No. Pixels	Min (m)	Max (m)	Mean (m)	Std Dev (m)	Description
2	1950	-33	54	1	9	Continuous urban fabric
3	27257	-925	49	-2	9	Discontinuous urban fabric
61	805201	-474	203	2	19	Areas of natural pasture and mountain meadows
64	232815	-203	143	-2	14	Coniferous forests
72	368990	-286	295	2	28	Bare rocks

Figure 91. Cat 2_diff_ASTER_Trent

Figure 91. Cat 2_diff_ASTER_Trent

Figure 92. Cat 3_ diff_ ASTER _Trent
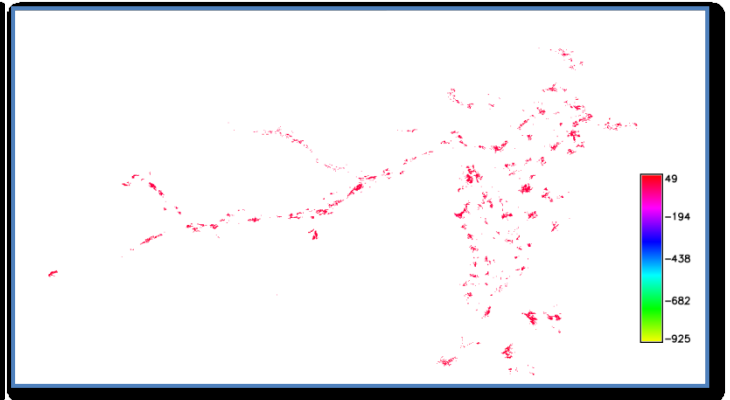
Figure 93. Cat 61_ diff_ ASTER _Trent
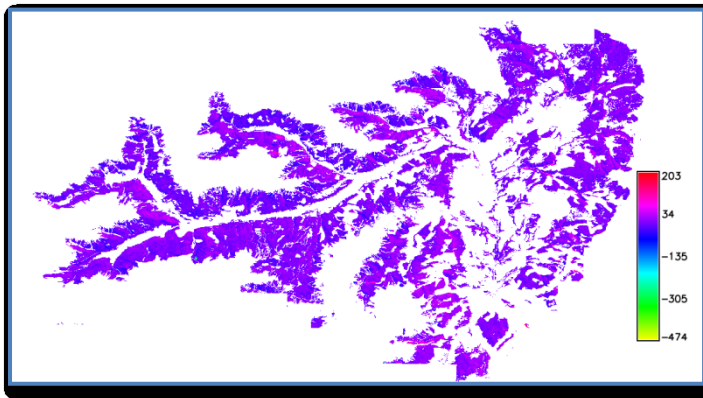
Figure 94. Cat 64_ diff_ ASTER _Trent
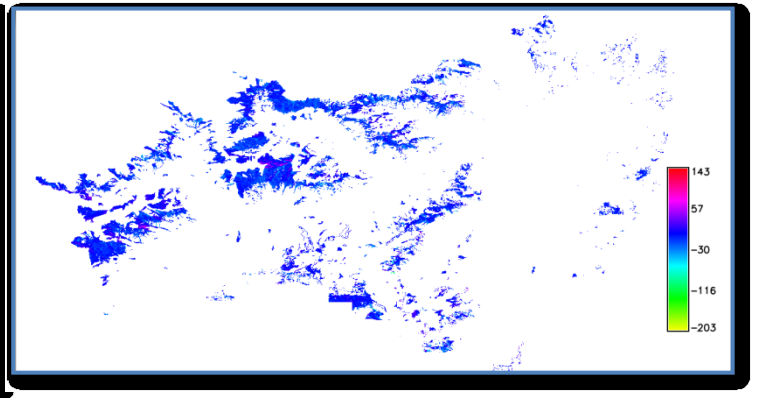
Figure 95. Cat 72_ diff_ ASTER _Trent
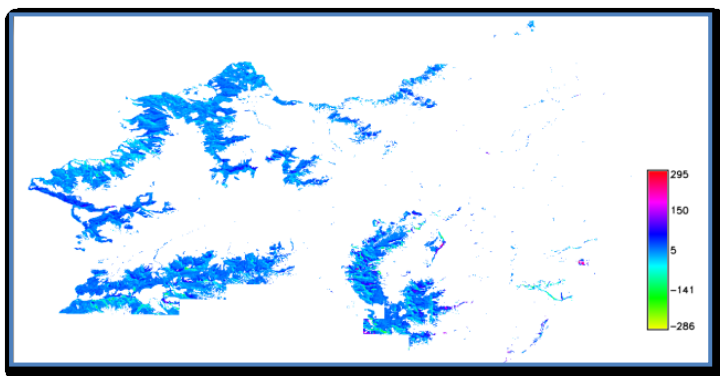
Regarding to the statistics, specifically the mean, is possible to conclude that as ASTER model has a higher resolution and a higher accuracy than SRTM or GMTED models then the errors between the LiDAR model and ASTER model are so bigger.

114

Chapter 10

10. Comparisons DSM local / DSM global GMTED2010

As the previous later chapter, here the upgraded set of data of the Val di Sol – Basin region is going to be employed to produce a final map that displays differences between the GMTED DEM with respect to the local DEM. Furthermore, it is calculated a correlation index between the height (h) and the variation of height between the two models (Δh).

10.1. Data source

The same process for importing data is required then, working from GlobalTrento-Location and the mapset Basin, the script mentioned in the section 8.1.2.is run with some specific characteristics using the following sintaxys.

Example for GMTED model,

GRASS 7.0. svn (TrentoGlobal):~/Scrivania/Thesis-Poli/Script > ./ trentino_demGMTED 275 301 0.002083333 nearest

For this example, the files than are going to be transform range from the 275 to 301, the resolution that is going to be reach is 0.002083333 degrees which means 7.5 seconds and finally the nearest method is the interpolation method used.

As is mentioned in section 8.1.2 , the 400 raster are transformed in subsets therefore when are obtained all the subsets, a merge function is needed trough the function r.patch.

At the end, the generated rasters are:

$$Trentinor\_7.5sec\_nearest@Basin$$
$$Trentinor\_7.5sec\_cubicf@Basin$$

The next figures present the mentioned maps; one is generated using the nearest method and the second the cubic_f method.

Figure 96. Trentino basin – 7.5 sec resolution – nearest method
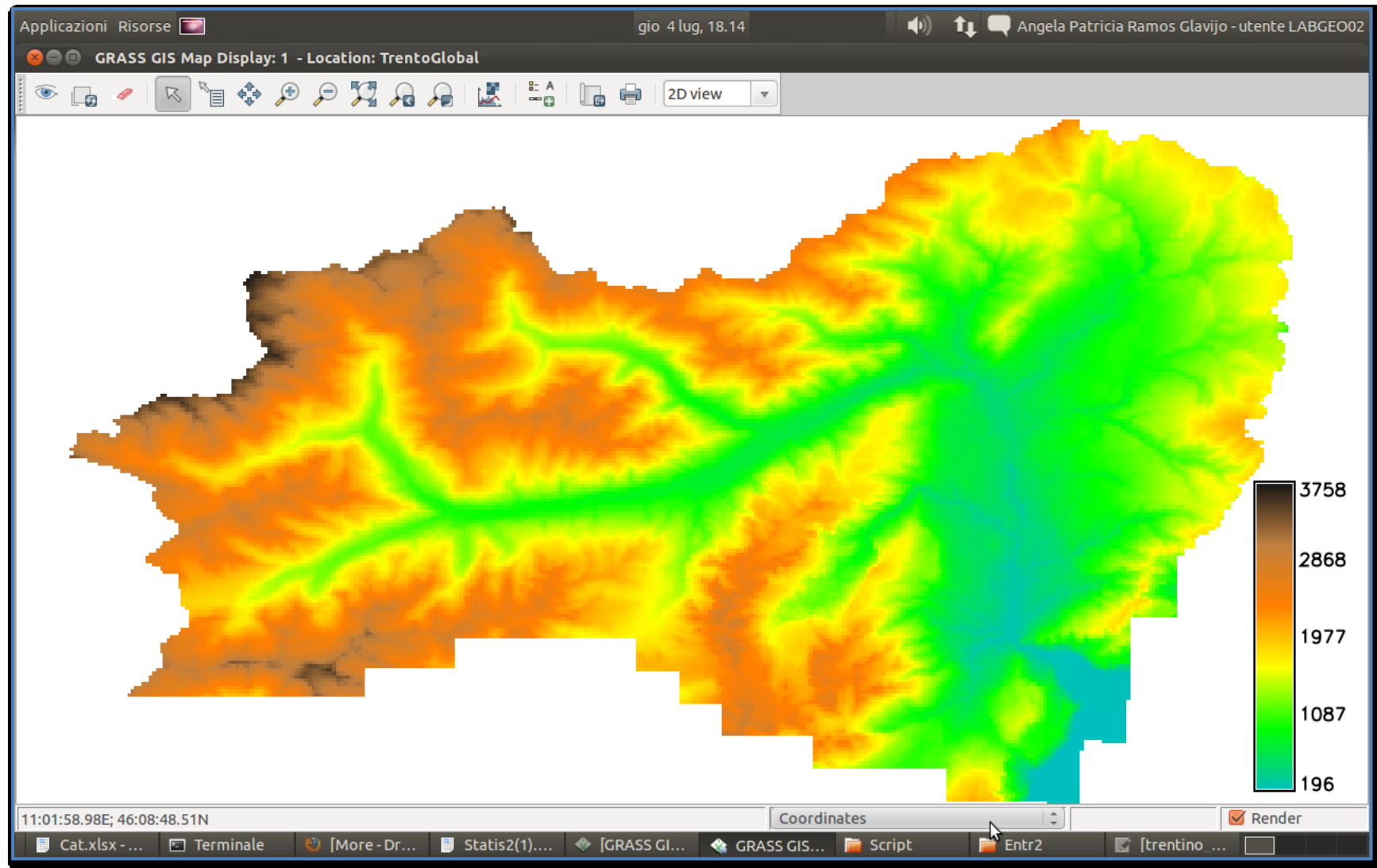
Figure 97. Trentino basin – 7.5 sec resolution – cubic_f method
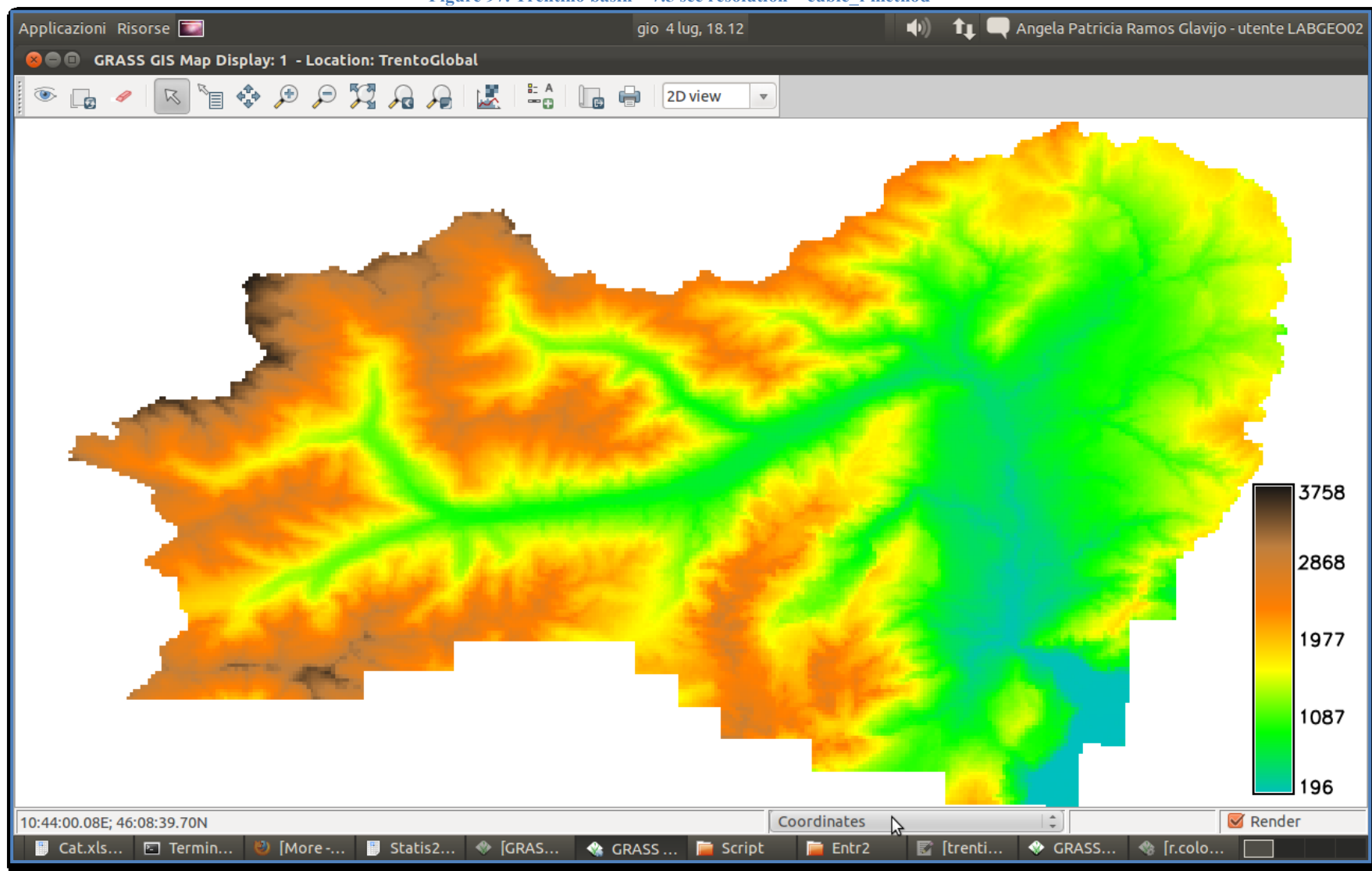
At first glance there is not much difference between the two methods; the variations can be detected, at the end of the process, just when the difference maps are done.

10.2. Data processing, analysis and final results

The main step is to run the subsequent command with its instructions:

r.mapcalc expression 'Diffr_GMTED_Trenear = GMTED - Trentinor_7.5sec_nearest'
r.mapcalc expression 'Diffr_GMTED_Trecubf = GMTED - Trentinor_7.5sec_cubicf'

Remark: As is mentioned in Chapter 2 and 7 GMTED has seven products, therefore just one of them, explicitly, the mean is used to perform the operation.

The respective maps obtained by these operations are shown below, but before print them, it is going to be presented the basic statistics of these products:

Table 22.Differences GMTED – Trentino-Basin_nearest

Table 23. Differences GMTED – Trentino-Basin_cubic_f

GMTED -Trentinor_7.5sec_nearest	
N. pixels	**39718**
Minimum (m)	-246.16
Maximum (m)	208,38
Range (m)	454,54
Mean (m)	0.012
Standard Deviation (m)	19.86

GMTED -Trentinor_7.5sec_cubic_f	
N. pixels	**39718**
Minimum (m)	-246.137
Maximum (m)	207.13
Range (m)	453,267
Mean (m)	0.018
Standard Deviation (m)	19.82

Also with the statistics, the histograms are presented

Figure 98. Histogram Differences GMTED – Trentino-Basin_nearest
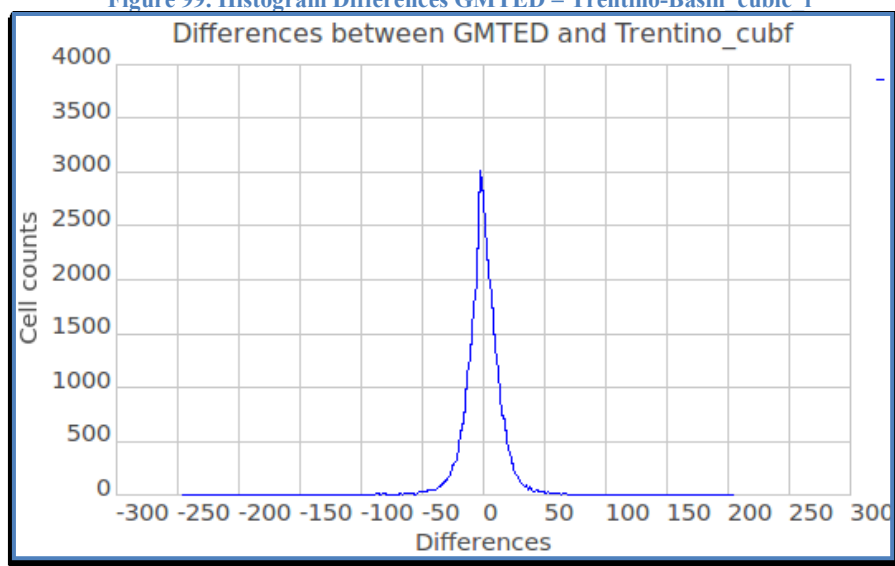
By analyzing these results, it can be said that the mean values are close to zero which means that the majority of the global DEMs are not so far from the local DEM, even though exist some outliers evidenced in the minimum and the maximum values. With respect to the standard deviation, it is taken as a parameter to construct the mentioned classification; this was explained in the chapter 8.

The results are the maps illustrate in Figure 101 and Figure 102

To complete this analysis an extra calculus is done, this corresponds to the difference between the differences produced by cubic_f method and the ones produced by the nearest method. The operation is performed through the command r.mapcal.

Example:

r.mapcalc expression= "Diff_GMTED_near_cubf= Diffr_GMTED_Trenear@Basin - Diffr_GMTED_Trecubf@Basin"

Once the differences are is obtained, the statistics tools are used to have a clear behavior of the results. This is done using the command r.univar.

r.univar map= Diff_GMTED_near_cubf@Basin

Then the statistics are presented in the following tables:

GMTED_cubf_near	
N. pixels	**39718**
Minimum (m)	-15 m
Maximum (m)	18 m
Range (m)	33 m
Mean (m)	0 m
Standard Deviation (m)	1 m

Table 24. Statistics of the differences between nearest diff and cubic_f diff

Also for this statistics, the respective histogram is shown

Figure 100. Histrogram of the differences between nearest diff and cubic_f diff
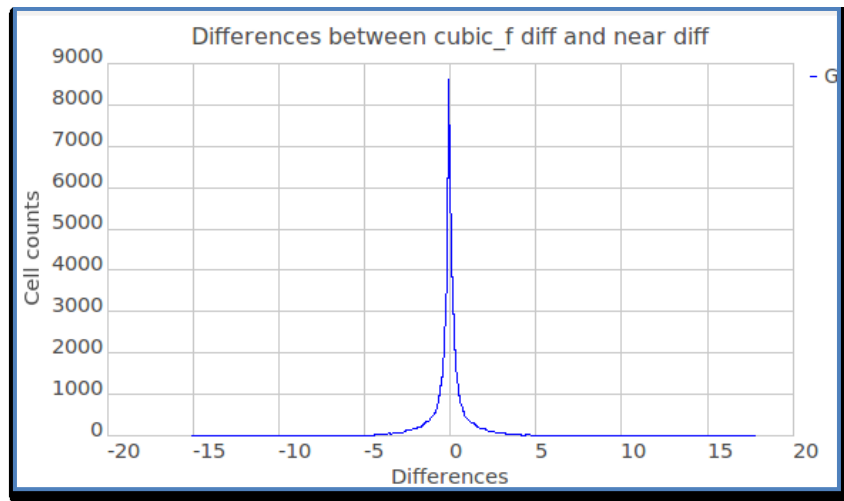
From the statistics and the histogram is observed that the methods either nearest or cubic_c are not so divergent because the mean is zero which means that most of the data do not have differences.

Figure 101.Differences of Trentino-Basin (Nearest method)

Color	Filter
	-19.86 < Differences map < 19.86
	-59.59 < Dm < -19.86 and 19.86 < Dm < 59.59
	-99.32 < Dm < -59.59 and 59.59 < Dm < 99.32
	-99.32 > Differences map > 99.32
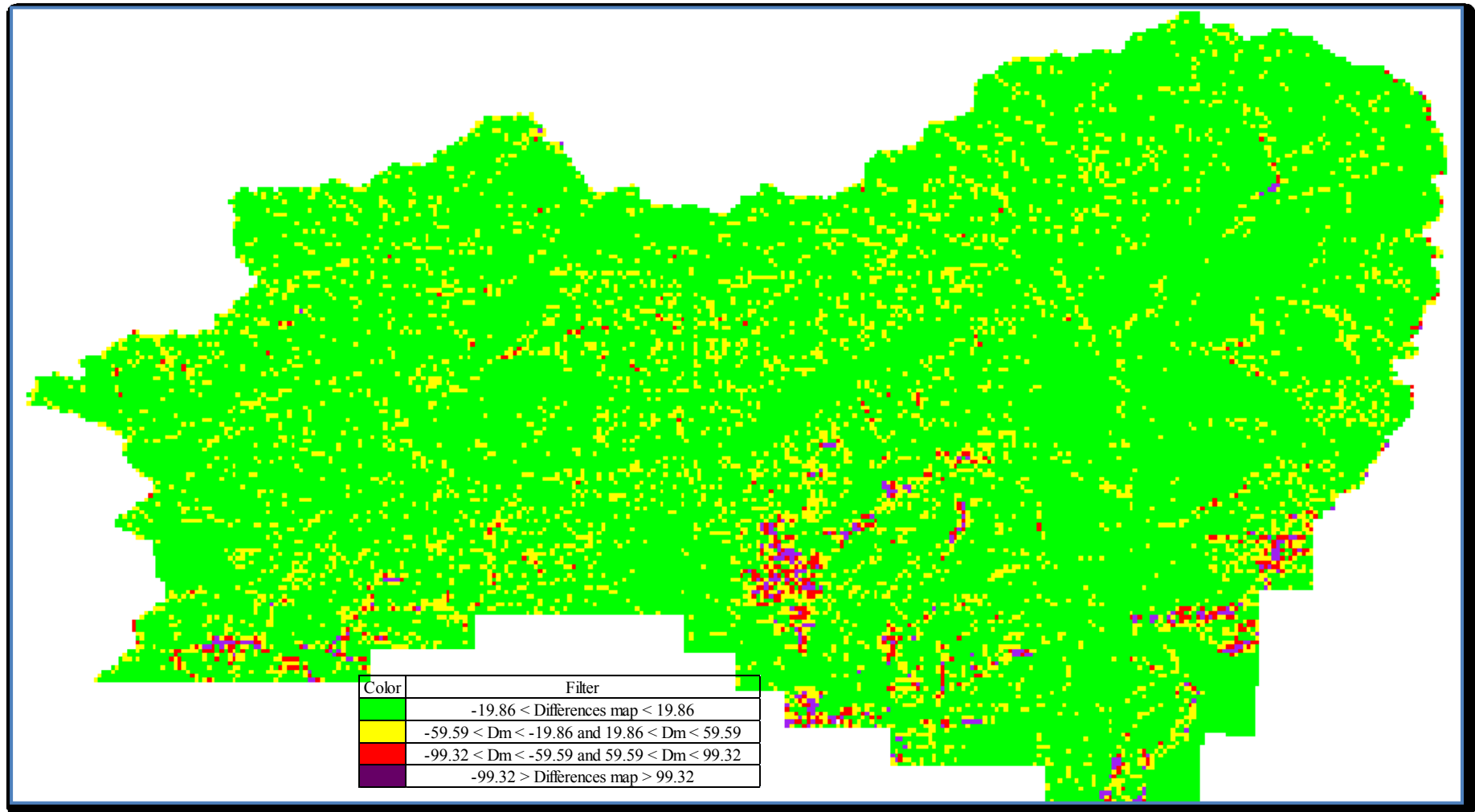
Figure 102. Differences of Trentino-Basin (Cubic_f method)

Color	Filter
	-19.82 < Differences map < 19.82
	-59.45 < Dm < -19.82 and 19.82 < Dm < 59.45
	-99.08 < Dm < -59.45 and 59.45 < Dm < 99.08
	-99.08 > Differences map > 99.08
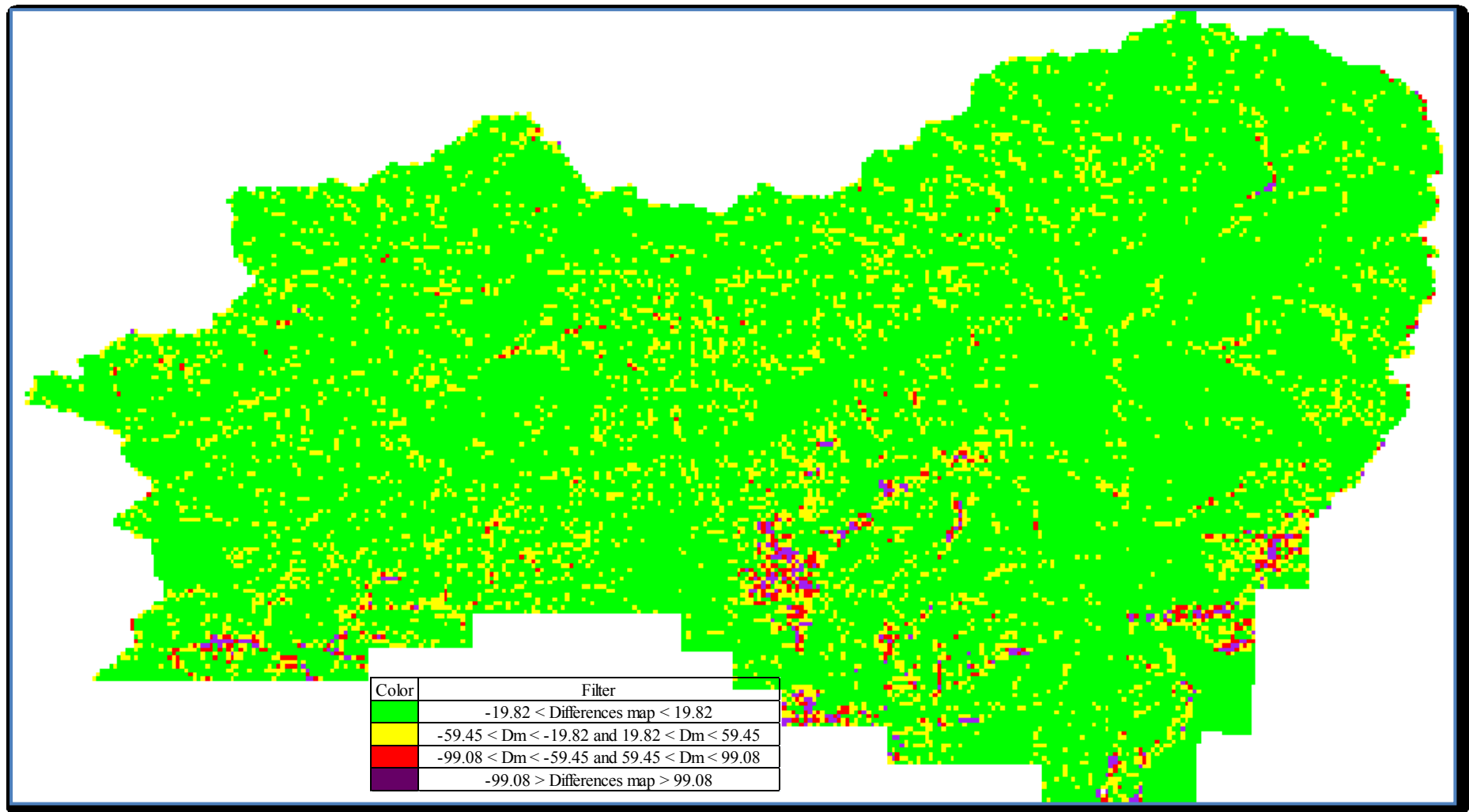
Figure 101 and Figure 102 evidence outliers that could be generated at the moment of the flight or in the data interpolation. The amount of outliers is greater that the once presented in the previous chapter but is less with respect to SRTM/LiDAR outliers due to the lower resolution of the data. Here the atypical points are also drawn with yellow, red and purple colors. These anomalies are located in the mountain areas where acquiring data process is not an easy task; while in plains the results are positively accepted.

To complete this general analysis, two new parameters are calculated:

- Correlation coefficient between Trentino elevations (h) and the differences GMTED/LiDAR (Δh): The idea of this coefficient is to verified if the differences between SRTM and LiDAR model depend on the terrain height. Otherwise another coefficient must be calculated.

  GRASS has two useful tools that allow visualizing the mentioned value; the first is Bivariate Scatterplot Tool which draws a point cloud of two variables, in this case h and Δh.

Figure 103.Bivariavate Scatterplot Trentino elevations (h) vs differences GMTED/LiDAR (Δh)
Near method (left) & Cubic_f method (right)

These plots indicate a certain trend for the majority of the points as well as few of them spread along the graphs. In order to give a number to this interpretation, the second tool of GRASS is used, namely, r. covar which returns the covariance matrix and implicitly the **correlation value**.

r.covar -r map=trentino_r2_7.5sec_cubf@Basin,Diff2_GMTED_trenrcubf@Basin

Finally, it is obtained the matrix

$$\begin{vmatrix} 1 & -0.14 \\ -0.14 & 1 \end{vmatrix}$$

The value -0.14 indicates an insignificant correlation, meaning that the differences between the two models do not depend on the height.

- Correlation coefficient between the slope (degrees) and the differences GMTED/LiDAR (∆h): With the calculation of this coefficient it is pretended to find a relationship inasmuch as, in some cases, the errors of the low accuracy DEMs do depend on the topography rather than the absolute heights. The first step is to produce the slope map through the "r.slope.aspect" GRASS command.

```
r.slope.aspect elevation=trentino_r2_7.5_cubf@Basin slope=Slope23_7.5sec_cubf format=degree
```

The following images are the outputs of the near and cubic_f method, respectively:

Figure 104. Slope (degrees) – Trentino_7.5sec_near

0	⬜
2	🟨
5	🟩
10	🟦
15	🟦
30	🟪
50	🟥
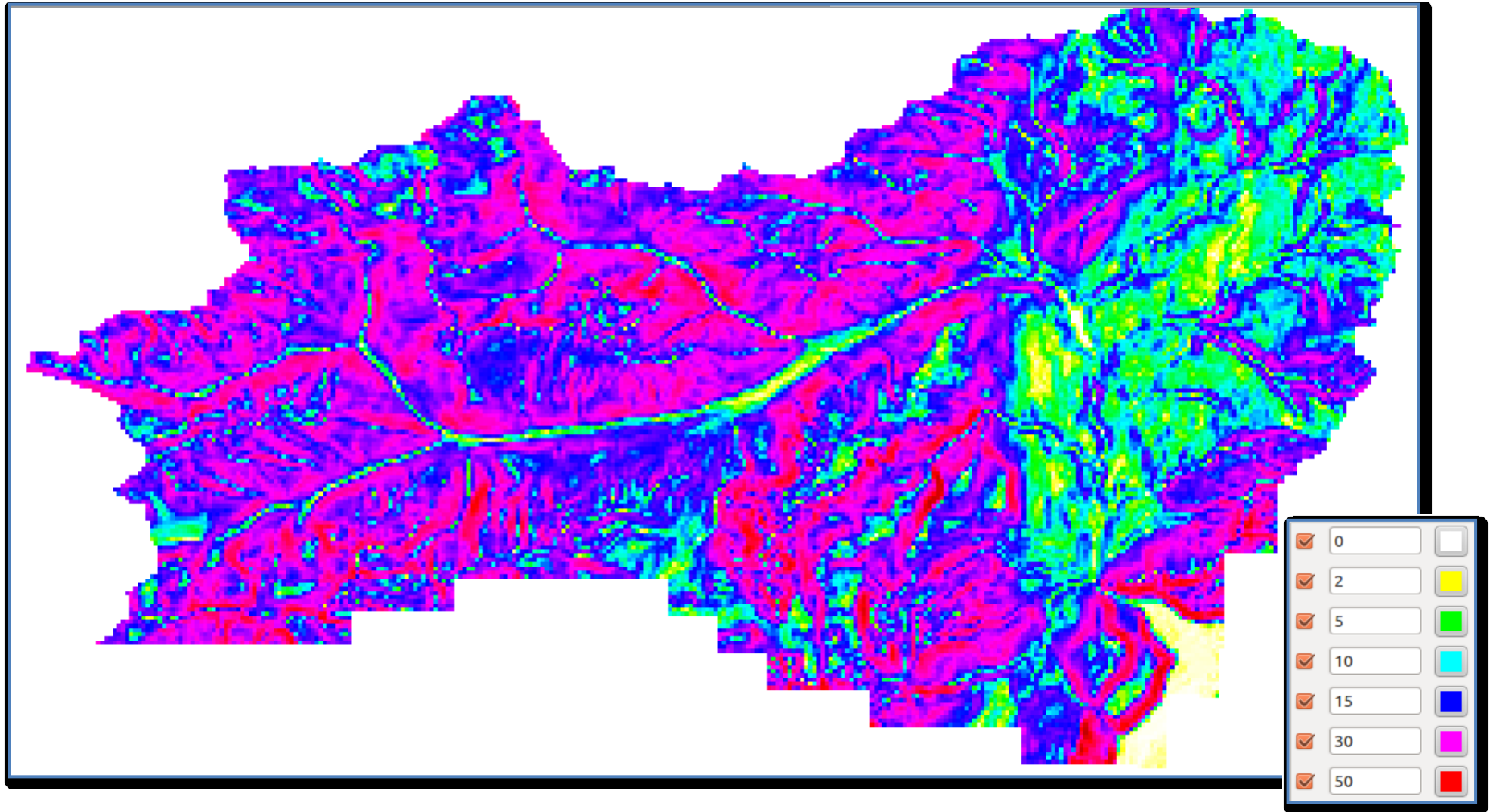
Figure 105. Slope (degrees). Trentino_7.5sec_cubic_f
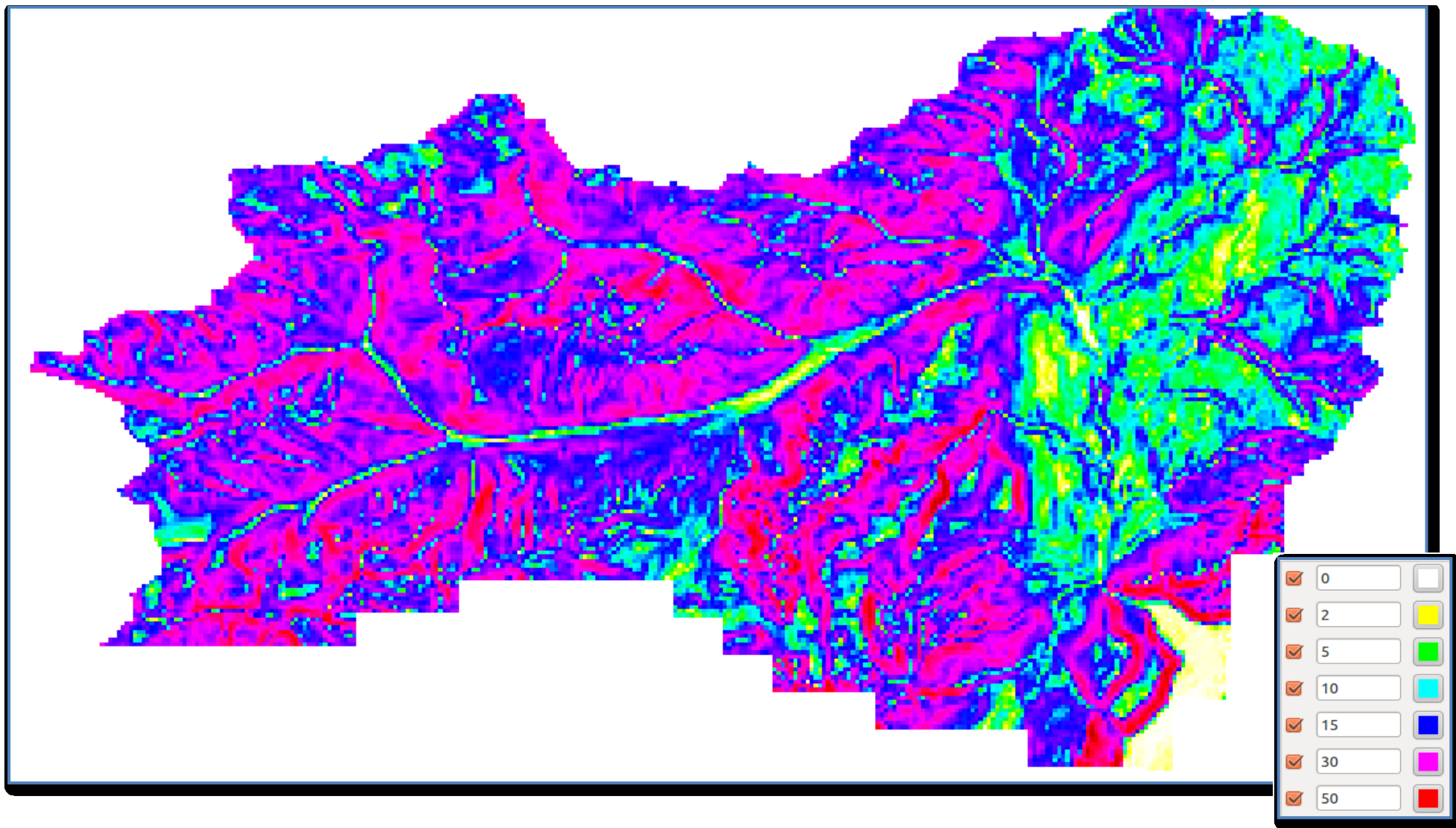
☑	0	⬜
☑	2	🟨
☑	5	🟩
☑	10	🟦
☑	15	🟦
☑	30	🟪
☑	50	🟥

The pink-purple color presents a terrain with a slope between 20 – 30 degrees which in fact belongs to a mountain region while yellow-green color marks flat areas between 0-10 degrees. Last but not least blue color shows a slope from 10-20 degrees approximately.

Once again Bivariate Scatterplot and r.covar are used and they produce the following results:

Figure 106. Bivariate Scatterplot Slope_7.5sec (degrees) vs differences GMTED/LiDAR (Δh) Near method (left) & Cubic_f method (right)
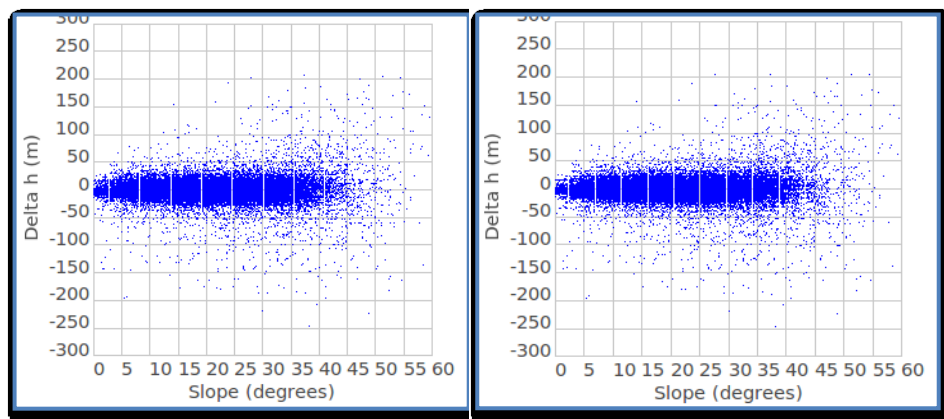
The figures do not illustrate a specific shape or tendency; however the correlation coefficient can be useful to confirm it.

r.covar -r map= Slope23_7.5sec_cubf@Basin, Diff2_GMTED_trenrcubf@Basin

Getting,

$$\begin{vmatrix} 1 & 0.04 \\ 0.04 & 1 \end{vmatrix}$$

The results obtained by the covariance matrix produce a correlation coefficient equal to 0.04 which indicates a null relationship between these two parameters.

10.3. Classification of differences with respect to Land Uses Classes

This additional step is done to verify if the differences between the LiDAR DSMs and the GMTED DSMs depend or better can be affected due to the land uses, the land uses map used is the same that in the Chapter 6 was rasterized and then transform in Chapter 8. The same five classes are used for the classification, which represent urban and non urban classes:

- Category 2 = Continuous urban fabric - Urban
- Category 3 = Discontinuous urban fabric - Urban
- Category 61 = Coniferous forests – Non urban
- Category 64 = Areas of natural pasture and mountain meadows – Non urban

- Category 72 = Bare rocks – Non urban

Once the Land Uses map is rasterized and transform, the classification process begins.

Example for Cat 72

```
r.mapcalc expression="Cat_72= if( usr_urb_tot@Basin==72 , Diff2_GMTED_trenrcubf@Basin, null() )"
```

This syntax just store in an output raster, for example Cat_61@Basin, the difference values, in this case_GMTED_trenrcubf@Basin, that belongs to land use category (usr_urb_tot@Basin) that is been considered and the values that do not fulfill the condition are filled with a null value.

This process is done for the five categories mentioned.

Once the differences are classified () some statistics tools are applied to the output maps Cat_72@Basin.

Table 25. Statistics of classification

Cat	No. Pixels	Min (m)	Max (m)	Mean (m)	Std Dev (m)	Description
2	35	-15	11	-2	6	Continuous urban fabric
3	498	-28	40	-2	7	Discontinuous urban fabric
61	14290	-177	190	2	13	Areas of natural pasture and mountain meadows
64	4168	-172	149	-2	15	Coniferous forests
72	6583	-246	207	-4	31	Bare rocks

Figure 107. Cat 2_diff_GMTED_Trent

Figure 108. Cat 3_ diff_ GMTED _Trent
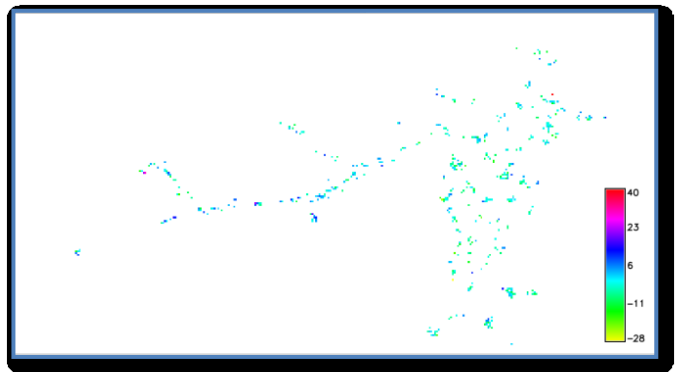
Figure 109. Cat 61_ diff_ GMTED _Trent
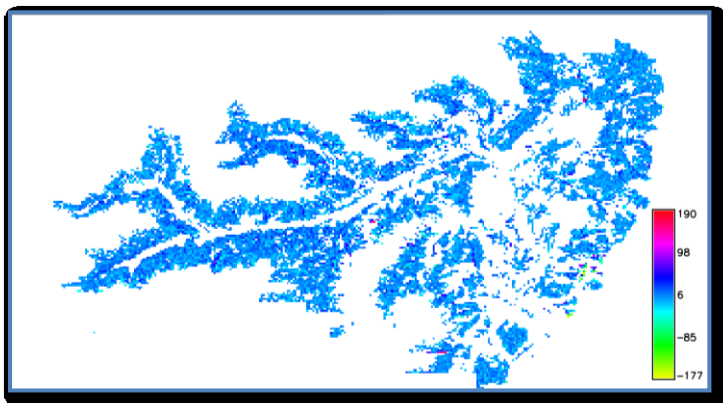
Figure 110. Cat 64_ diff_ GMTED _Trent
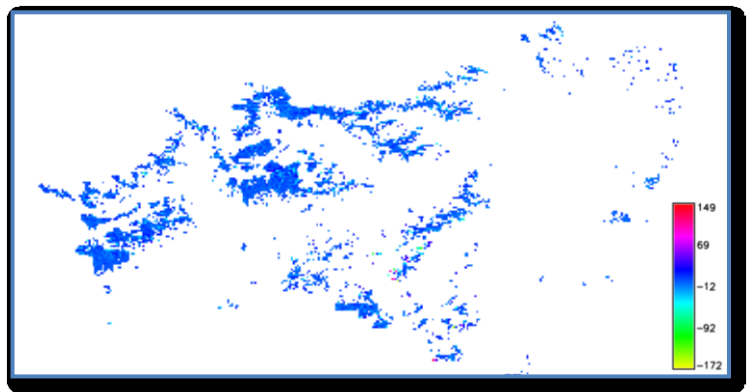
Figure 111. Cat 72_ diff_ GMTED _Trent
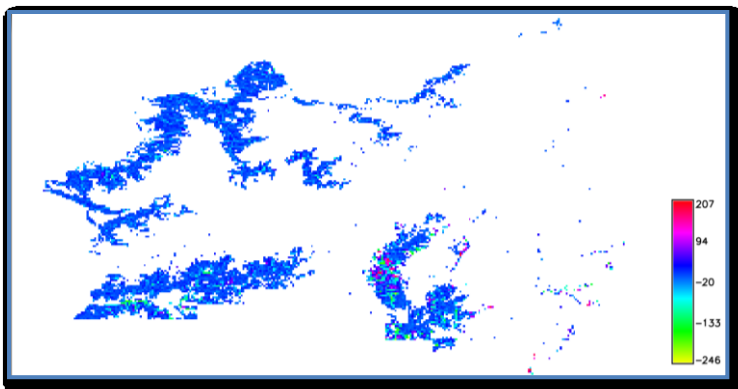
Regarding to the statistics, specifically the mean, seems that the model is quite good, but the real problem is that as the resolution is the biggest one between ASTER, SRTM and GMTED then some detail information is lost and then the outputs are not the expected.

Conclusions

The scope of the thesis is the study, comparison and problem identification of global Digital Elevation Models. In analysis of global models we decided to use LiDAR DSM/DTM of the Trento region as a benchmark data. They are characterized by high accuracy, available and free to be used: LiDAR data set is referenced in UTM zone 32N – WGS 84 .

More in particular, Val di Sole valley has been used as a case study which is part of the Trentino region. The Val di Sole has been chosen because here is possible to evaluate the mountain area and the valley area in terms of data acquisition (DSM-DTM), taking into account the information supplied by the Trentino Province as OpenData.

The first step, we collected and described all the data needed for this thesis:

Model	Resolution	Year of production
ASTER – global DEM	1 arc-second (≈30 m)	1999
SRTM – global DEM	3 arc-second (≈90 m)	2000
GMTED – global DEM	7.5 arc-second (≈250 m)	2010
LiDAR – local DEM	1 m	2006,2007,2008

Regarding to the calculation processes, the first comparison was done between the LiDAR DSM/DTM.

For the comparisons several scripts have been developed using GRASS GIS software because it has some important capabilities. <u>GRASS GIS</u>, commonly referred to as GRASS (Geographic Resources Analysis Support System), is a **free** Geographic Information System (GIS) software used for geospatial data management and analysis, image processing, graphics/maps production, spatial modeling, and visualization.

When we performed the local differences between DSM/DTM, we obtained two kind of situation. The first one is due to the natural characteristics of the terrain: vegetation, buildings, bridges, etc. The second one is due to some anomalies in DSM that are outliers, mainly caused by the interpolation method used to smooth the surface.

The second comparison concerned to local and global DSMs. In this case we are able to analyze a bigger area of Val di Sole. To do this operation the local DSMs have been transformed to the same resolution and coordinate system of global DSMs that are in Lat/Lon – WSG 84. Two interpolation methods were used: nearest and cubic_f. The nearest method, which performs a nearest neighbor assignment, is the fASTER than the cubic method. The cubic_f method determines the new value of the cell based on a weighted distance average of the 16 surrounding cells in the input map. *cubic_f* interpolation method can be used if thinning along null edges is not desired. This method "falls back" to simpler interpolation methods along null borders.

Regarding to the statistics results, the mean values are close to zero which means that the majority of the global DEMs have not biases with respect to the local DEM, even though exist some outliers. The standard deviation values are satisfactory since they are comparable with the values of the global DEMs.

To complete this general analysis we calculate the correlation between the differences and the heights and slopes, obtaining that no significant correlation exits.

Figures

Tables

Bibliography

AAVV. (2009). *Allegato Tecnico - Progetto HELI - DEM (Helvetia - Italy Digital Elevation Model)* .

ArcGIS Resources. (2013, March 6). Retrieved May 24, 2013, from What is Lidar?: http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_lidar/015w00000041000 000/

ASTER GDEM Readme File – ASTER GDEM Version 1. (2009).

ASTER Global DEM Validation. (2009).

Bamler, R. (1999). The SRTM Mission: A World-Wide 30m Resolution DEM from SAR Interferometry in 11 Days. 145-154.

Brovelli, M. A. (2012). GIS advanced functionalities: exploring data/field analysis., (p. 120). Como.

Carter, J., & others. (2012). *An Introduction to Lidar Technology, Data and Applications*. Charleston.

Danielson, J., & Gesch, D. (2011). *Global Multi-resolution Terrain Elevation Data 2010*. Virginia.

De Carolis, G. (2012). Active Microwave Instruments. Como.

De Carolis, G. (2012). From radiance to images: RS sensors characteristics., (p. 47). Como.

Dold, C. (2000). *Analyse und Implementierung eines Verfahrens zur Referenzierung geographischer Objekte*. London.

Farr, T. G., & others. (2007). *The Shuttle RadarTopography Mission*.

Geomatics Group. (2013). Retrieved May 24, 2013, from LIDAR: https://www.geomatics-group.co.uk/GeoCMS/Products/LIDAR.aspx

GIScience at SFU. (2003). Retrieved May 22, 2013, from Simon Fraser University: http://www.sfu.ca/gis/geog_x55/web355/icons/10_lec_web.pdf

GRASS GIS. (2012, November 13). Retrieved May 24, 2013, from General Overview: http://grass.osgeo.org/documentation/general-overview/

Hengl, T., & Evans, I. (2009). *Geomorphometry.*

Jet Propulsion Laboratory . (2004, Julio 9). Retrieved May 23, 2013, from ASTER Global Digital Elevation Map Announcement: http://ASTERweb.jpl.nasa.gov/gdem.asp

Jet Propulsion Laboratory. (2009, June 17). Retrieved May 23, 2013, from Shuttle Radar Topography Mission: http://www2.jpl.nasa.gov/srtm/

Knippers, R. (2009, August). *AGILE (Association of Geographic Information Laboratories in Europe)*. Retrieved July 6, 2013, from Geometric Aspects of Mapping: http://kartoweb.itc.nl/geometrics/Coordinate%20systems/coordsys.html

Knowles, K. (n.d.). *Geospatialmethods.org*. Retrieved July 6, 2013, from Points, Pixels, Grids, and Cells: http://geospatialmethods.org/documents/ppgc/ppgc.html

Lerma Garcia, J. L. (2002). *Fotogrametría Moderna: Analítica y Digital.* Valencia.

Li, S., Sun, H., & Yan, L. (2011). *A Filtering Method for Generating DTM based on Multiscale Mathematic Morphology.* Beijing.

Longley, P. A., Goodchild, M. F., Maguirre, D. J., & Rhind, D. W. (1999). *Geographical Information Systems.* Johm Willey & Sons, Inc.

Lucca, S. (2011). *Validation and Fusion of Digital Surface Models.* Como.

Lucca, S., & Luana, V. (2012). GRASS GIS intro. Como.

Neteler, M. (2000). *Introduction to GRASS.* Hannover.

Open Knowledge Foundation. (2012). Retrieved May 24, 2013, from Open Data: http://okfn.org/opendata/

OSGeo. (2013, July). Retrieved July 7, 2013, from About the Foundation: http://www.osgeo.org/content/foundation/about.html

Peucker, T. K., & Douglas, D. H. (1975). Detection of surface specific points by local parallel processing of discrete terrain elevation data. *Computer graphics and image processing* .

R2V. (1994). Retrieved May 22, 2013, from ABLE SOFTWARE CORP.: http://www.ablesw.com/r2v/rasvect.html

Schenk, T. (2005). *Introduction to Photogrammetry.* Ohio.

Strutture Dati. (n.d.). Retrieved May 23, 2013, from Triangulated Irregular Network: http://www.webalice.it/andrea.aime/lucidi/6%20-%20StruttureDati%20-%20parte%202.pdf
Team, A. G. (2009). *ASTER Global DEM Validation.*

United State Geological Survey. (2010, September 1). Retrieved May 23, 2013, from Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010): http://topotools.cr.usgs.gov/gmted_viewer/

USGS. (2012). Elevation Model Availability., (pp. 12-18).

Wikipedia. (2013, June 27). Retrieved May 24, 2013, from Provincia Autonoma di Trento: http://it.wikipedia.org/wiki/Provincia_autonoma_di_Trento

Wikipedia. (2013, June 11). Retrieved July 7, 2013, from Scots pine: http://en.wikipedia.org/wiki/Scots_pine

Wikipedia. (2013, July 2). Retrieved July 7, 2013, from Larch: http://en.wikipedia.org/wiki/Larch

Wikipedia. (2013, March 25). Retrieved July 7, 2013, from Abies Alba: http://en.wikipedia.org/wiki/Abies_alba

Wikipedia. (2013, June 20). Retrieved July 7, 2013, from Swiss Pine: http://en.wikipedia.org/wiki/Swiss_pine

Wilson, J. P. (2012). Digital Terrain Modeling. *Geomorphology* .