2.4 Fundamentals of GHSL

The GHSL consists of three main information components hierarchically placed at three different levels of abstraction: Global Human Settlement built-up areas (GHS-BU), the GHS population grids (GHS-POP) and the GHS urban/rural classification model (GHS-SMOD).

At the base of the hierarchy - including the most spatially accurate and the least abstract information level - we have a layer collecting concrete evidences about the human presence on the planetary surface as seen from global Earth Observation systems. In the GHSL paradigm, the fundamental link between Earth Observation sensor data and the human presence is the observable presence of built-up structures or buildings. From the GHSL perspective, the “building” makes the physical part of the human settlement fabric or spatial extension that is observable and measurable using the available global sensors. At this basic level the GHSL reports about built-up areas (GHS-BU), as areas (spatial units) where buildings can be found {Pesaresi_al_2013}. The concept of “buildings” formalized by the GHSL are enclosed constructions above ground which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure constructed or erected on its site {ref: Pesaresi_al_2013}. This abstraction is very similar to the standard topographic definition of the “building” class as compiled in the INSPIRE directive13, except for the fact that the condition of the permanency of the structure it is not in the GHSL definition. This fact allows to include also refugee camps, informal settlements, slums and other temporary settlements and shelters in the notion of built-up area in the GHSL concept.

The intermediate abstraction information layer of the GHSL is the population grid or GHS-POP that is produced in an in-between spatial resolution. This information layer is derived from the combination of global collections of national population census data and global built-up areas as extracted from Earth Observation data analytics (GHS-BU). In the approach taken by the GHSL, the population data collected by national censuses with heterogeneous criteria and heterogeneous update time are harmonized in the space and time domains in to the GHS-POP grids, by systematic and consistent application of the same set of data interpolation and spatial disaggregation methods to the best available global spatial baseline data {Freire_al_2016}.

The top abstraction information layer of the GHSL it is the urban/rural classification model (GHS-SMOD). It is provided with the least spatial detail (1 km) by combining the two less-abstract and more-spatially-detailed built-up and population grids, GHS-BU and GHS-POP, respectively. The GHS-SMOD model implemented by the GHSL it is consistent with the "Degree of urbanisation" (DEGURBA) model adopted by EUROSTAT14. It discriminates 3 settlement class abstractions: 1) Cities, 2) Towns and suburbs and 3) Rural areas. The discrimination is based on the population density in the square kilometre grid15, total settlement population and other spatial generalization parameters.

In the GHSL paradigm, the base layer GHS-BU it is designed to be the most stable against different visions and approaches, while GHS-SMOD is the most abstract and as such exposed to conceptual changes and alternative problem settings proposed by the different stakeholders involved in the post-2015 international framework processes. The modular hierarchical abstraction schema used in the GHSL design allows to protect the investment made in the global, fine-scale information gathering from perturbations on the abstract classification schema that may be introduced by different decision-makers involved in the process and potentially producing different problem setting and abstractions. On the other side, the modular hierarchical abstraction schema facilitates the test of alternative abstract

---

14 http://ec.europa.eu/eurostat/web/degree-of-urbanisation/overview
15 densely, intermediate density and thinly populated areas
models on the same agreed information baseline, facilitating the discussion and the comparison of the results also between international stakeholders not necessary sharing the same high abstraction definitions.

The following section helps the reader to understand fundamental concepts of GHSL and its data. The first subparagraph deals with extraction of information from satellite imagery (2.4.1) and built-up definition.

The second paragraph explore the process allows to combine built-up grids with census data to produce the population grids (2.4.2).

The third paragraph (2.4.3) illustrates the key elements and rules of the settlement model, derived from the New Degree of Urbanization (Lewis Dijkstra and Hugo Poelman 2014): specifically, the rules for defining Urban Centres, Urban Clusters and rural settlements are illustrated.

The forth paragraphs show with simple images, and example of three GHSL datasets (GHS Built-up, GHS POP and S-MOD) for the city of Madrid, Spain (2.4.4).
2.4.1 From Earth’s surface to built-up area

Built up area is typically expressed with a continuous values representing the proportion of building footprint area within the total size of the cell.

Built-up extraction

Human settlement are characterized by constructed, man-made objects - that include buildings and associated structures and civil works. For settlement analysis, the location and spatial size of the building surface area - referred as building footprint area - is modelled into built up areas.

Satellite imagery

A satellite image is a raster file which represents Earth’s surface. In order to be used to obtain useful information about urban settlements, many steps have to be done, such as: ortho-rectification, georeferencing, spectral calibration and radiometric corrections.

Earth Surface

Earth observation satellites regularly provide images of its surface. These images have different resolution and characteristics.
2.4.2 From Built-up area to population grid

**INPUT**

GHS BUILT-UP

GSH built-up uses small grid cells to measure human settlements regardless of administrative boundaries.

Population censuses provide accurate information on the characteristics and number of residents for administrative or finer numeration areas (census tracts).

These data sets are typically available as a total count for units varying widely in size and shape, while frequently residents occupy only specific zones of these units, at different densities.

**METHOD**

The GHSL method is design to combine information from population censuses with built-up and to downscale population into a grid of 1Km of resolution, according to the presence or absence of built-up in the grid cell.

**OUTPUT**

GHS POP

The combined information result into a new layer (resolution 1Km) which disregards administrative boundaries, and represents the presence and density of population. In the GHS pop grid, the grid cell value represents the absolute number of inhabitants.
2.4.3 The GHSL Settlement Model

**URBAN CENTRE**
- ** Required conditions **
  - EACH GRID CELL
    - 1 Km
    - \( >1.500 \) minimum of 1,500 inhabitants
    - 1 Km
    - Built-up \( >50\% \) density of built-up greater than 50%
  - AND
  - TOTAL POPULATION \( > 50,000 \) INHABITANTS
  - contiguous grid cells (4-connectivity, gap filling) with minimum population of 50,000 inhabitants

**URBAN CLUSTER**
- ** Required conditions **
  - EACH GRID CELL
    - 1 Km
    - \( >300 \) minimum of 300 inhabitants
  - AND
  - TOTAL POPULATION \( > 5,000 \) INHABITANTS
  - contiguous grid cells (8-connectivity) with minimum population of 5,000 inhabitants

**RURAL**
- ** Required conditions **
  - EACH GRID CELL
    - 1 Km
    - \( >0 \) cell with inhabitants
  - AND
  - single or contiguous grid cells with total population of less than 5,000 inhabitants
2.4.4 An example from the city of Madrid, Spain

The image on the left is a satellite image of the city of Madrid, Spain in 2015. The overlapped administrative boundaries (in blue) show their differences in size and unevenness of borders.

Built-up (resolution 38m)
Built up area is typically expressed with a continuous values representing the proportion of building footprint area within the total size of the cell.

The value of the cells in this area are significantly different, from 0 to 98.

Population grids (resolution 250m)
In the population grid, grid cell value represents the number of inhabitants.

In this specific area, the number of inhabitants varies from 3 to about 12,000 per sqKm.

Settlement Model (resolution 1Km)
The GHS S-MOD aims at classifying human settlements according to certain rules of population and built-up density and contiguity of grid cells.

In the example on the left, the urban centre of Madrid, with relative urban clusters and rural settlements.