European Commission

JRC TECHNICAL REPORT

Applying the Degree of Urbanisation Method to National Population and Housing Census with GHSL Tools

The example of Colombia

Melchiorri Michele, Schiavina Marcello,
Freire Sergio, Maffenini Luca, Kemper Thomas,
Dijkstra Lewis

2021

Joint Research Centre

EUR 30692 EN

Contact information
Name: Thomas Kemper
Address: Via Fermi, 2749 21027 ISPRA (VA) - Italy - TP 130
European Commission - DG Joint Research Centre
Space, Security and Migration Directorate
Disaster Risk Management Unit E.1
Email: thomas.kemper@ec.europa.eu
Tel.: +39 0332 78 55 76

EU Science Hub
https://ec.europa.eu/jrc

JRC124765

EUR 30692 EN

PDF ISBN 978-92-76-37556-2 ISSN 1831-9424 doi:10.2760/939289

Luxembourg: Publications Office of the European Union, 2021

How to cite this report: Melchiorri Michele, Schiavina Marcello, Freire Sergio, Maffenini Luca, Kemper Thomas, and Dijkstra Lewis, Applying the Degree of Urbanisation Method to National Population and Housing Census with GHSL Tools, EUR 30692 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-37556-2, doi:10.2760/939289, JRC124765.

Contents

Authors

Michele Melchiorri[2], Marcello Schiavina[1], Sergio Freire[1], Luca Maffenini[3], Thomas Kemper[1] and Lewis Dijkstra[4]

[1] European Commission-Joint Research Centre, Ispra, Italy, [2] Engineering S.p.A, [3] UniSystems Luxembourg SàRL

[4] European Commission Directorate General for Urban and Regional Policy, Bruxelles, Belgium.

Abstract

The Degree of Urbanisation is a method to delineate urban and rural areas for international statistical comparison. It relies on geospatial data that are generally available in National Statistical Offices and can be implemented with open and free data and tools. This report explains the application of the Degree of Urbanisation method to the 2018 edition of the National Census of Colombia, using tools produced in the framework of the Global Human Settlement Layer project of the European Commission Directorate General Joint Research Centre. The Degree of Urbanisation is applied in a two-stage process to a population grid (obtained by combining the 2018 Colombian census data with a built-up surface layer extracted from Copernicus Sentinel-2 imagery): first grid cells are classified based on population size and density criteria, then small local administrative units are classified into cities, towns and semi-dense areas, and rural areas. The completeness and open data policy of the 2018 census of Colombia allows to classify small territorial units, and to disaggregate six demographic and housing variables by Degree of Urbanisation. These include sex ratio and access to infrastructure, like connectivity to internet and access to electricity among others. This technical report contains practical guidance for the application of the Degree of Urbanisation method to national demographic and housing census/surveys and complements the EUROSTAT manual on how to apply the Degree of Urbanisation, produced at the request of the 51st Session of the United Nations Statistical Commission.

Executive Summary

In the past few years the European Commission led a voluntary commitment of six International Organisations and agencies – the Food and Agriculture Organization of the United Nations, the United Nations Human Settlements Programme, the International Labour Organization, the Organisation for Economic Co-operation and Development and The World Bank – to develop a harmonised definition of urban and rural areas for statistical comparison, the Degree of Urbanisation. The commitment was launched at the Habitat III conference in 2016 and after few years of testing at global level and consultations with over 80 countries, the 51st Session of the United Nations Statistical Commission adopted the Degree of Urbanisation as method to delineate cities, urban and rural areas for international statistical comparisons. A common approach to define urban and rural areas is key to monitor progress made on the 2030 Development Agenda, development and cooperation policies, and disaster risk reduction among other sectors.

The Joint Research Centre (JRC) facilitated the development of the Degree of Urbanisation and its testing outside Europe with the Global Human Settlement Layer (GHSL) providing global and multi-temporal built-up area and population grids, and developing a model to apply the Degree of Urbanisation to these grids. With the adoption of the Degree of Urbanisation the JRC developed a suite of open and free tools to apply this method to nationally produced data, such as national census or survey data. The tools support the capacity enhancement activities carried out by the European Commission and its partners.

This report applies the Degree of Urbanisation method to the 2018 edition of the Colombia Population and Housing Census by creating a population and a settlement classification grid, plus the classification of local units by Degree of Urbanisation and finally the disaggregation of selected population and housing variables important in the Sustainable Development Goals (SDG) perspective by Degree of Urbanisation. Results show, among other things, that share of urban population based on classified territorial units is 77.9%, with share of women being lower than men only in rural areas; that access to the Internet is three times higher in cities than towns, with Semi-dense towns being under-served respect to peri-urban areas; and that rate of connection to the sewage system is 3.8 times higher in cities than in rural areas.These statistics can be internationally compared with other ones collected from population and housing census and aggregated according to the standard methodology defined by the Degree of Urbanisation.

The principal purpose of this technical report is to explain the typologies of input data, chains of operations, and tools that can be used to apply the Degree of Urbanisation to nationally produced data. The work is made possible by the open data policy of the National Statistical Office of Colombia (DANE) and is a technical follow-up to the EUROSTAT manual "Applying the Degree of Urbanisation" produced at the request of the 51st United Nations Statistical Commission.

1 Introduction

Demographic, housing and health surveys are key sources of information for policymakers that supply essential data on population and their living conditions. These data, usually collected with national censuses, are indispensable for several policy domains and they are also required to monitor the implementation of the 2030 Agenda for Sustainable Development, which requires the reporting of several indicators disaggregated by urban and rural areas. Accordingly, the capacity to define urban and rural areas is needed both to collect data, and to analyse the different conditions in these areas. The definition of urban and rural areas was not unique and commonly agreed worldwide before the 51st United Nations Statistical Commission that adopted such a harmonised definition for international comparison: the Degree of Urbanisation. The Degree of Urbanisation method helps to standardise information collection and reporting across countries by offering a common method to classify urban and rural areas for statistical comparison based on population size and density criteria and by a geospatial method relying on contiguity.

Geospatial data are powerful assets linking information to their specific geographical location (e.g. a building or a larger designation like an administrative area). Geospatial data application to official statistics is widely recognised, and increasingly used by National Statistical Offices (UN-GGIM 2020). Oftentimes official geospatial data on demographic and housing census is restricted information available only within National Statistical Offices due to security and privacy concerns associated with fine scale and sensitive information. In other cases, no recent census or survey data are available due to adverse situations (e.g. conflicts, ethnic tensions or swift economic changes). In these cases population modelling is used to estimate populations to support essential planning and delivery of services, infrastructure and aid (Wardrop et al. 2018).The Degree of Urbanisation is a method that classifies administrative or statistical areas into three classes starting from population information generally produced for census rounds. EUROSTAT and other international Organisations produced a the request of the United Nations Statistical Commission a methodological manual about the Degree of Urbanisation (European Commission and Statistical Office of the European Union 2021).

The basic requirement to apply the Degree of Urbanisation method is a population raster on a regular grid in equal area projection that can be derived from population data in geospatial format (population counts per polygon or per census point i.e. household locations). The first stage is the application of a set of criteria to the population grid (settlement classification definitions –see 2.3.1) to obtain a settlement classification grid. The settlement classification grid, combined with the population grid, is then used in the second stage to classify local administrative units based on population majority per class (at the grid level). Local units classified by Degree of Urbanisation allow the analysis of statistics by Degree of Urbanisation. This process is named application of the Degree of Urbanisation, and the latter step is the disaggregation of variables by Degree of Urbanisation. The Degree of Urbanisation has been tested at global level by Dijkstra et al. (Dijkstra et al. 2020), and the Organisation for Economic Cooperation and Development (OECD) and the European Commission analysed the relationships between urbanisation and development at global level (OECD and European Commission 2020).

The Departamento Administrativo Nacional de Estadística of Colombia (DANE) conducted a new round of population and housing census in a multi-year process and released the results as detailed, open and free geospatial data[1] during 2020. The availability of such recent and rich data source has presented an opportunity to apply the Degree of Urbanisation method to Colombia, and analyse some demographic and housing variables by Degree of Urbanisation. This implementation builds upon the European Commission guide "Applying the Degree of Urbanisation - a methodological manual to define cities, towns and rural areas for international comparisons" requested by the 51st Session of the United Nations Statistical Commission upon adoption of the Degree of Urbanisation method. This report serves as practical implementation of the method to post 2015 census information and is a showcase exercise of the reach of the Degree of Urbanisation to make the most impact for international comparison out of national demographic and housing census.

The implementation presented in this report relies on the Global Human Settlement Layer (GHSL) framework of data and tools that support the implementation of the Degree of Urbanisation. The process is performed using the GHSL suite of free tools[2] that operationalise the Degree of Urbanisation method. These GHSL tools enable the production of population grids (GHS-POP2G Tool), settlement grids (GHS-DUG Tool), and the classification of local units by Degree of Urbanisation (GHS-DU-TUC).

[1] Geospatial census data access portal: https://geoportal.dane.gov.co/servicios/descarga-y-metadatos/visor-descarga-geovisores/

[2] https://ghsl.jrc.ec.europa.eu/tools.php

2 Data and Methods

This section describes the two main data sources used in this study: the 2018 National Population and Housing Census of Colombia, and the GHSL framework of data and methods.

2.1 Colombian *Censo Nacional de Población y Vivienda 2018*

The national population and housing census for Colombia was driven by three main objectives: to know how many people live in Colombia, where these people are located, and how they live. The production of periodic census is required by law and the census constitutes the main source of the national statistical system. The whole process unfolded between 2013 and 2020, and included legal, logistical, practical and analytical phases. Data collection included several variables grouped in thematic areas of location (i.e. hierarchical from department to municipality, etc.), *vivienda* (typology, access to infrastructure and utilities, etc.), households (structure of the household etc.) and people (i.e. sex, age, education etc.). These data are publically available and were obtained through the DANE Geoportal[3].

The 2018 edition of the census has been the first one in Colombia also facilitated by digitalisation (12.5% of the total population was surveyed digitally).

2.2 Global Human Settlement Built-up Layer from Sentinel-2 Convolutional Neural Networks

Global mapping of human settlements has improved considerably with the open availability of Earth Observation (EO) data provided by programmes like Landsat and Copernicus. In parallel, also the capacity to process these data and extract valuable information has leaped thanks to advances in machine learning and artificial intelligence. The GHS-BUILT-S2 R2020A is the first global product mapping built-up areas for the reference year 2018 using Convolutional Neural Networks (GHS-S2Net, Corbane et al. 2020). The layer builds on a new deep learning framework for pixel-wise large-scale classification of built-up areas named GHS-S2Net to produce a global map of built-up areas expressed in terms of a probability grid at 10 m spatial resolution derived from a Sentinel-2 global image composite.

The main improvement of the GHS-BUILT-S2 R2020A layer compared to other remote sensing derived human settlement maps are the high balanced accuracy (exceeding 0.7) and the 2018 reference year, which makes it very suitable as covariate for disaggregating recent demographic data.

2.3 The Degree of Urbanisation

The Degree of Urbanisation is a method to classify the entire territory of a country or any smaller area into three classes: 1) cities, 2) towns and semi-dense areas and 3) rural areas based on a combination of population density and size, and a geospatial concept of grid cells contiguity. The 51st United Nations Statistical Commission adopted the Degree of Urbanisation as method to delineate cities, urban and rural areas for international statistical comparisons.

The method is applied to a regular kilometric population grid in equal area projection (e.g. World Mollweide EPSG: 54009) in two stages: first grid cells are classified into one of the three classes producing a settlement grid, then small local spatial units are classified based on the grid typology where the majority of units' population is located.

2.3.1 Grid Cell classification

Groups of 1 km^2 population grid cells are plotted in relation to their neighbouring cells to identify:

— An **urban centre (high-density cluster),** a cluster of contiguous grid cells of 1 km^2 (using four-point contiguity, in other words, excluding diagonals) with a population density of at least 1,500 inhabitants per km^2 and collectively a minimum population of 50,000 inhabitants. Gaps in this cluster are filled (if smaller than 15 km^2) and edges are smoothed (3-by-3 majority filter). If needed, contiguous cells (using four-point contiguity) that are at least 50% built-up may be added. In the geospatial raster grid, urban centre cells are coded with value "3".

[3] https://geoportal.dane.gov.co/servicios/descarga-y-metadatos/visor-descarga-geovisores/

- An **urban cluster (moderate-density cluster),** a cluster of contiguous grid cells of 1 km² (using eight-point contiguity, in other words, including diagonals) with a population density of at least 300 inhabitants per km² and a minimum population of 5,000 inhabitants. In a last step, after the verification of the minimum population, the urban centre cells are removed from the urban cluster. In the geospatial raster grid, urban cluster cells are coded with value "2".
- **Rural grid cells (or mostly low-density cells),** grid cells that are not identified as urban centres or as urban clusters. In the geospatial raster grid, rural grid cells are coded with value "1".

The Degree of Urbanisation has also an extension, named Level 2 (whereas the above description is the Level 1 – 3 classes) that has 7 classes as hierarchical sub-classification of the Level 1. The second level has been introduced to capture the full settlement hierarchy of large, medium and small settlements or, in simpler terms, cities, towns and villages. The definition of the first class remains unchanged:

- An **urban centre** consists of contiguous (using four-point contiguity) grid cells with a density of at least 1,500 inhabitants per km². An urban centre has a collective population of at least 50,000 inhabitants. Gaps in this cluster are filled (if smaller than 15 km²) and edges are smoothed (3-by-3 majority filter). If needed, contiguous cells (using four-point contiguity) that are at least 50% built-up may be added. In the geospatial raster grid, urban centre cells are coded with value "30".

The urban cluster cells <u>that are not part of an urban centre</u> can be subdivided into three types.

- A **dense urban cluster** consists of contiguous (using four-point contiguity) grid cells with a density of at least 1,500 inhabitants per km², with a collective population of at least 5,000 inhabitants and less than 50,000 inhabitants in the cluster. In the geospatial raster grid, dense urban cluster cells are coded with value "23".
- A **semi-dense urban cluster** consists of contiguous (using eight-point contiguity) grid cells with a density of at least 300 inhabitants per km² and has a collective population of at least 5,000 inhabitants (in other words, an urban cluster) and this cluster is neither contiguous with nor within 2 km of a dense urban cluster or an urban centre([4]). In the geospatial raster grid, semi-dense urban cluster cells are coded with value "22".
- **Suburban or peri-urban cells** are the remaining urban cluster cells, in other words those not part of a dense or semi-dense urban cluster. These grid cells are part of an urban cluster that is contiguous (using eight-point contiguity) or within 2 km of a dense urban cluster or an urban centre. In the geospatial raster grid, suburban or peri-urban cells are coded with value "21".

Rural grid cells can be categorised into three types.

- A **rural cluster** consists of contiguous (using eight-point contiguity) grid cells with a density of at least 300 inhabitants per km² and a collective population between 500 and 5,000 inhabitants in the cluster. In the geospatial raster grid, rural cluster cells are coded with value "13".
- **Low density rural grid cells** are rural grid cells with a density of at least 50 inhabitants per km² and are not part of a rural cluster. In the geospatial raster grid, low density rural grid cells are coded with value "12".
- **Very low density rural grid cells** are rural grid cells with a density of less than 50 inhabitants per km². In the geospatial raster grid, very low density rural grid cells are coded with value "11".

2.3.2 Classification of local units

The Degree of Urbanisation classifies local units based on population majority[5] applied to the grid level classification. Each local unit belongs exclusively to one class at Level 1 and one at Level 2 (hierarchy based). Local units can be administrative units - such as municipalities - or statistical units - such as census enumeration areas.

[4] Measured as outside a buffer of three grid cells of 1 km² around dense urban clusters and urban centres. This ensures that adjacent, but not contiguous suburbs are taken into account.

[5] To classify the spatial units without any population, the same rules explaned below should be applied to their area instead of to their population

Once all grid cells have been classified as urban centres, urban clusters and rural grid cells the next step concerns overlaying these results onto local units:

- **Cities** (or densely populated areas): local units that have at least 50% of their population in urban centres, are coded in their attribute table under degree of urbanisation Level 1 with the code "3".
- **Towns and semi-dense areas** (or intermediate density areas): local units that have less than 50% of their population in urban centres and less than 50% of their population in rural grid cells, code "2".
- **Rural areas** (or thinly populated areas): local units that have at least 50% of their population in rural grid cells, code "1".

Urban areas consist of cities plus towns and semi-dense areas.

At the Level 2 the same logic is implemented as hierarchical sub-classification of the Level 1. Local units classified as "towns and semi-dense areas" can be divided into three classes:

- **Dense Towns** have a larger share (>= 50%) in dense plus semi-dense urban clusters than in suburban or peri-urban cells (i.e. it is a town) and a larger share (>= 50%) of their population in dense urban clusters than in semi-dense urban clusters (i.e. it is dense), code "23".
- **Semi-dense Towns** have a larger share (>= 50%) in dense plus semi-dense urban clusters than in suburban or peri-urban cells (i.e. it is a town) and a larger population share (> 50%) in semi-dense urban clusters than in dense urban clusters (i.e. it is semi-dense) and, code "22".
- **Suburban or peri-urban areas** have a larger population share (> 50%) in suburban or peri-urban cells than in dense plus semi-dense urban clusters, code "21".

Dense and semi-dense towns can be combined into a single class of **towns**. This reduces the number of classes and is especially useful if the population share in semi-dense towns is low.

Local units classified as "rural areas" can be divided into three classes:

- **Villages** have the largest share of their rural grid cell population living in a rural cluster[6], code "13".
- **Dispersed rural areas** have the largest share of their rural grid cell population living in low density rural grid cells[7], code "12".
- **Mostly uninhabited areas** have the largest share of their rural grid cell population living in very low density rural grid cells, code "11".

Once local units are classified by Degree of Urbanisation it is possible to summarize statistics by classes at Level 1 or Level 2. This is particularly useful to harmonise reporting for Sustainable Development Goals indicators, or cross compare statistics across countries.

[6] Special cases: share equal to the proportion of rural grid cell population living in low density rural grid cells and to the proportion of rural grid cell population living in very low density rural grid cells; or share equal to one and larger than the other.

[7] Special case: share larger then the proportion of rural grid cell population living in a rural cluster and equal to the proportion of rural grid cell population living in very low density rural grid cells

2.4 GHSL Tools to implement the Degree of Urbanisation

The Degree of Urbanisation is a simple geospatial concept implemented through geospatial and geostatistical processing. Such processing requires a set of notions and a specific workflow to handle geospatial data in a rigorous way. In order to streamline this processing, and to expand as much as possible the users of the Degree of Urbanisation, the Joint Research Centre of the European Commission developed a set of tools for the implementation of the Degree of Urbanisation method, along its various steps. GHSL Tools were developed as standalone executables relying on MATLAB Runtime, and as ArcGIS Toolboxes.

The tools described below address three production steps also applied in this study. The first step is the construction of a regular-spaced population grid from available geospatial population data in the form of polygons or points with the GHS-POP2G Tool. The second step is the application of the Degree of Urbanisation methodology with the GHS-DUG Tool to the population grid resulting from the previous step. Next, the derived grid cell classification is used to classify small spatial units into cities, towns and semi dense areas, or rural areas with the GHS-DU-TUC (Figure 1).

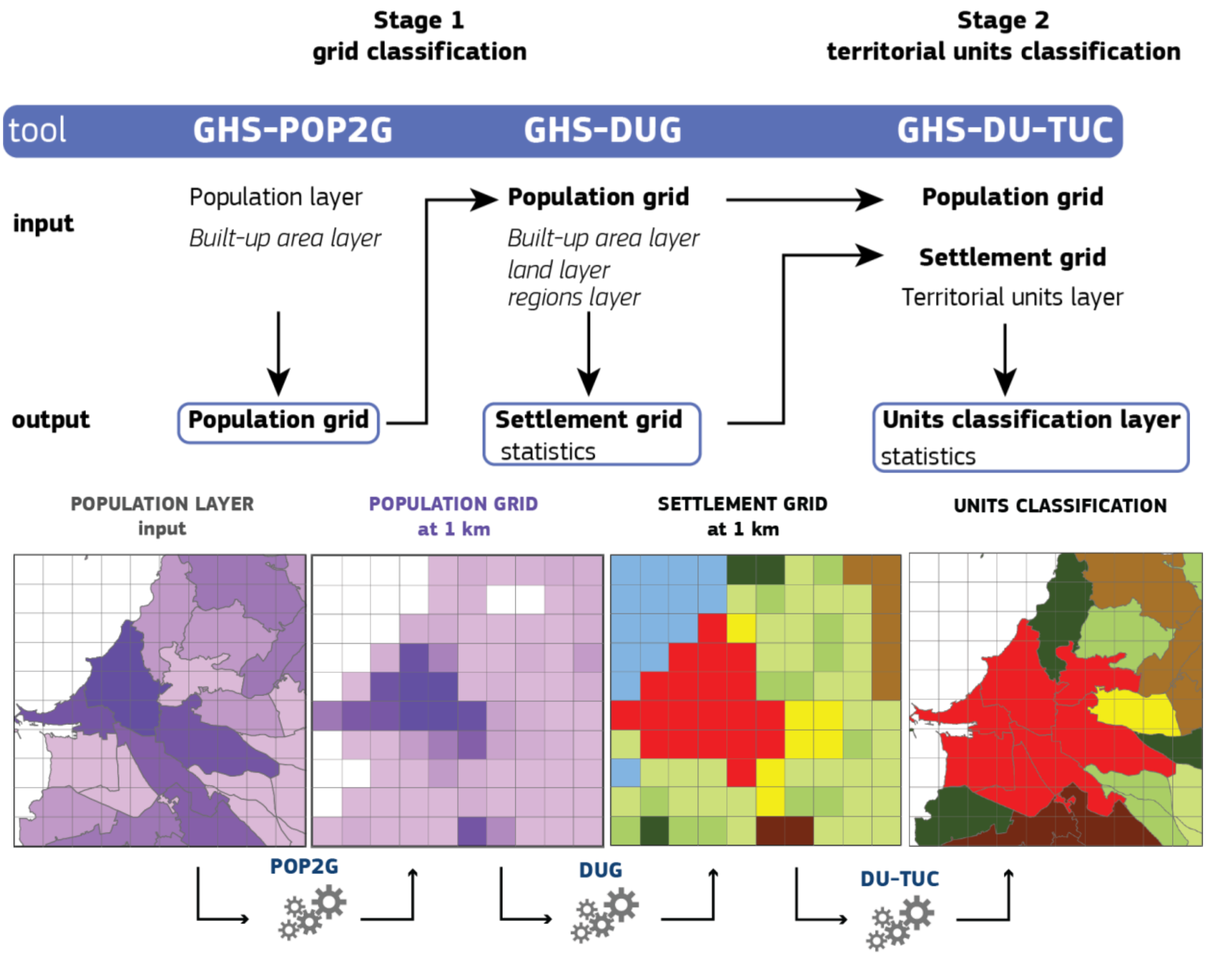


Figure 1 Applying the Degree of Urbanisation with GHSL Tools

2.4.1 Creating a population grid with the GHS-POP2G Tool

The Population-To-Grid tool (GHS-POP2G) produces geospatial raster grids of population counts at different spatial resolutions. The tool allows the creation of population grids at 50 m, 100 m, 250 m, and 1 km spatial resolutions, starting from census data stored as point or polygon vector data (in latter case requires additional covariate as input for dasymetric disaggregation).

The GHS-POP2G allows to produce a population grid both by the method of dasymetric disaggregation of population data stored as polygons (using a built-up area layer as covariate), and by the method of aggregation of population data stored as points (like micro-census data).

The GHS-POP2G user guide (Maffenini, Schiavina, Freire, et al. 2020) explains in detail the operations to produce a population grid.

2.4.2 Producing a settlement classification grid with the GHS-DUG Tool

The Degree of Urbanisation Grid (GHS-DUG) Tool produces geospatial grids at 1 km by mapping settlement classes and extracts related statistics. GHS-DUG requires a population grid as input (i.e. produced with the GHS-POP2G) and optionally other ancillary data like a built-up area layer or a custom land/water layer used to compute densities.

The GHS-DUG user guide (Maffenini, Schiavina, Melchiorri, et al. 2020a) explains in detail the operations to produce a settlement classification grid.

2.4.3 Classifying local units by Degree of Urbanisation with the GHS-DU-TUC

The Degree of Urbanisation Territorial Units Classifier (GHS-DU-TUC) Tool classifies the territorial units based on the Degree of Urbanisation and extracts related statistics. The tool classifies territorial units by Degree of Urbanisation at level 1 (3 classes) and level 2 (7 classes) based on population majority by gridded settlement classes produced with GHS-DUG Tool.

The GHS-DU-TUC is designed as an operational tool to perform the classification of local administrative units based on the previous described settlement classification grid. The tool requires as input the settlement classification grid (e.g. produced with the GHS-DUG Tool), the population grid used to produce the settlement classification grid (e.g. produced with the GHS-POP2G Tool) at its highest resolution (between 50 m and 1 km)[8] and a geometry of territorial units to be classified by Degree of Urbanisation.

The GHS-DU-TUC user guide (Maffenini, Schiavina, Melchiorri, et al. 2020b) explains in detail the operations to classify local units by Degree of Urbanisation.

[8] The GHS-DU-TUC solves the small spatial unit issue (see chapter 6.3 of European Commission and Statistical Office of the European Union 2021) by using a raster with a 50 m spatial resolution. Therefore, all population grids with resolution that are multiple of 50 m are suitable.

2.5 Methodology

The application of the Degree of Urbanisation to the *2018 Censo Nacional de Población y Vivienda* of Colombia has been implemented in four distinct phases. First, the demographic and built-up area data were pre-processed to obtain data compatible with the GHSL Tools workflow (1 and 2 in Table 1). Next, the tools were used to produce the Degree of Urbanisation layers (3). Last, the demographic and housing information was obtained by Degree of Urbanisation and analysed (4).

In the first phase (1), the built-up layer has been pre-processed transform the GHS-BUILT Sentinel 2 built-up probability tiles hosted on the GHSL website to a continuous layer of built-up surface (i.e. suitable for population disaggreagation) at 50 m resolution. This was achieved through four actions with which the single tiles were first binarised (i.e. transform the probability into built-up presence/absence) by thresholding with a value of 0.2 – as recommended in Corbane et al. 2020 (1.1), the thresholded tiles where then warped and oversampled to 5 m resolution in World Mollweide projection (1.2). Next, the binary warped tiles at 5 m were mosaicked into a single continuous raster covering the full extent of Colombia census (1.3), and last the continuous thresholded layer was warped to 50 m (1.4).

The second phase has been the preparation of Colombia census by merging the finest nationally defined rural and urban available geometries in the census (the *Manzana* and *Seccion Rural*) into a single shapefile (2.1). Once this continuos shapefile containing population counts per unit was available, a new field containing a new unique identifier (UID) for each of the unit was added (2.2).

The third phase consisted in running the GHS-Tools to produce the population grid with the GHS-POP2G tool (3.1), to produce the settlement classification grid with the GHS-DUG (3.2), and to classify local units by Degree of Urbanisation with the GHS-DU-TUC (3.3). The GHS-POP2G standalone tool was configured to output the grid at 50 m, to source population counts from the TP_27 field, to identify polygons by the UID field, and to use as built-up layer the output of step 1.4 (Figure 2). The processing has run with the multicore option for about 4 hours on a server with 256 GB of RAM and Intel® Xeron ® CPU E-5 2650.

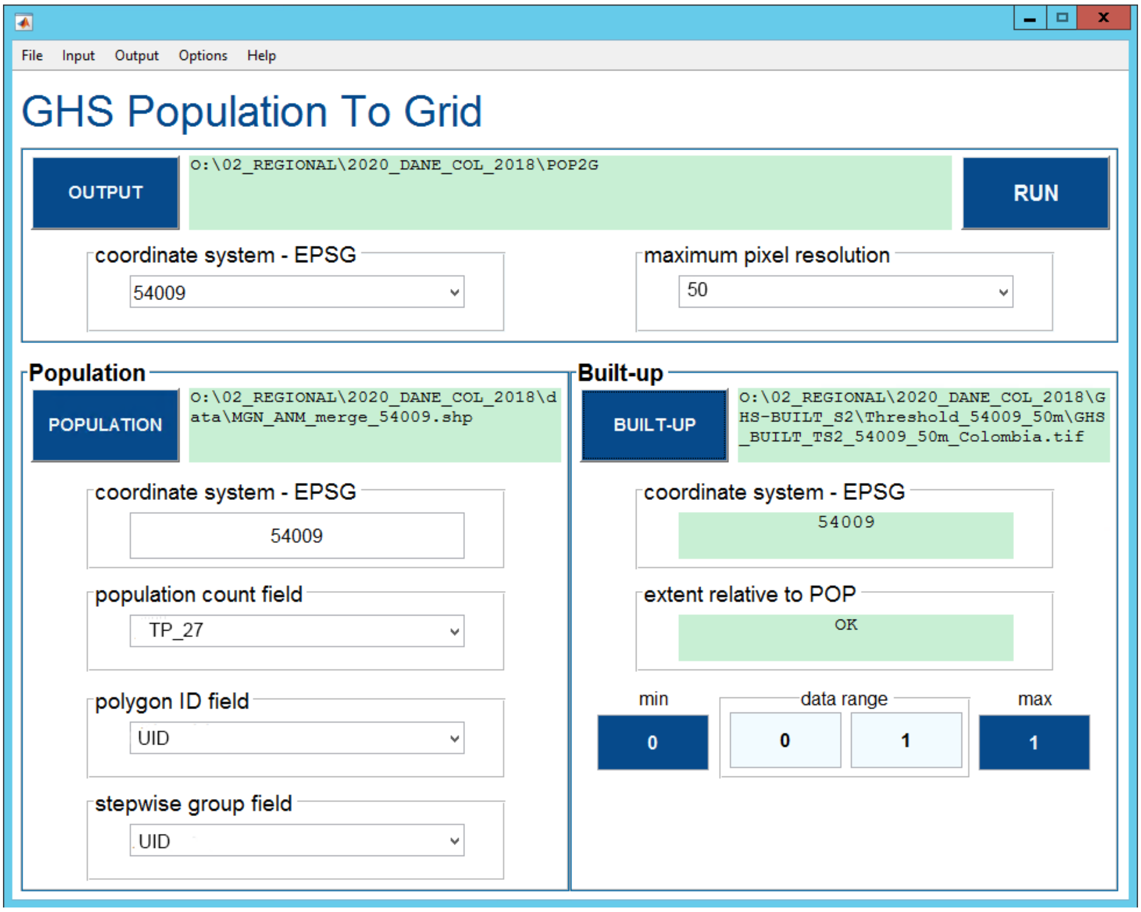


Figure 2 GHS-POP2G interface with the parameters set for the production of the population grid

The GHS-DUG (Figure 3) was configured to obtain the settlement classification grid using the population grid (1 km) obtained in 3.1 and the built-up surface from 1.4 to apply the 50% built-up rule in urban centres. In the same computing infrastructure mentioned above the process took about 2 minutes to complete.

The GHS-DU-TUC (Figure 4) was configured to obtain the classification of local units (2.2) by Degree of Urbanisation based on the settlement classification grid (3.2), and the population grid at the finest available resolution (3.1). Using the same infrastructure, the process completed in about 4 hours.

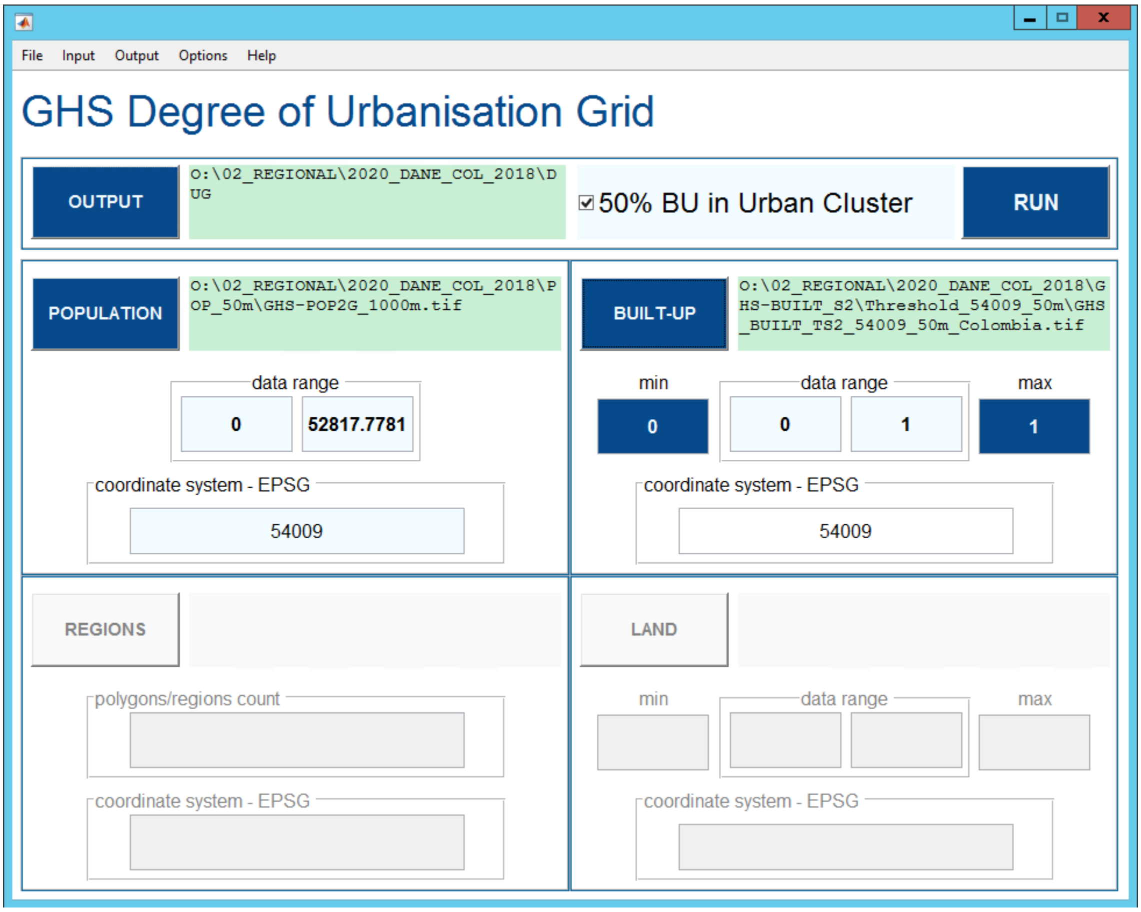


Figure 3 GHS-DUG interface with the parameters set for the production of the settlement classification grid

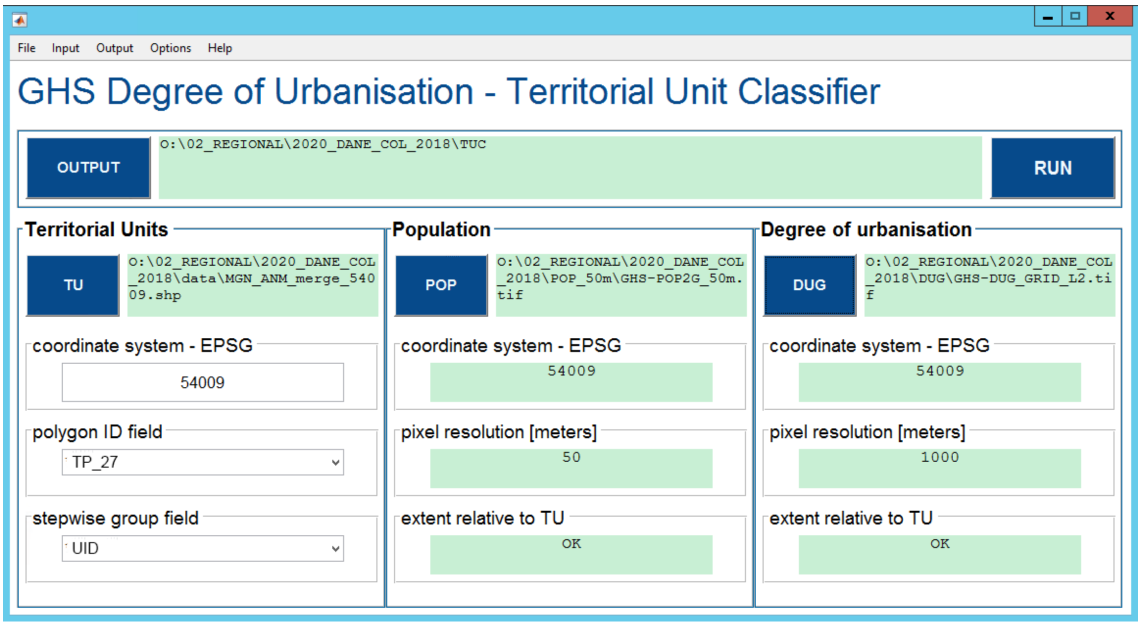


Figure 4 GHS-DU-TUC interface with the parameters set for the classification of the local units by Degree of Urbanisation

In the last phase, the spreadsheet resulting from the territorial units classification were analysed with a pivot table to disaggregate demographic and housing statistics by Degree of Urbanisation.

Some selected variables available for *viviendas* relevant in the Sustainable Development Goals context were analysed by Degree of Urbanisation:

- Number of housing units with and without electric energy infrastructure (TP19_EE_1, and TP19_EE_2);
- Number of housing units with and without water infrastructure (TP19_ACU_1, and TP19_ACU_2);
- Number of housing units with and without a connection to the sewage system (TP19_ALC_1, and TP19_ALC_2);
- Number of housing units with and without solid waste collection (TP19_RECB1, and TP19_RECB2);
- Number of housing units with, without, and unknown internet service (TP19_INTE1, TP19_INTE2, TP19_INTE9);

To obtain the number of *viviendas* with the characteristics above, we performed a statistical analysis counting all the *viviendas* within each Degree of Urbanisation class (at Level 1 and Level 2). This operation was performed with pivot tables. This calculation has been the basis for obtaining the shares of units with and without access per class and at national level.

Row Labels	Sum of 0-9 years (TP34_1_EDA)	Sum of 10-19 years (TP34_2_EDA)	Sum of 20-29 years (TP34_3_EDA)	Sum of 30-39 years (TP34_4_EDA)	Sum of 40-49 years (TP34_5_EDA)	Sum of 50-59 years (TP34_6_EDA)	Sum of 60-69 years (TP34_7_EDA)	Sum of 70-79 years (TP34_8_EDA)	Sum of 80+ years (TP34_9_EDA)
1	1,820,738	2,047,108	1,570,471	1,351,034	1,184,269	1,029,764	717,319	423,112	212,063
2	1,596,642	1,844,181	1,667,273	1,411,261	1,190,038	1,029,986	671,154	374,380	197,966
3	3,470,090	4,083,922	4,808,179	4,176,396	3,349,427	3,102,729	2,013,645	1,040,497	542,129
Grand Total	**6,887,470**	**7,975,211**	**8,045,923**	**6,938,691**	**5,723,734**	**5,162,479**	**3,402,118**	**1,837,989**	**952,158**
	0-9 years	**10-19 years**	**20-29 years**	**30-39 years**	**40-49 years**	**50-59 years**	**60-69 years**	**70-79 years**	**80+ years**
1	17.6%	19.8%	15.2%	13.0%	11.4%	9.9%	6.9%	4.1%	2.0%
2	16.0%	18.5%	16.7%	14.1%	11.9%	10.3%	6.7%	3.8%	2.0%
3	13.1%	15.4%	18.1%	15.7%	12.6%	11.7%	7.6%	3.9%	2.0%
Country	*14.7%*	*17.0%*	*17.1%*	*14.8%*	*12.2%*	*11.0%*	*7.2%*	*3.9%*	*2.0%*

Figure 5 Pivot table used to compute the counts and shares of population age classes by degree of urbanisation: (1 = Rural Area; 2 = Towns and Semi-dense areas; 3 = City)

Table 1 Workflow to apply the Degree of Urbanisation with GHS-Tools with pre-processing and analysis steps

Step	Action	Input files	Method	Output files
1. Pre–processing Built-up layer	1. Binarization of built-up probability layer in original UTM zone	17N_PROB.tif, 17P_PROB.tif, 18M_PROB.tif, 18N_PROB.tif, 18P_PROB.tif, 19M_PROB.tif, 19N_PROB.tif, 19P_PROB.tif	Thresholding	17N_PROB_T02.tif, 17P_PROB_T02.tif, 18M_PROB_T02.tif, 18N_PROB_T02.tif, 18P_PROB_T02.tif, 19M_PROB_T02.tif, 19N_PROB_T02.tif, 19P_PROB_T02.tif
	2. Warp with oversampling to 5 m World Mollweide	17N_PROB_T02.tif, 17P_PROB_T02.tif, 18M_PROB_T02.tif, 18N_PROB_T02.tif, 18P_PROB_T02.tif, 19M_PROB_T02.tif, 19N_PROB_T02.tif, 19P_PROB_T02.tif	Nearest neighbour resampling	17N_54009_5m.tif, 17P_54009_5m.tif, 18M_54009_5m.tif, 18N_54009_5m.tif, 18P_54009_5m.tif, 19M_54009_5m.tif, 19N_54009_5m.tif, 19P_54009_5m.tif
	3. Mosaic the 5 m tiles into a one continuous layer	17N_54009_5m.tif, 17P_54009_5m.tif, 18M_54009_5m.tif, 18N_54009_5m.tif, 18P_54009_5m.tif, 19M_54009_5m.tif, 19N_54009_5m.tif, 19P_54009_5m.tif	Mosaic with average algorithm	GHS_BUILT_TS2_54009_5m_Colombia.tif
	4. Aggregate the 5 m layer into a 50 m layer	GHS_BUILT_TS2_54009_5m_Colombia.tif	Average resampling	GHS_BUILT_TS2_54009_50m_Colombia.tif
2. Pre-processing Population census data	1. Merge urban and rural shapefiles at the most detailed geometry level	MGN_ANM_SECCION_RURAL.shp MGN_ANM_MANZANA.shp	Merge	MGN_ANM_merge_54009.shp
	2. Add new unique ID field	MGN_ANM_merge_54009.shp	Add field	MGN_ANM_merge_54009.shp
3. Run GHS-Tools	1. Produce population grid	MGN_ANM_merge_54009.shp, GHS_BUILT_TS2_54009_50m_Colombia.tif	GHS-POP2G	GHS-POP2G_1000m.tif, GHS-POP2G_50m.tif
	2. Produce settlement classification grid	GHS-POP2G_1000m.tif, GHS_BUILT_TS2_54009_50m_Colombia.tif	GHS-DUG	GHS-DUG_GRID_L1.tif, GHS-DUG_GRID_L2.tif, GHS-DUG_L1_stats.xls, GHS-DUG_L2_stats.xls, GHS-DUG_URBAN_CENTRE.shp, GHS-DUG_SEMI_DENSE_URBAN_CLUSTER.shp, GHS-DUG_RURAL_CLUSTER.shp, GHS-DUG_DENSE_URBAN_CLUSTER.shp
	3. Classify local units by Degree of Urbanisation	GHS-POP2G_50m.tif, GHS-DUG_GRID_L2.tif, MGN_ANM_merge_54009.shp	GHS-DU-TUC	MGN_ANM_merge_54009_GHS-DU-TUC.shp, MGN_ANM_merge_54009_GHS-DU-TUC.xlsx
4. Disaggregate statistics	1. Disaggregate demographic and housing information by Degree of Urbanisation	MGN_ANM_merge_54009_GHS-DU-TUC.xlsx	Pivot table	MGN_ANM_merge_54009_DEGURBA_stats.xlsx

3 Results

This section presents the results and statistics of the different outputs generated by the application of the Degree of Urbanisation method to census 2018 of Colombia.

3.1 Population grid

The total population contained in the population grid is 46,925,773 people. Overall 4.6% of this figure is counted in polygons, where GHS-BUILT from Sentinel 2 detected no built-up areas (in such cases population is spread evenly in the cells forming the respective rasterised polygon – areal weighting).

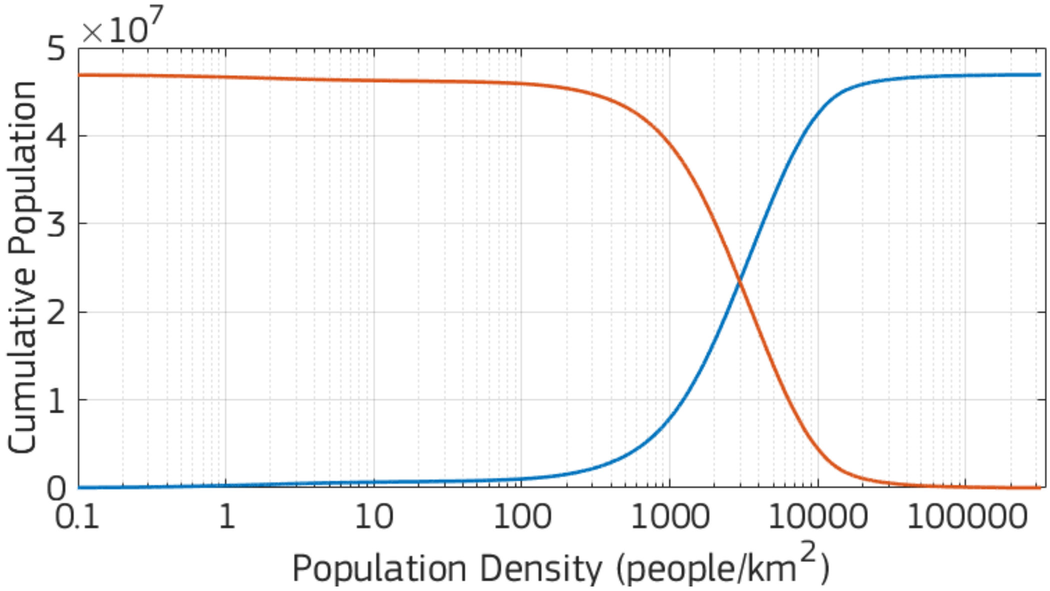


Figure 6 Cumulative (blue) and reverse cumulative (red) population distributions per levels of density in the population grid at 50 m

Table 2 Population distribution by density classes in both 50 m and 1 km population grids

Population density (people/km²)	**50 m Population grid**		**1 km Population grid**	
0 - 50	814,794	1.7%	1,589,006	3.4%
50 - 100	204,113	0.4%	1,094,135	2.3%
100 - 200	531,285	1.1%	1,665,624	3.5%
200 - 400	1,359,687	2.9%	2,229,843	4.8%
400 - 800	3,255,748	6.9%	2,735,425	5.8%
800 - 1,600	7,159,852	15.3%	3,178,089	6.8%
1,600 - 3,200	11,607,296	24.7%	3,867,562	8.2%
3,200 - 6,400	12,153,439	25.9%	5,709,058	12.2%
6,400 - 12,800	7,248,326	15.4%	8,644,510	18.4%
12,800 - 25,600	1,876,363	4.0%	10,360,733	22.1%
25,600 - 51,200	480,359	1.0%	5,746,725	12.2%
51,200 - 102,400	159,392	0.3%	105,063	0.2%
102,400 - 204,800	61,713	0.1%	-	0.0%
204,800 - 409,600	13,405	0.0%	-	0.0%

Table 2 and Figure 6 show the population distributions in the population grids produced. Half of population lives in areas of 50 m by 50 m (pixels) accommodating more than 3,200 people per km^2, while less than 20% of people in both grids lives in areas with a density lower than 1,600 people per km^2. The punctual density (at 50 m) shows a peak (25.9% of total population) of more than 12 million people between 1,600 people per km^2 and 3,200 people per km^2, while the density on a broader scale (at 1 km) has a peak (22.1% of total population) of more than 10 million people between 12,800 and 25,600 people per km^2. Figure 7 shows a detail of 50-m population grid in the centre of Bogotá.

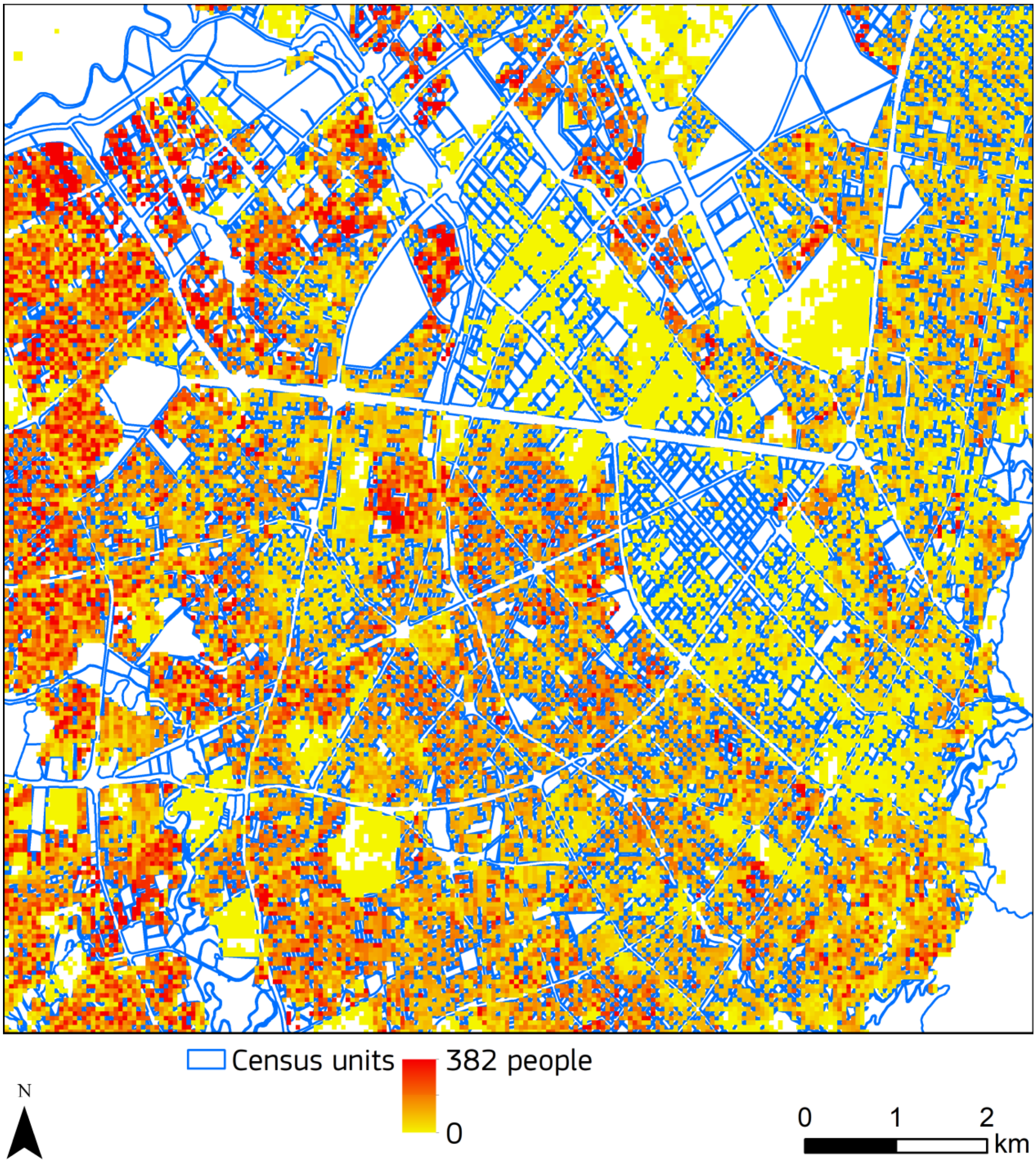


Figure 7 Centre of Bogotá, population grid at 50 m

3.2 Degree of Urbanisation grid

The Degree of Urbanisation has been produced at Level 1 and Level 2 (summary in Table 3). At Level 1, 22.4% of the population is accounted in rural grid cells (about 10.5 million people), 20.9% in urban cluster grid cells (about 9.8 million people), and 56.6% in urban centres grid cells (about 26.5 million people). At Level 2, urban cluster breaks down into 13% of the total population in dense urban clusters, 2.8% in semi-dense urban cluster and 5% in suburban or peri-urban grid cells. The rural domain at Level 2 breaks down into 9.1% of the total population in rural clusters, 10% in low density rural grid cells and the remaining 3.4% in very low density rural grid cells. The share of urban population based on Degree of Urbanisation grid is 77.5%.

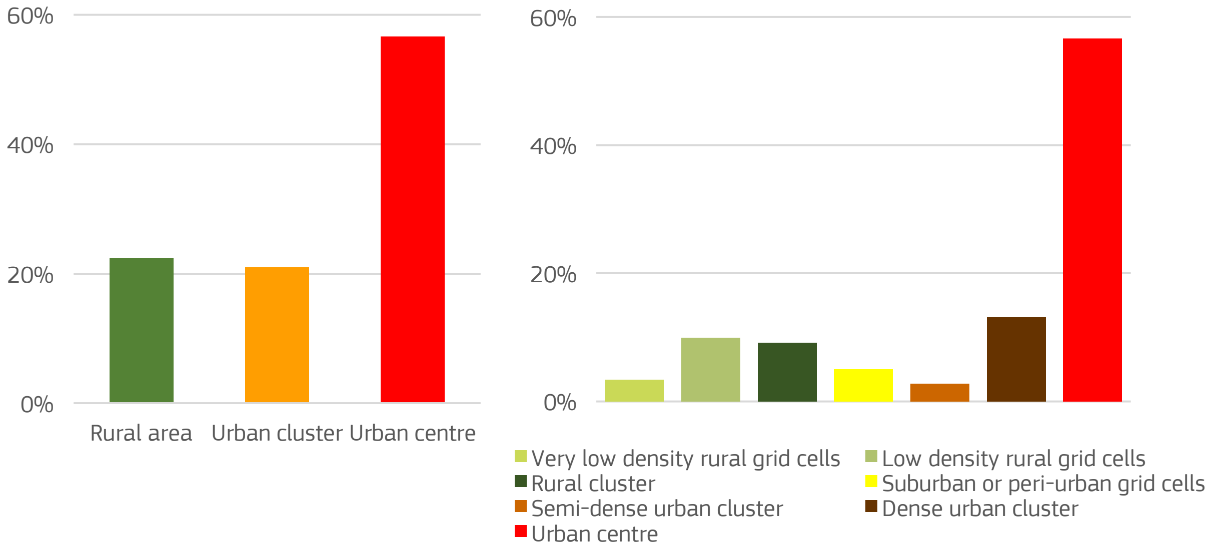


Figure 8 Population share by Degree of Urbanisation Level 1 (left) and Level 2 (right)

Concerning the number of cells in each class (i.e. surface), the vast majority of grid cells are classified as rural grid cells (99.7% of the total), 0.2% as urban cluster and 0.1% as urban centres. The class with the largest extent is low density rural grid cells. In the urban domain, the largest is the urban centre class (exceeding 2,800 km^2).

Population density (people/km^2) varies considerably across settlement types, from less than one person per km^2 in very low density grid cells to almost 10,000 people in urban centre grid cells (Table 3).

Table 3 Degree of Urbanisation Level 1 and Level 2 population and area statistics at grid level

Level 1 Class	Population count	Area (km^2)	Population share	Level 2 Class	Population count	Area (km^2)	Population share	Population density people/km^2
Rural area	10,529,239	2,976,980	22.4%	Water	-	796,250	0.0%	-
				Very low density rural grid cells	1,578,418	2,141,415	3.4%	0.7
				Low Density Rural grid cells	4,669,827	34,039	10.0%	137.1
				Rural cluster	4,280,994	5,276	9.1%	811.4
Urban cluster	9,832,695	5,886	21.0%	suburban or peri-urban grid cells	2,351,571	3,067	5.0%	766.7
				Semi-dense urban cluster	1,307,343	1,150	2.8%	1136.8
				Dense urban cluster	6,173,781	1,669	13.2%	3699.1
Urban centre	26,563,840	2,836	56.6%	Urban Centre	26,563,840	2,836	56.6%	9366.6

3.2.1 Settlement entities

Regarding the delineation of individual settlements (entities), the application of the Degree of Urbanisation method to the 2018 census identifies 74 urban centres in Colombia. The largest (both in surface and population) corresponds to the capital city of Bogotá, having a surface of 449 km^2, more than 7.73 million inhabitants and a built-up area surpassing 212 km^2. The second largest urban centre in area overlays with Medellin, it has a surface of 261 km^2, it hosts about 3.59 million people and its built-up area exceeds 46 km^2.

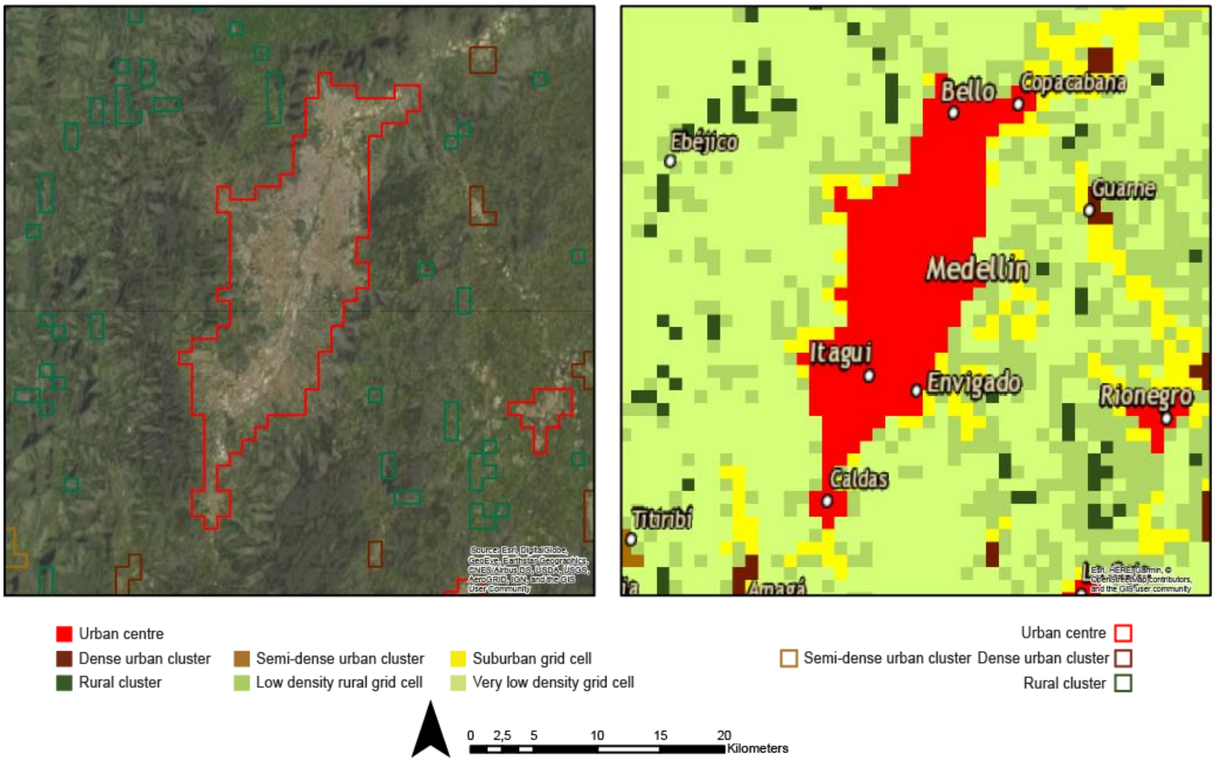


Figure 9 Delineation of urban centres in the area of Medellin (left) and settlement grid over the same area (right)

Forty-eight urban centres (65% of the total) are in a population size class between 50 and 150 thousand inhabitants; other nine urban centres are in the class 150 – 250 thousand inhabitants, four in the class 350 – 450, and three in the class 750 – 850 thousand. The remaining classes (to185 thousand people upper bound) count only one urban centre. The great majority (90%) of urban centres are smaller than 60 km^2 while six are larger than 100 km^2 (Table 4).

Table 4 Ten largest urban centres (population rank) in Colombia

Place	Area (km^2)	Population count
Bogota	449	7,733,295
Medellin	261	3,594,872
Cali	164	1,837,624
Barranquilla	127	1,597,096
Cartagena	154	831,565
Cúcuta	118	760,038
Bucaramanga	62	758,334
Pereira	66	524,457
Ibagué	49	422,133
Valledupar	49	398,845

The Degree of Urbanisation method identifies 440 dense urban clusters with a population ranging between 5,034 and 48,605 people (ten largest in population in Table 5). The largest one is in the area of Santa Cruz de

Lorica. Almost half the dense urban clusters (213 of them) are in the population class below 10,000 inhabitants, 20% between 10 and 15 thousand, 12% between 20 and 25 thousand.

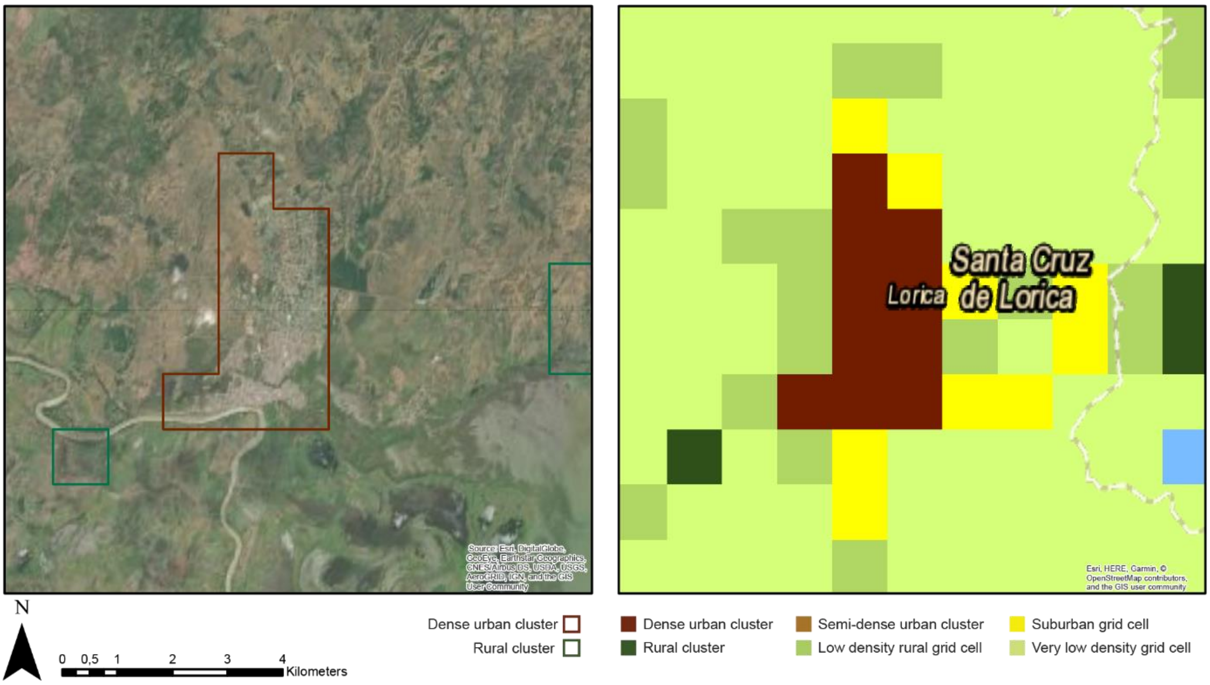


Figure 10 Delineation of dense urban cluster in the area of Santa Cruz de Lorica (left) and settlement grid over the same area (right)

Seventy-five dense urban clusters are in the size classes between 25 and 40 thousand people, 14 between 40 and 50 thousand. Among the largest dense urban clusters are the coastal settlements in the area of Santiago de Tolú (40 km^2 and hosting about 25,000 people) and Puerto Colombia (22 km^2 and about 33,000 people). The extent of dense urban cluster varies between 2 and 48 km^2, of which 85% are less than 5 km^2.

Table 5 Ten largest dense urban clusters (population rank) in Colombia

Place	**Area (km^2)**	**Population count**
Santa Cruz de Lorica	10	48,605
San Andres	48	48,119
Chigorodó	6	47,679
Espinal	10	47,579
San Gil	11	47,173
Turbo	8	46,621
El Carmen de Bolívar	17	45,960
Sahagún	7	44,885
Santander de Quilichao	9	44,378
Chiquinquirá	9	44,024

The Degree of Urbanisation method identifies 192 semi-dense urban clusters with a population ranging between 5,002 and 27,849 people. Most of these clusters (60%) are in a population size class between 5,000 and 6,500 inhabitants, 112 clusters (25%) between 6,500 and 8,000, the remaining 16 clusters have a population between 8,000 and 30,000 inhabitants, whereas only two exceeds 15,000. The size of semi-dense urban cluster varies between 2 and 35 km^2. The majority of these settlements (60%) is smaller than 5 km^2. About 30% are between 5 and 10 km^2. Semi-dense urban clusters between 10 and 20 km^2 are about 12% of the total, and the remaining two clusters are greater than 20 km^2. Other 55 clusters are between 6 and 10 km^2. Seven are between 11 and 20 km^2, while three are larger than 20 km^2.

In the rural domain, the degree of urbanisation maps 2,905 rural clusters. Almost half the clusters (1,311) have a population between 500 and 1,000 inhabitants. About 20% of the rural clusters have a population below 1,500 people and 4% host a population between 4,000 and 5,000 people. Half of these clusters are very small (1 km^2), or small (30%, 2 km^2), about 20% are between 3 and 5 km^2, 26 are between 6 and 8 km^2 and the remaining three rural clusters are 9 km^2.

3.3 Classification of territorial units by Degree of Urbanisation

Based on the Degree of Urbanisation grid and the population grid derived from the *Censo Nacional de Población y Vivienda Colombia 2018* geometries it was possible to code the *Manzana* and *Seccion Rural* administrative units by Degree of Urbanisation. The Degree of Urbanisation classifies local units based on population majority applied to the grid level classification.

The 551,028 units comprising the territorial units layer (merging *Seccion Rural* and *Manzana*) are classified at Degree of Urbanisation Level 1 in 151,855 rural area units, 152,294 towns and semi-dense areas and 246,849 city units. The corresponding population split is 23% in rural area units, 21% in towns and semi-dense areas and 56% in city units (Figure 11 and Table 6). The rural areas and towns and semi-dense areas classes account for a higher share of units (28% in both classes), but host lower shares of the national population compared to cities.

The share of urban population based on classified territorial units is 77.9%.

At level 2 of the Degree of Urbanisation, 7% of the units are classified as mostly uninhabited areas hosting about 1 million people. Dispersed rural areas are 8% of the units and host 4.7 million people, villages are 12% of the units (almost 66,000) and host 4.4 million people. In the urban domain, 6% of census polygons are classified as suburban or peri-urban areas (and host 2.1 million people), semi-dense towns are 4% of the census units and host 1.3 million people, dense towns are 18% of the census units and host 6.1 million people (the city class remains the same as in level 1).

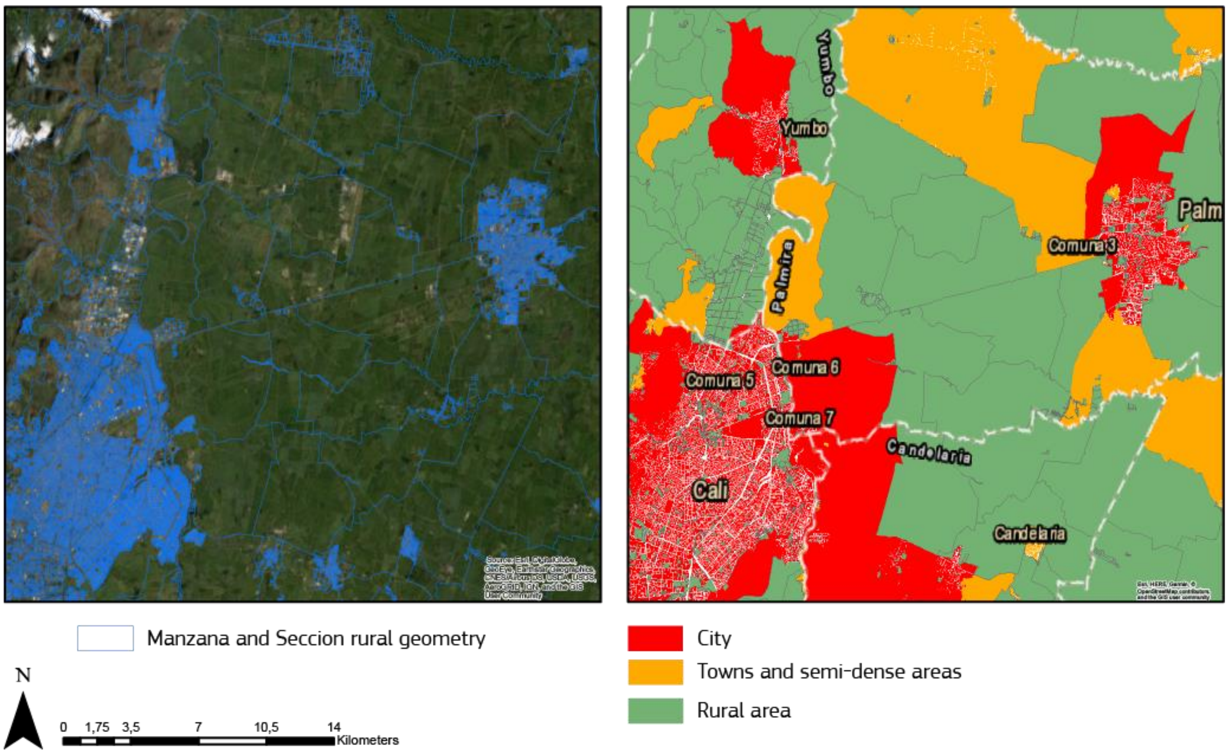


Figure 11 Classification of Manzanas and Seccion Rural by Degree of Urbanisation at level 1in the area of Cali

Table 6 Degree of Urbanisation level 1 and level 2 population and area statistics at territorial unit level

Level 1 type	number of units	population	% units	% pop	Level 2 type	number of units	population	% units	% pop
Rural area	151,885	10,278,891	28%	23%	Mostly uninhabited areas	40,386	1,055,683	7%	2%
					Dispersed rural areas	45,541	4,771,144	8%	11%
					Villages	65,958	4,452,064	12%	10%
Towns and suburbs	152,294	9,612,470	28%	21%	Suburban or peri-urban areas	32,527	2,156,241	6%	5%
					Semi-dense towns	21,015	1,332,600	4%	3%
					Dense towns	98,752	6,123,629	18%	14%
City	246,849	24,977,425	45%	56%	City	246,849	24,977,425	45%	56%

3.3.1 Comparing grid based statistics to territorial units statistics

The Degree of Urbanisation is applied in two stages, first at the grid level and second at territorial units level. Population statistics derived from these two processes can vary. This is due to scale (a factor that affect geospatial analysis), but also to the fact that at the grid level, population per class is calculated based only on the cells classified as rural, urban cluster or urban centre. For local units, once the unit is classified based on the population majority at the grid level, all the population in that unit is accounted in that class.

In the case of Colombia, population in grid cells in the urban domain (grid cell codes 2 and 3) is 76.7% compared to 77.9% when the calculation is based on local units. A similar difference is appreciable when we consider classes Urban Centre (55.6%) and Cities (56.7%).

3.3.2 Comparing the national definition with the Degree of Urbanisation

When information available in the census contain the national classification of units in urban and rural domains, it is possible to compare the national classification with the one obtained by applying the Degree of Urbanisation, by doing a cross tabulation. This tabulation intersects the units and their population as classified by the national definition and the Degree of Urbanisation (by population or by number of units). The result is indicative on the level of agreement or divergence between what is classified as urban or rural by the national definition and what it is by the Degree of Urbanisation. This helps understanding how the intermediate category of towns and semi-dense area is generally classified by the national definition. This intermediate category in some contexts may be more likely to be classified as rural, while in others as urban.

Overall, the share of population counted in units classified urban in the census of Colombia is 78.8% compared to 77.1% in the Degree of Urbanisation.

In the case of Colombia (Table 7), 97% of the territorial units classified as rural in the census, are also classified as rural areas by the Degree of Urbanisation, and 78% of the units classified as urban in the census are also in the urban domain according to the Degree of Urbanisation method. In particular, 47.1% of the urban units in the census are cities, and 30.6% are towns and semi-dense areas. Some 22.3% of the units classified as urban in the census are rural in the Degree of Urbanisation, and 2.6% of the units classified as rural in the census are towns and semi-dense area in the Degree of Urbanisation. Only 0.4% of the units are classified rural in the census and are cities in the Degree of Urbanisation.

Table 7 Cross tabulation between the national definition classification and the Degree of Urbanisation classification (Level 1) of territorial units per share of population and share of territorial units:
"how many census defined Urban/Rural (U/R) are classified Rural area/Towns and semi-dense areas/City by Degree of Urbanisation?"

	Population		**Polygons**	
	R	***U***	***R***	***U***
Rural area	75.1%	7.9%	97.0%	22.3%
Towns and semi-dense areas	20.9%	21.6%	2.6%	30.6%
City	4.0%	70.6%	0.4%	47.1%

The match is also high when classifications are compared by population: 92.1% of the population considered urban by the census designation is also classified in the urban domain by the Degree of Urbanisation. About 71% of the population counted as urban by the national definition for the Degree of Urbanisation is in cities, and an additional 21.6% is in towns and semi-dense areas. More than 75% of the population defined rural in the census is counted in rural areas by the Degree of Urbanisation. About 21% of the population classified rural in the census is in the intermediate category towns and semi-dense areas in the Degree of Urbanisation.

The Degree of Urbanisation level 2 offers some insights about the cases in which there are the slight discrepancies (Table 8) mentioned above. About 45% of the population counted in villages in the Degree of Urbanisation is classified urban population in the census; in this case, more than 95% of the territorial units that are villages are classified urban in the census.

In total, the difference between the national definition and the one by the Degree of Urbanisation of the urban population is 0.9% (based on territorial units). The match by population remains high when comparing the results at grid level (difference of 2.1%).

Table 8 Cross tabulation between the Degree of Urbanisation classification (Level 1 and Level 2) and the national definition per share of population and share of territorial units:
"How many classified Rural area/Towns and semi-dense areas/City by Degree of Urbanisation are census defined Urban/Rural (U/R)?"

	Population		**Polygons**	
	R	***U***	***R***	***U***
Rural area	73.3%	26.7%	33.8%	66.2%
Towns and semi-dense areas	21.8%	78.2%	1.0%	99.0%
City	1.6%	98.4%	0.1%	99.9%

	Population		Polygons	
	R	***U***	***R***	***U***
Mostly uninhabited areas	97.9%	2.1%	84.4%	15.6%
Dispersed rural areas	85.1%	14.9%	23.9%	76.1%
Villages	54.8%	45.2%	4.5%	95.5%
Suburban or peri-urban areas	38.3%	61.7%	2.0 %	98.0%
Semi-dense towns	37.1%	62.9%	1.5%	98.5%
Dense towns	12.6%	87.4%	0.5%	99.5%
City	1.6%	98.4%	0.1%	99.9%

3.4 Classification of demographic and housing variables by Degree of Urbanisation

The classification of territorial units allows to relate specific indicators contained in the demographic and housing census to the Degree of Urbanisation classification of the survey/reporting units (this process is named disaggregation by Degree of Urbanisation).

The source dataset contains a number of variables related to demographic (sex and age) and housing attributes (infrastructural endowment). These variables can be disaggregated by Degree of Urbanisation to identify variations across the urban-rural gradient. The disaggregation of indicators by Degree of Urbanisation is particularly relevant to monitor Sustainable Development Goals indicators, and to compare the results across countries. The results presented in the next sub-sections are based on the number of *viviendas* (i.e. dwelling or residence) with access to a specific infrastructure or service. Several SDG targets refer to the proportion of the population with access to a specific infrastructure or service. This information cannot be extracted directly from the information publicly available in Colombia census. Other approaches could attempt to model an average person per *viviendas*, but we do not perform this estimate to reduce assumptions. With the Degree of Urbanisation it is possible to determine which *viviendas* are located in cities, towns and semi-dense areas and rural areas to disaggregate data and eventually compute the indicator in combination with other data available at local level.

3.4.1 Population age and sex by Degree of Urbanisation

Population counts by age and sex are typical census variables. This information is vital to derive population age pyramids, and sex ratio. Both these indicators are widely used across policy sectors (including education, labour, health, etc.).

3.4.1.1 Population age

The population age structure in Colombia varies across the urban-rural gradient (Figure 12). Younger ages (0-9 years and 10-19 years) are dominant in rural areas, and more than half the population in these areas is 29 years or younger. Urban areas have a higher share of the age ranges between 20 and 69 years. Cities host the majority of population groups in the ranges 20-29 years, 30-39 years, 40-49 years, 50-59 years and 60-69 years. Up to 18% of the individuals between 20 and 29 years are in cities. Cities also host the highest share in the range 30-39 years (15.7% compared to 14.1% in towns and semi-dense areas and 13% in rural areas). These characteristics may relate to higher access to opportunities offered by cities (e.g. jobs and education) as reported in literature. Elderly people also account for a considerable share of city population: about 25% of city population is older than 50. Towns and semi-dense areas have a blended pattern between rural areas and Cities. These areas have an higher share of younger generations (<19 years) of about 3% per class compared to cities, but account for a consistently lower share of middle age population (between 20 and 50 years) of about 1% across classes.

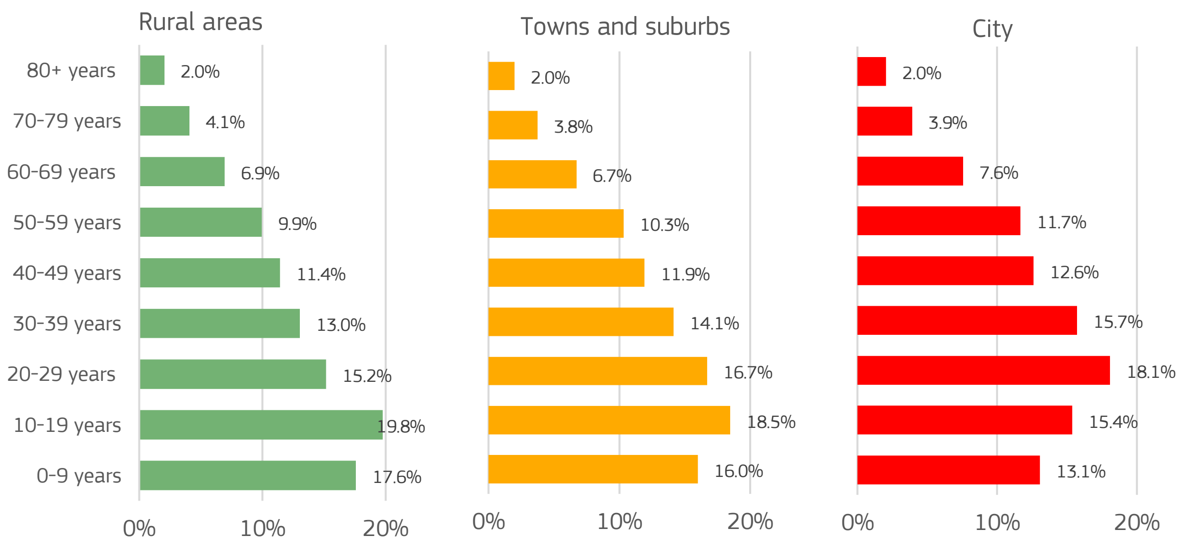


Figure 12 Population age classes by Degree of Urbanisation Level 1

At the second level of the Degree of Urbanisation, the pattern above is more evident in some types of areas. The dominance of the youngest age groups (0-9 and 10-19) emerges in mostly uninhabited areas (class 11), where each of these two classes are above 20% of the population (Table 9). The other classes of the Degree of Urbanisation Level 2 account a higher share of youngest groups in semi-dense towns (22) compared to dense towns (23) –age groups unitl 19 years of age, after the pattern reverses with a higher share of the age groups between 30 and 59. Again, a prevalence of the oldest population groups (60 to 80+) is observed in semi-dense towns compared to dense towns.

Table 9 Population age share by Degree of Urbanisation Level 2 and national average (NAT)

	0-9 years	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	70-79 years	80+ years
Mostly uninhabited areas	20.9%	20.7%	15.3%	12.8%	11.1%	9.1%	5.9%	3.0%	1.3%
Dispersed rural areas	16.9%	19.5%	14.9%	13.0%	11.6%	10.3%	7.3%	4.3%	2.1%
Villages	17.5%	19.8%	15.5%	13.2%	11.3%	9.7%	6.7%	4.1%	2.2%
Suburban or peri-urban areas	15.7%	18.2%	16.7%	14.3%	12.2%	10.4%	6.7%	3.7%	1.9%
Semi-dense towns	16.4%	19.0%	16.1%	13.6%	11.7%	10.1%	6.8%	4.1%	2.2%
Dense towns	16.0%	18.4%	16.8%	14.2%	11.9%	10.3%	6.7%	3.7%	2.0%
City	13.1%	15.4%	18.1%	15.7%	12.6%	11.7%	7.6%	3.9%	2.0%
NAT	*14.7%*	*17.0%*	*17.1%*	*14.8%*	*12.2%*	*11.0%*	*7.2%*	*3.9%*	*2.0%*

3.4.1.2 Population sex

The gender distribution between urban and rural areas in Colombia has a clear split. Male population is the majority in Rural areas (51.7%), and female is the majority in the urban domain (52.1%), and highest in cities (52.5%) compared to a national proportion of 48.7% of males and 51.3% of females.

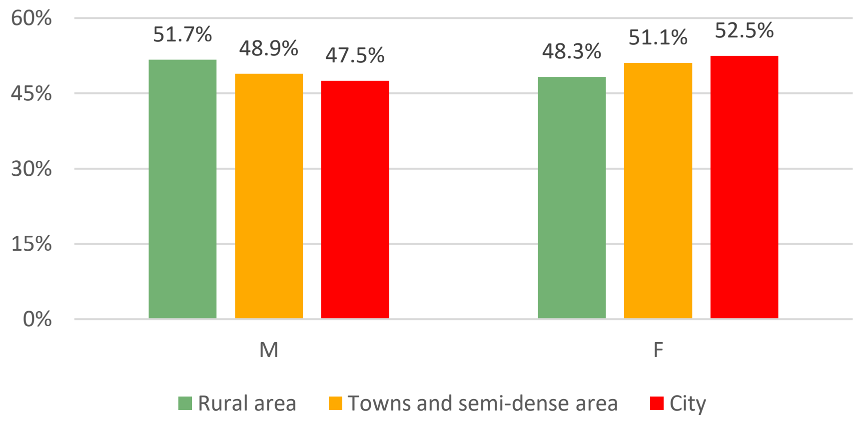


Figure 13 Sex ratio (M males; F females) at Degree of Urbanisation Level 1

At Level 2 of the Degree of Urbanisation, the share of female population is lowest in class 11-mostly uninhabited areas (45.8%), and dominant in classes 21-suburb or peri-urban, 23-dense and 22-semi-dense towns (Table 10).

Table 10 population gender share by Degree of Urbanisation Level 2 and national average (NAT)

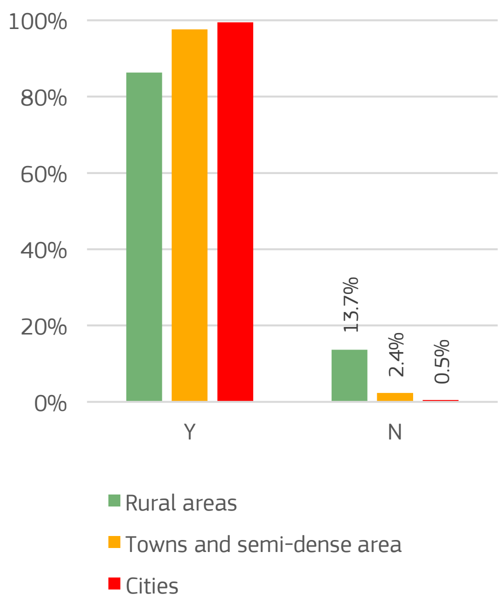
Class	Male	Female
Mostly uninhabited areas	54.2%	45.8%
Dispersed rural areas	52.2%	47.8%
Villages	50.6%	49.4%
Suburban or peri-urban areas	49.5%	50.5%
Semi-dense towns	49.6%	50.4%
Dense towns	48.6%	51.4%
City	47.5%	52.5%
NAT	*48.7%*	*51.3%*

3.4.2 Access to electricity

The Target 7.1 of Sustainable Development Goal 7 on "Affordable and Clean Energy" aims to *ensure universal access to affordable, reliable and modern energy services.* One of the indicators used to monitor this Target assesses the *proportion of population with access to electricity* (7.1.1). In Colombia, knowing the amount of *viviendas* having access to electricity is important information in that perspective, and understanding how the proportion varies by Degree of Urbanisation (cities, towns and semi-dense areas and rural areas) can help meeting the SDG targets and inform policy to speed transition.

In Colombia 96.3% of the *viviendas* are reported to have access to electricity. The Degree of Urbanisation helps disaggregating this national figure to identify variations. Urban areas have a higher access: for example, 99.5% of *viviendas* in cities are reported having access to electricity, compared to 97.6% in towns and semi-dense areas, and to 86.3% in rural areas (10% less compared to the national average) (Figure 14 left).

At the second level of the Degree of Urbanisation, the situation is more nuanced, but access to electricity still increases with urbanization. Mostly uninhabited areas report the lowest national share of *viviendas* with access to electricity (58.7%), while villages have the highest share in the rural domain (87.7%). In the urban domain, Dense towns have a higher share of *viviendas* with access to electricity (98%) compared to suburban or peri-urban areas (97.2%) and semi-dense towns (96.4%), but lower than cities (99.5%) (Figure 14 right).



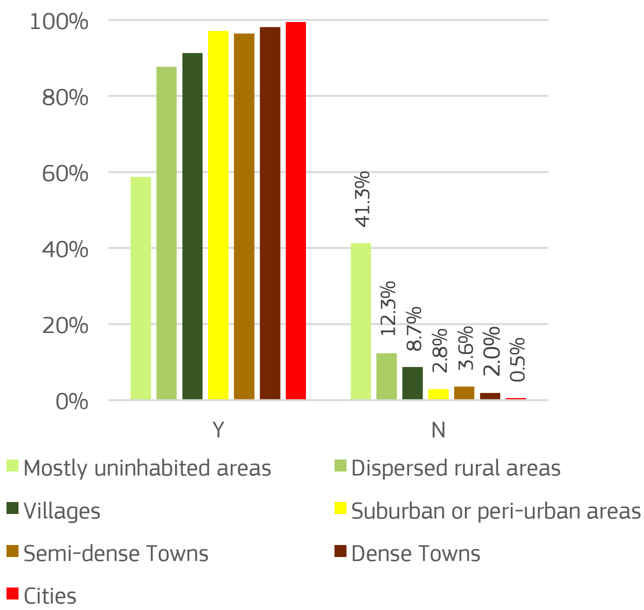


Figure 14 Share of viviendas with access (Y) and no access (N) to electricity by Degree of Urbanisation Level 1(left) and Level 2 (right)

3.4.3 Access to internet

Access to internet is a growing need to exploit the opportunities brought by digitalisation. Also in the SDG context specific goals aim at promoting peoples' access to internet connectivity with broadband and mobile phones (SDG 9 and SDG 17). The census of Colombia counts the number of *viviendas* with access to internet, and at national level 41.5% of the *viviendas* are reported with access and 57.6% with no access (0.9% of the *viviendas* are reported with an undetermined access). However, analysis by Degree of Urbanisation shows striking variations in this access. More than 60% of the *viviendas* in cities have access to the internet, while this share drops to 21.6% in towns and semi-dense areas and to 5.3% in rural areas.

The second level of the Degree of Urbanisation shows the lowest access to internet in mostly uninhabited areas (1.8% of the *viviendas*), and a significant increase in villages (6.9%). In the urban domain, access to internet in suburban or peri-urban areas is almost double the one in semi-dense towns (21% and 11% respectively) and increases to almost ¼ of the *viviendas* in dense towns. Access to internet in cities is above 60%, so more than 20% higher compared to national average.

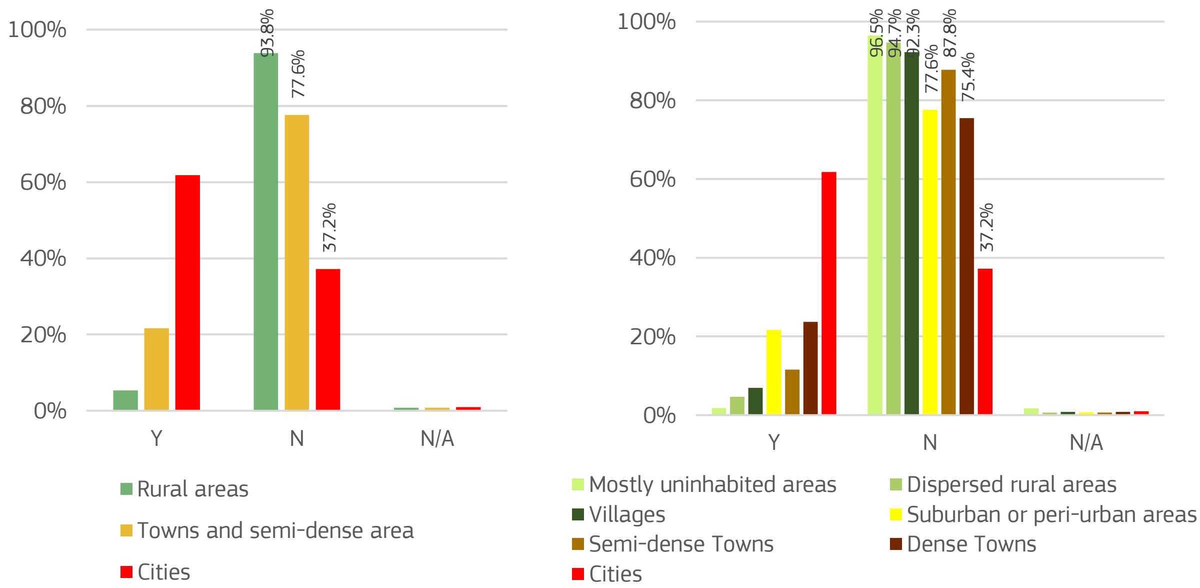


Figure 15 Share of *viviendas* with access (Y) no access (N) and not available (N/A) to Internet by Degree of Urbanisation Level 1(left) and Level 2 (right)

3.4.4 Access to water

The Target 6.1 of the Sustainable Development Goal 6. "Ensure availability and sustainable management of water and sanitation for all" aims at *by 2030, achieve universal and equitable access to safe and affordable drinking water for all.*

The 2018 census of Colombia reports the number of *viviendas* with access to water. Combining this data with the Degree of Urbanisation shows a split in access to water at Level 1, with access in 54.3% of the *viviendas* in rural areas, 85.6% of the *viviendas* in towns and semi-dense areas, and 97.1% of the *viviendas* in cities. The national average is 85.6%. Also in the case of access to water, cities are much more endowed compared to other areas. In rural areas, more than 45% of *viviendas* have no access to water.

The Degree of Urbanisation Level 2 further disaggregates this pattern, showing that less than 20% of the *viviendas* in mostly uninhabited areas have access to water. The share of serviced *viviendas* increases substantially in dispersed rural areas (50.2%) and in villages, where more than 65% of the *viviendas* have access to water (Figure 16). In towns and semi-dense area and suburban and peri-urban areas the access to water is about 81% compared to 88% in Dense towns.

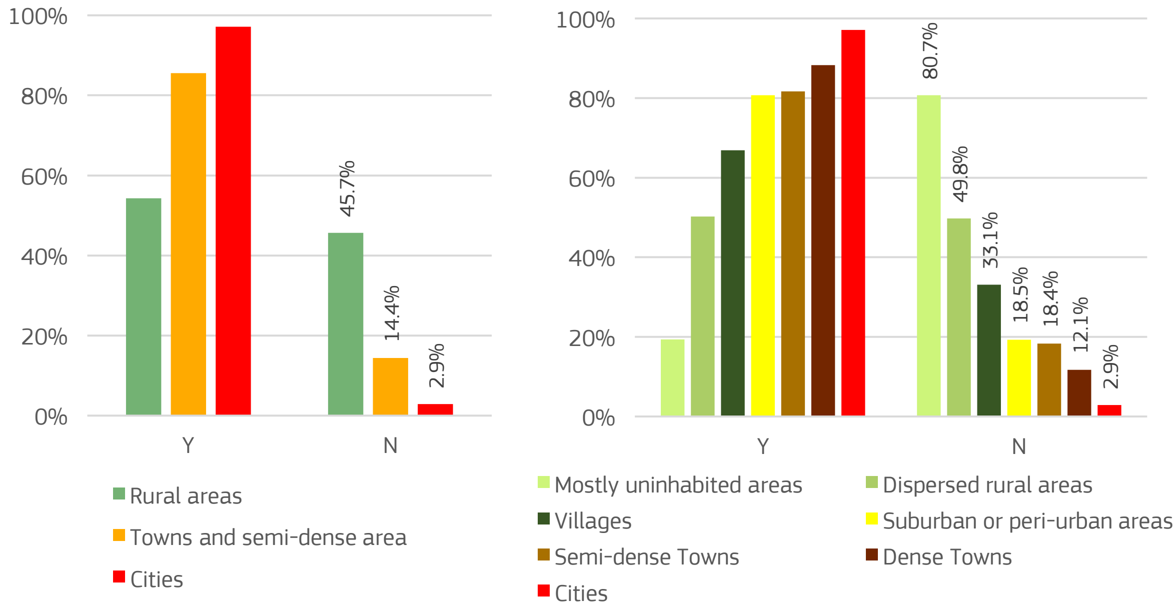


Figure 16 Share of viviendas with access (Y) and no access (N) to water by Degree of Urbanisation Level 1(left) and Level 2 (right)

3.4.5 Access to the sewage system

In the same policy framework in which access to water is a key target, also access to sanitation plays an important role (SDG 6). SDG Target 6.2 strives to ensure, by 2030, *access to adequate and equitable sanitation and hygiene for all and end open defecation, paying special attention to the needs of women and girls and those in vulnerable situations*. The 2018 Colombian census counts the number of *viviendas* with and without access to a sewage system. At the national level, 74.6% of the *viviendas* are connected to such systems. Analysing this by Degree of Urbanisation shows that the share of *viviendas* that are connected is much higher in cities (95.1%) compared to towns and semi-dense areas (68.3%) and rural areas (24.6%) (Figure 17).

At the second level of the Degree of Urbanisation, in the rural domain villages are the most connected (39.9%), while in dispersed rural areas only 15% of the *viviendas* are connected, and in mostly uninhabited areas this value drops to 5%. In the urban domain, dense and semi-dense towns have more *viviendas* connected to (75.5% and 56.4% respectively) compared to the ones in suburban or peri-urban areas (54.9%). Accordingly, the most significant differences are observed between dispersed rural areas and villages, and between suburban or peri-urban areas and cities.

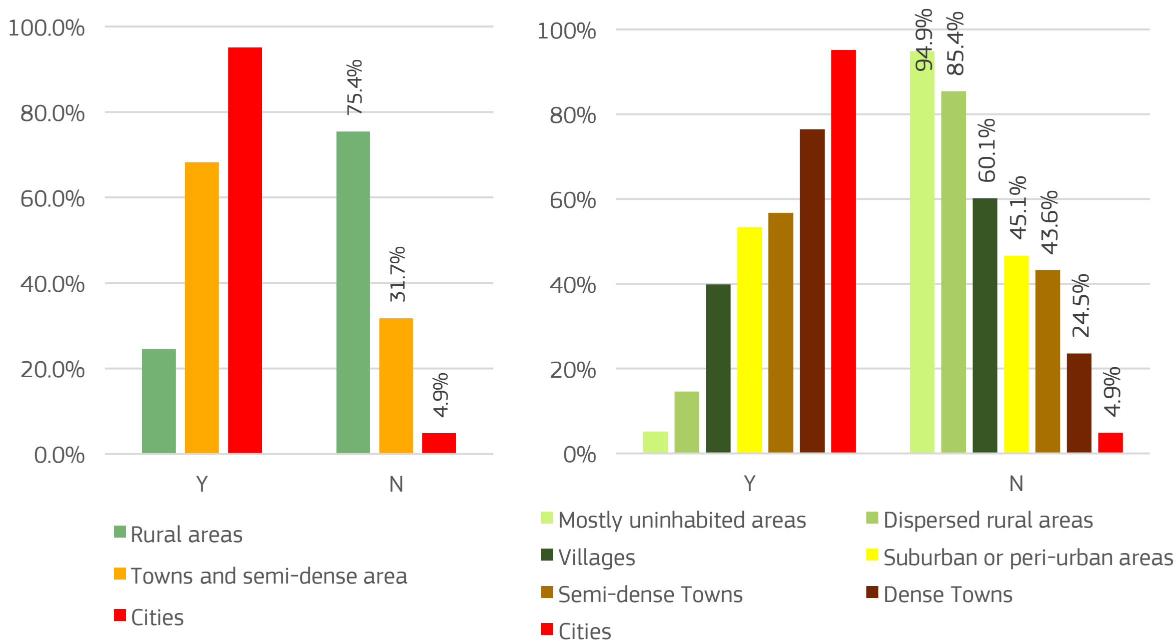


Figure 17 Share of viviendas connected (Y) or not (N) to a sewage system by Degree of Urbanisation Level 1(left) and Level 2 (right)

3.4.6 Access to waste management

Sustainable Development Goal 11 aiming to "Make cities and human settlements inclusive, safe, resilient and sustainable" has an Indicator (11.6.1) focusing on the share of city solid waste collected and managed in controlled facilities. Beside the need to produce local data for this indicator, it is important to harmonise the ways in which urban and rural areas are defined, to make sure this indicator is consistently reported across countries.

The 2018 census of Colombia reports the number of *viviendas* with access to waste management. Overall 80% of the *viviendas* have access, but this accessibility varies significantly by Degree of Urbanisation. The share is highest in cities (98%), higher than national average in towns and semi-dense areas (81.5%) and about 30% in rural areas.

At the second level of the Degree of Urbanisation, in the rural domain, villages stand out with 46.1% of the *viviendas* with access to waste management compared to less than half this figure in dispersed rural areas (20.1%) and to the 6.1% of *viviendas* in mostly unpopulated areas. In the urban domain, the share of *viviendas* with access to waste management is higher in suburban or peri-urban areas (73.2%) compared to semi-dense towns (66.7%), but lower compared to dense towns where more than 87.5% of the *viviendas* are served.

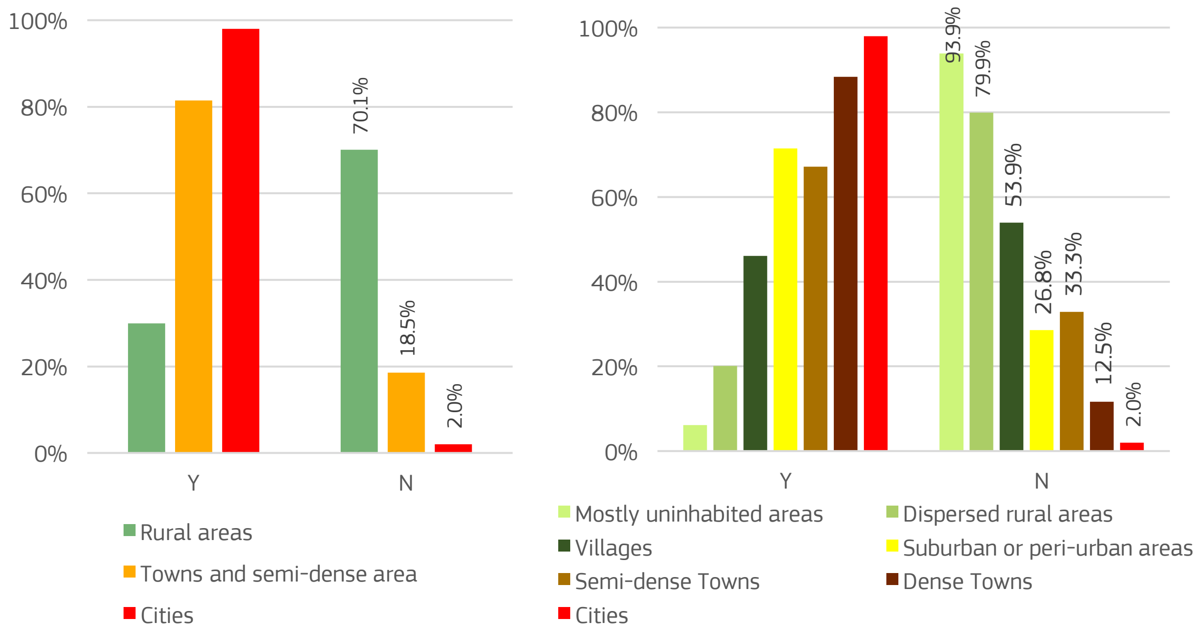


Figure 18 Share of *viviendas* with access (Y) and no access (N) to waste management by Degree of Urbanisation Level 1(left) and Level 2 (right)

4 Conclusion

The Degree of Urbanisation method to delineate urban and rural areas for international statistical comparison was endorsed and recommended at the 51st session of the United Nations Statistical Commission to support the monitoring of the Sustainable Development Goals. This report contains the first application of the Degree of Urbanisation method to a post 2015 national demographic and housing survey (census) with tools developed in the framework of the Global Human Settlement Layer. The showcase is made possible by the open data policy of the *Departamento Administrativo Nacional de Estadística of Colombia* (DANE) that makes fully available on the internet the 2018 census information in geospatial and statistical format.

The report presents the whole process to apply the Degree of Urbanisation method to census data: first by generating a population grid with the GHS-POP2G Tool using census geospatial data and the most updated GHSL built-up layer to date, obtained by processing Copernicus Sentinel 2 imagery; second, by producing a settlement classification grid using the GHS-DUG Tool based on the population grid obtained in the previous step; third, by classifying local units (in this report the finest census units available, *Manzana* and *Seccion Rural*) by Degree of Urbanisation; and fourth, by disaggregating other demographic and housing indicators contained in the census by Degree of Urbanisation (population sex and age structure, and *viviendas* access to electricity, internet, water, a sewage system and waste management).

The results show a very close match between the classification of local units based on the national definition or urban/ural areas and the Degree of Urbanisation method, and a gradient between urban and rural areas across all the variables analysed by Degree of Urbanisation.

With the methods and application framework presented in this report, the aim is to explain the practical implementation of the Degree of Urbanisation method to census rounds and demographic and housing surveys. The report serves as implementation extension to the report “Applying the Degree of Urbanisation - a methodological manual to define cities, towns and rural areas for international comparisons” jointly developed by the European Commission, FAO and OECD (European Commission and Statistical Office of the European Union 2021).

5 References

Corbane, Christina, Vasileios Syrris, Filip Sabo, Panagiotis Politis, Michele Melchiorri, Martino Pesaresi, Pierre Soille, and Thomas Kemper. 2020. "Convolutional Neural Networks for Global Human Settlements Mapping from Sentinel-2 Satellite Imagery." *Neural Computing and Applications*, October. https://doi.org/10.1007/s00521-020-05449-7.

Dijkstra, Lewis, Aneta J. Florczyk, Sergio Freire, Thomas Kemper, Michele Melchiorri, Martino Pesaresi, and Marcello Schiavina. 2020. "Applying the Degree of Urbanisation to the Globe: A New Harmonised Definition Reveals a Different Picture of Global Urbanisation." *Journal of Urban Economics*, November, 103312. https://doi.org/10.1016/j.jue.2020.103312.

European Commission, and Statistical Office of the European Union. 2021. *Applying the Degree of Urbanisation - a Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons*.

Maffenini, L., M. Schiavina, S. Freire, M. Melchiorri, M. Pesaresi, and T. Kemper. 2020. *GHS-POP2G User Guide*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/500887.

Maffenini, L., M. Schiavina, M. Melchiorri, M. Pesaresi, and T. Kemper. 2020a. *GHS-DUG User Guide*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/916217.

———. 2020b. *GHS-DU-TUC User Guide*. Luxembourg: Publications Office of the European Union. https://data.europa.eu/doi/10.2760/633624.

OECD, and European European Commission. 2020. *Cities in the World*. https://www.oecd-ilibrary.org/content/publication/d0efcbda-en.

UN-GGIM. 2020. *Future Trends in Geospatial Information Management: The Five to Ten Year Vision*. United Nations.

Wardrop, N. A., W. C. Jochem, T. J. Bird, H. R. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman, and A. J. Tatem. 2018. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *Proceedings of the National Academy of Sciences* 115 (14): 3529–37. https://doi.org/10.1073/pnas.1715305115.

List of abbreviations and definitions

DANE	Departamento Administrativo Nacional de Estadística of Colombia
DEGURBA	Degree of Urbanisation
EUROSTAT	The statistical office of the European Union
F	Females
GHS-BUILT	Multitemporal information layer on built-up area presence as derived from satellite data, from GHSL
GHS-DUG	GHSL Degree of Urbanisation Grid Tool
GHS-DU-TUC	GHSL Degree of Urbanisation – Territorial Units Classifier Tool
GHSL	Global Human Settlement Layer
GHS-POP	Spatial raster dataset that depicts the distribution of population, expressed as the number of people per cell, from GHSL
GHS-POP2G	GHSL Population to Grid Tool
GHS-SMOD	Settlement classification raster dataset applying the Degree of Urbanisation method to GHS-POP and GHS-BUILT data
M	Males
N	Viviendas with no access to the infrastructure
N/A	Accessibility to infrastructure data not available in the census
NAT	National average
R	Rural
RAM	Random-access memory
SDG	Sustainable Development Goals
U	Urban
UID	Unique Identifiers
UN-GGIM	United Nations Committee of Experts on Global Geospatial Information Management
Y	viviendas with access to the infrastructure

List of GHSL Resources

GHS-BUILT	Dataset information: https://ghsl.jrc.ec.europa.eu/ghs_bu_s2_2018.php
GHS-DUG	Tool: https://ghsl.jrc.ec.europa.eu/tools.php
	User Guide: https://ghsl.jrc.ec.europa.eu/documents/GHS-DUG_4_User_Guide.pdf?0.6375801745442322
GHS-DU-TUC	Tool: https://ghsl.jrc.ec.europa.eu/tools.php
	User Guide: https://ghsl.jrc.ec.europa.eu/documents/GHS-DUTUC_User_Guide.pdf?0.6827332891117774
GHS-POP	Dataset information: https://ghsl.jrc.ec.europa.eu/ghs_pop2019.php
GHS-POP2G	Tool: https://ghsl.jrc.ec.europa.eu/tools.php
	User Guide: https://ghsl.jrc.ec.europa.eu/documents/GHS-POP2G_2_User_Guide.pdf?0.9026048752029705
GHS-SMOD	Dataset information: https://ghsl.jrc.ec.europa.eu/ghs_smod2019.php

List of figures

List of tables

KJ-NA-30692-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/939289

ISBN 978-92-76-37556-2