

# Caribbean Handbook on Risk Information Management



**GFDRR**  
Global Facility for Disaster Reduction and Recovery

**ACP-EU Natural Disaster Risk Reduction Program**  
An initiative of the African, Caribbean and Pacific Group, funded by the European Union and managed by GFDRR

[Home](#) > 6.3 Data homogenization

## 6.3 Data homogenization

**Author:** Michiel Damen

Data homogenization is the process bringing all data into a common geospatial framework to ensure consistency of data, integrity of analysis and validity of results. In a geospatial database, data are used from various sources; data that were captured using different technologies; and data published in different coordinate systems. All of which must be harmonized before using in a GIS. This can be vector data but also raster data from satellite images or a digital elevation model; also old scanned topographical paper maps with many unknown distortions can be used for historic change analysis.

Unfortunately not all people working with GIS data fully understand the collection methodologies of the (end) products, which can lead to unrealistic expectations, misuse of the data and acceptance of poor results. At the same time, the demands and expectations of geospatial information increases. Only if data is homogenized correctly then would spatial queries and analysis become practical and feasible.

### Geo-referencing and geo-coding of raster data

For a GIS analysis to be useful it is crucial that all datasets are referenced to the same spatial coordinate system and correctly overlay each other. This alignment of coordinate systems is referred to as geo-referencing. Geo-referencing may involve translations in the form of shifts, rotations, skewing and sometimes data transformations using non-linear methodologies. The term Geocoding is used when not only the coordinate system, but the whole row-column system of for instance a satellite image is transformed. For vector data is in most cases geo-referencing sufficient. But to rotate the view of a satellite image towards the North, also geocoding is required.



Figure 1: Illustrating the problems geo-referencing a remote sensing image.

The most recent remote sensing images from satellites come with tie-point file data from orbital parameters and other parameters to facilitate geo-referencing and ge-coding. This data are stored in separate files to be used in most modern image processing software programs. The WorldView-3 satellite sensor with a spatial resolution of only 30 cm claims a horizontal geolocation accuracy of less than 3.5 meters without ground control. For the latest Landsat-8 this is 12.0 meters; for the older Landsats this value will be much higher.

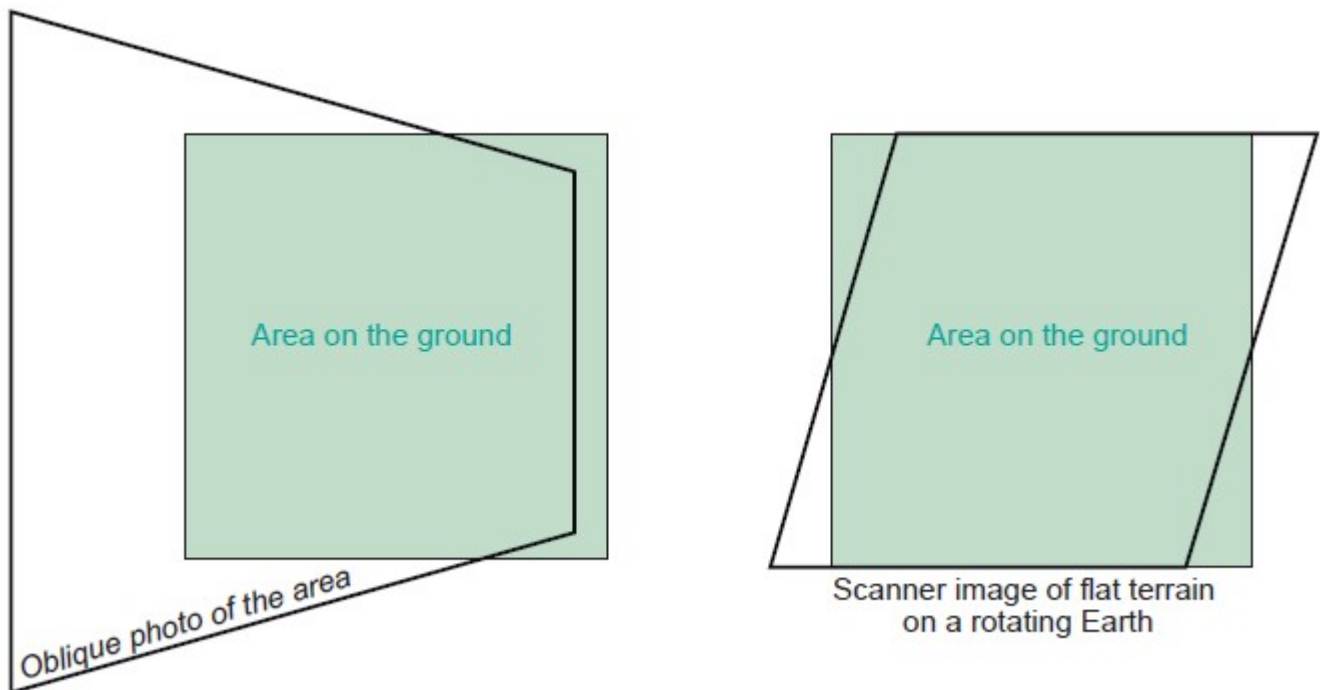


Figure 2: Examples of geometric image distortions. (1) Effect of oblique viewing; (2) Effect of Earth Rotation and (3) "relief displacement".

Therefore it is advised to use accurate ground control points for a maximum geolocation accuracy and apply image-to-image rectification of the oldest images using the more accurate recent ones. It is even better to use a sufficient

accurate digital elevation model (DEM) of the same area as the satellite images, so that tie-points can be found in both a horizontal X, Y and vertical Z direction. It will be clear that also such a DEM should be geometrically corrected, for instance with the help of a topographical map with elevation information. So has the very popular Shuttle Radar Topography Mission (SRTM) for instance a horizontal geolocation error of at least 5-6 meters and a vertical error of 15 meters or more.

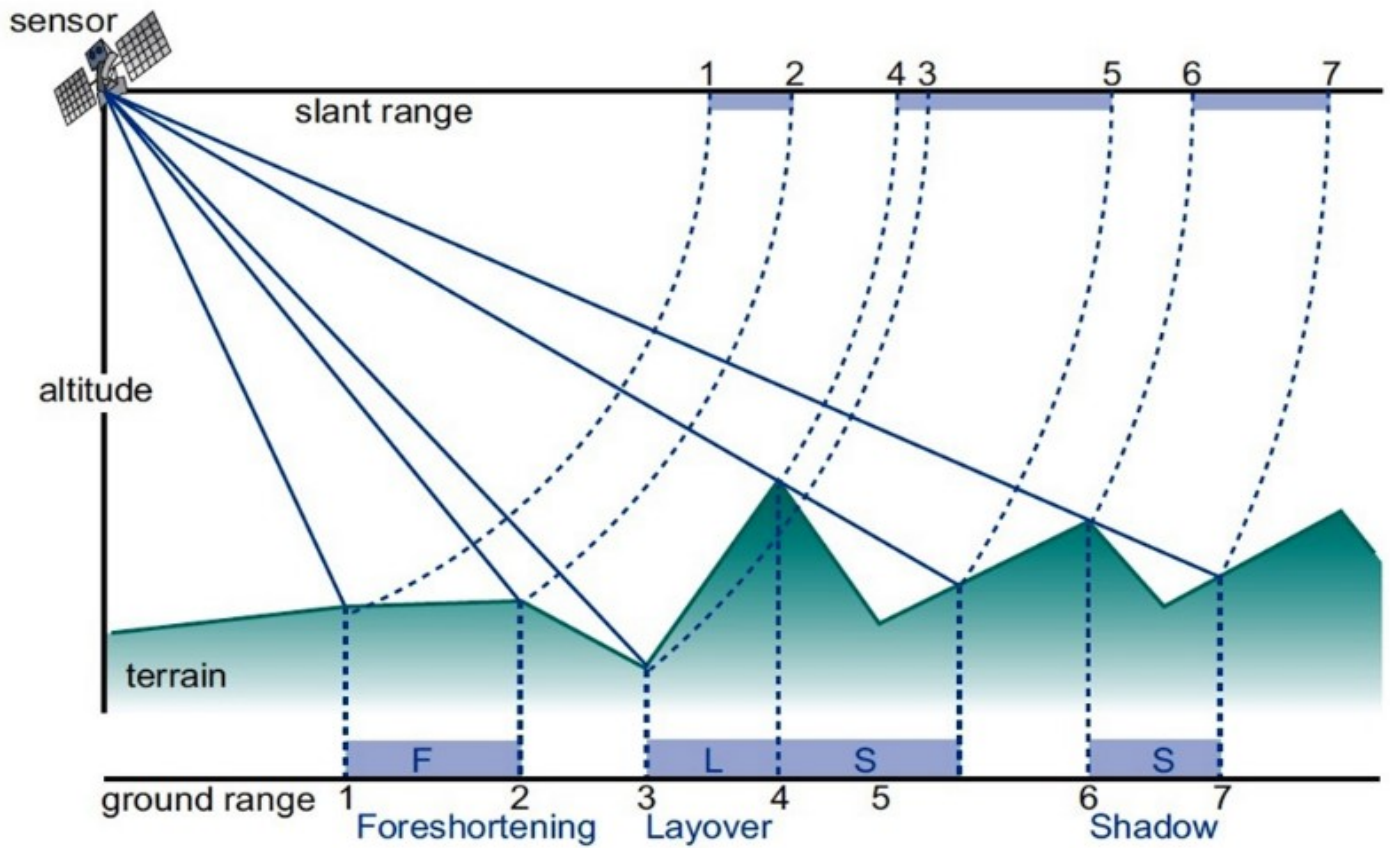


Figure 3. Radar images have specific distortions due to its side-looking viewing geometry.

Foreshortening: Slope facing radar is compressed

Layover: If the radar beam reaches the top of a slope earlier than the bottom

Shadow: No illumination of slopes facing away from the sensor.

GeoReference Editor: Test\_Import\_GDAL

File Edit Options Help

Third Order    Sigma = 1.294 pixels

Global tools  
Test\_Import\_GDAL  
Background

Continuous

Coordinate 39°  
Test\_Import 45  
Test\_Import 512,

	X	Y	Row	Col	Active	DRow	DCol
9	18.754	40.058	0.00	10560.00	True	0.01	-1.80
10	18.907	40.079	0.00	11880.00	True	0.01	-1.79
11	19.059	40.100	0.00	13200.00	True	0.01	-1.77
12	19.212	40.120	0.00	14520.00	True	0.01	-1.76
13	19.364	40.140	0.00	15840.00	True	0.01	-1.76
14	19.517	40.160	0.00	17160.00	True	0.01	-1.75
15	19.670	40.180	0.00	18480.00	True	0.01	-1.76
...	...	...	...	...	...	...	...

Customize the Pixel Information Window    512,2862    39°58'45.8"N, 17°51'21.91"E    39°58'45.82"N, 17°51'21.91"E

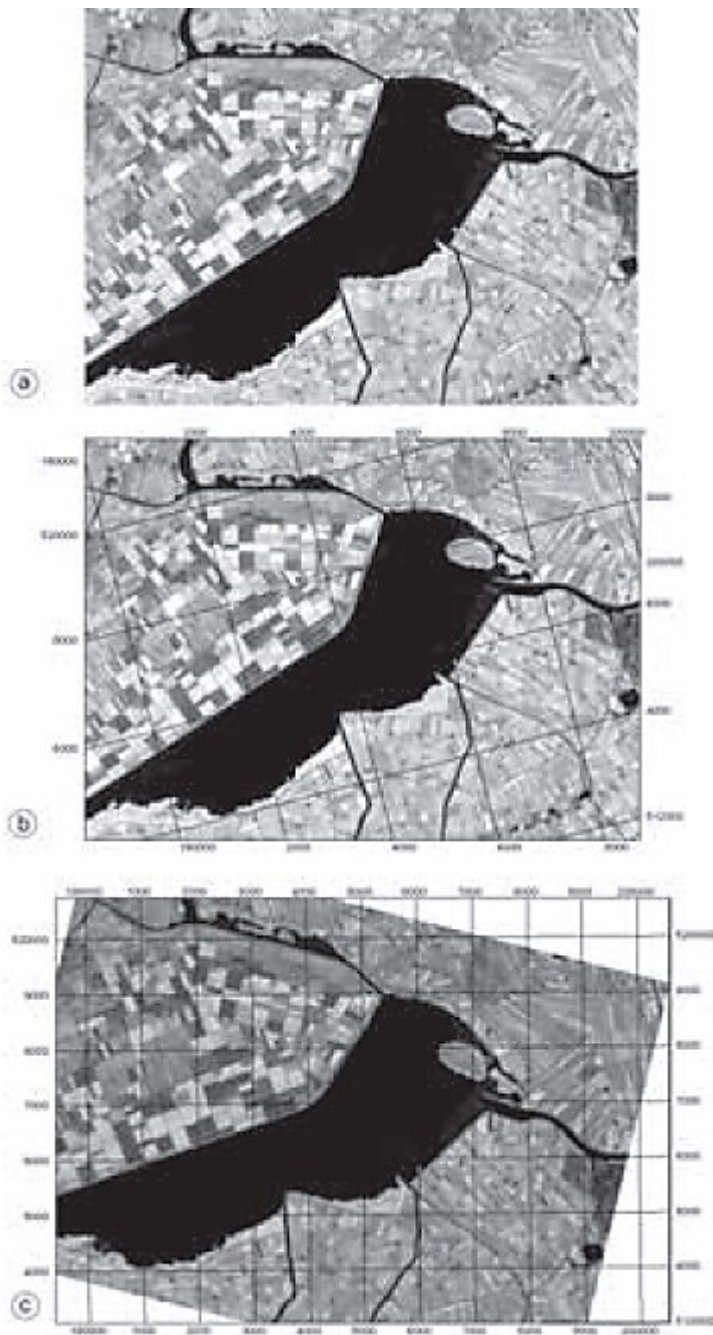


Figure 4 (LEFT) Example of Tie-point file for geometric correction and geocoding of Sentinel-1A radar data

Figure 5 (RIGHT) Top image original; middle image: georeferenced; bottom image: geo-coded.

Some extra information is given by following the following links:

[INTRODUCTION](#)

## THREE DIMENSIONAL APPROACHES

## GEOREFERENCING AND RESAMPLING

## GEOMETRIC TRANSFORMATIONS

### Geo-referencing of vector data

Also for vector data, in many cases geo-referencing is needed. This can include data translations to remove shifts, rotations, skewing and data transformations of coordinate systems. Geo-coding is however in most cases not necessary, as all professional software (for instance ArcGIS and QGIS) are capable to display the vector data in an orientation towards the North.

In the Use Case Book are in [Chapter 9.1](#) on Homogenization of base data three very illustrative examples given of the removal of horizontal data shifts between satellite raster data and vectorised base data. It is advised to visit this chapter for further information on this subject. Examples: (1) Base data from Dominica including a road layer and 2014 Pleiades satellite image; (2) base data from St. Lucia including a coastline dataset and satellite image; and (3) Base data from St. Vincent including two building maps & satellite image.



Figure 6: Mismatch of digitized roads (red lines) and high resolution Pleiades satellite image of 2014. Example from West Dominica.

In the example of Figure 6 there exist not only a mismatch between the roads and the satellite image, but some roads are clearly also outdated.

It is advised to first analyse of all provided data layers their geographic accuracy. After this one should take the most accurate dataset as a basis to geo-rectify the other data layers. This can be done by (1) re-projecting the data to one common base and after this perform additional image-to-image rectification and re-digitizing / updating if this is needed.

For the example of Figure 6 this could be to (1) measure extra accurate tie-point with differential GPS in the field, to geo-rectify the satellite image as good as possible, and (2) to perform image-to-image rectification of the roads using an image interpretation of the rectified image. Also the roadmap should be updated using this image.

### Software for data homogenization

The selection of the software to be used to process vector and raster data, including the geographic re-projection from one datum to the other, should include the following considerations:

#### **The data type to be processed.**

It is advised to use professional software for the import and processing of high resolution satellite imagery. This has to do with the import and use of the additional tie-point and other data files about the positional accuracy of the satellite during time of acquisition. Import of the data file in only GeoTiff format or HDF will not always be sufficient to acquire the most accurate geo-positional accuracy. These software have also modules for the use of 3D point data for ortho-

rectification (for instance from differential GPS or high resolution DEM's). They can also process LiDAR point data and RADAR Imagery with additional modules. Examples of such software are [ERDAS Imagine](#) and [ENVI](#). Please note that a licence of this software is costly. Open source remote sensing software which is freely available can also be an option. Examples are SAGA GIS, GRASS and [ILWIS](#). Click [HERE](#) for an overview.



For the import and processing of vector data [ArcGIS](#) from ENVI might be a good choice. This software can import and manage the already pre-processed satellite imagery without difficulties, together with the vector data. It is also capable to reproject vector data together with raster data image-to-image. Also here the licence costs can be a limiting factor. A very good alternative can be the free installation of [Quantum GIS \(QGIS\)](#). This software has most of the options of ArcGIS, has dozens of plugins available, also plugins for satellite image processing such as GRASS. To read the differences between ArcGIS and QGIS click [HERE](#).

#### **Software knowledge of local GIS and Remote sensing experts.**

Of course every expert has its own preferences, in most cases based on educational background and working experience. It is however advised to consider for new projects also open source software, such as ILWIS, GRASS and QGIS. This kind of software has no licence limitations and can be installed on any system in the office or at home. Unexperienced new staff can train itself easily in an informal way.

#### **Costs of the data**

High resolution satellite imagery (less than 1 m. resolution) is expensive. It mostly depends on the deals made with the National Mapping Agency and the Satellite Data provider which data can be used. Of course Google Earth imagery can be an alternative, but keep in mind that this data is "only a picture" without spectral bands to process for instance land use/land cover and flooded areas with supervised classification. Another option is the use of (stereo) aerial photography.

---

**Source URL:** <http://www.charim.net/datamanagement/63>