European Commission

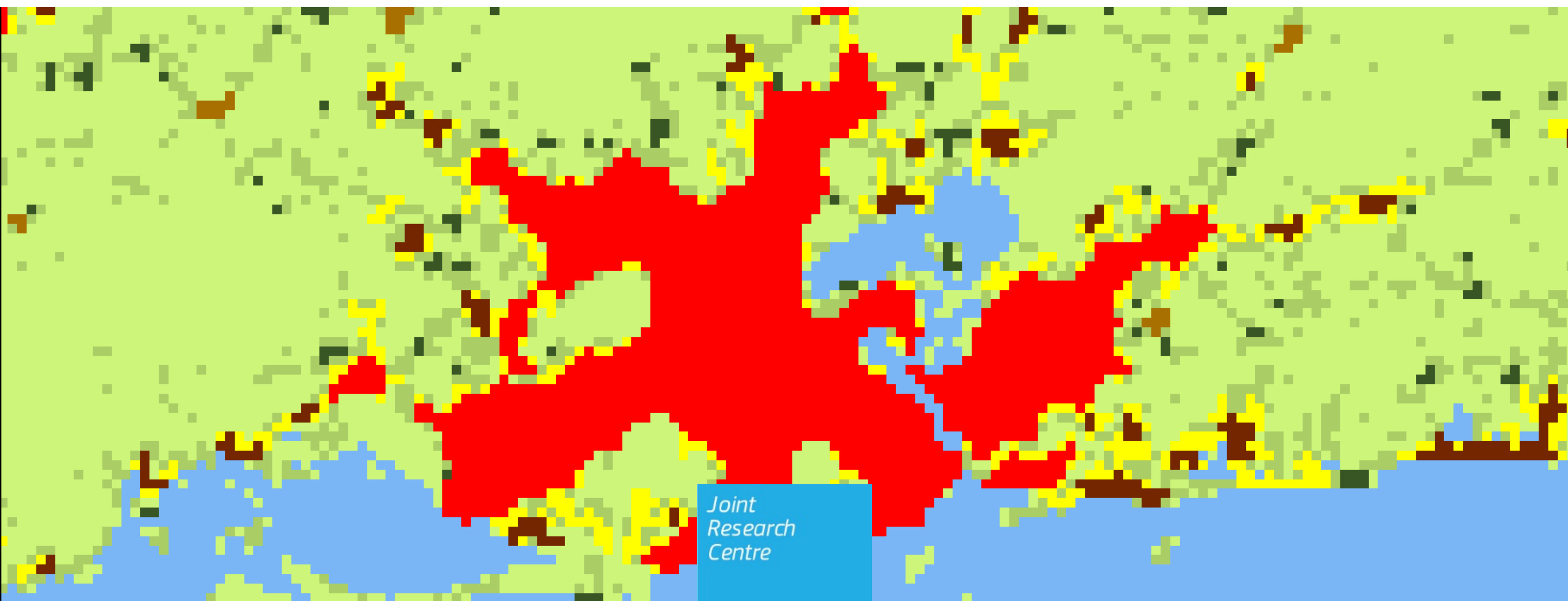
JRC SCIENTIFIC INFORMATION SYSTEMS AND DATABASES REPORT

GHS-DUG User Guide

Degree of Urbanisation Grid
User Guide
Version 4

Maffenini, L.; Schiavina, M.; Melchiorri, M.;
Pesaresi, M.; Kemper T.

2020



Contact information
Name: Thomas Kemper
Address: Via Fermi, 2749 21027 ISPRA (VA) - Italy - TP 267
European Commission, Joint Research Centre (JRC), Ispra
Email: Thomas.kemper@ec.europa.eu
Tel.: +39 0332 785576

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

JRC121484

PDF	ISBN 978-92-76-22416-7	doi:10.2760/916217
Print	ISBN 978-92-76-22415-0	doi:10.2760/645307

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Maffenini, L., Schiavina, M., Melchiorri, M., Pesaresi, M. and Kemper, T., GHS-DUG User Guide, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-22416-7, doi:10.2760/916217, JRC121484

Contents

Authors

Maffenini, Luca[b]; Schiavina, Marcello[a]; Melchiorri, Michele[c]; Pesaresi, Martino[a]; Kemper Thomas[a]

[a] European Commission-Joint Research Centre, Ispra, Italy
[b] UniSystems Luxembourg SàRL
[c] Engineering S.p.a

The following institutions provided useful reviews to improve the usability of the tool: Statistics South Africa (Policy Research & Analysis), EUROSTAT Unit E-4 (Regional statistics and geographical information), UN-Habitat (Research and Capacity Development Branch), Federal Competitiveness and Statistics Authority (United Arab Emirates)

Abstract

The Degree of Urbanisation Grid (GHS-DUG) Tool (– version 4) is an information system developed in the framework of the Global Human Settlement Layer (GHSL) to produce geospatial grids to map settlement classes and extract related statistics. The settlement classes are derived from the *"Degree of Urbanisation"* method and ported to the GHSL environment through the *GHSL Settlement Model (GHSL SMOD).*

The GHS-DUG 4 is designed as a scalable tool allowing the application of the *GHSL Settlement Model* to the input data available to the user or to data made available in the GHSL Data Package 2019. This document contains the description of the GHS-DUG Tool use, the rationale of the differentiation between settlement classes and the comprehensive description of the outputs. The tool is a capacity enhancement asset in the framework of the multi-stakeholder effort for the uptake of the Degree of Urbanisation, the people-based harmonised definition of cities and settlements recommended by the 51st Session of the United Nations Statistical Commission as the method to delineate cities and rural areas for international statistical comparison.

The GHS-DUG, as all GHSL Tools, is issued with an end-user licence agreement, included in the download package.

1 Introduction

The Global Monitoring Framework of the 2030 Agenda for Sustainable Development includes several indicators that require disaggregation in urban and rural classes, and several others that are sensitive to how an urban area is delineated. The broad array of different criteria applied in national definitions of rural-urban areas poses serious challenges to cross-country comparisons (ILO 2018). Both the Action Framework of the Implementation of the New Urban Agenda (UN-Habitat 2017) and the Global Strategy to improve Agricultural and Rural Statistics (GSARS 2018) highlight the need for a harmonised method to facilitate international comparisons and to improve the quality of rural and urban statistics in support of national policies and investment decisions.

Under these circumstances the European Union, The Food and Agriculture Organization of the United Nations (FAO), the International Labour Office (ILO), the Organisation for Economic Co-operation and Development (OECD), UN-Habitat and the World Bank have joined forces to develop a new method to delineate cities, urban and rural areas in a harmonised way. This work was launched at the Habitat III conference in 2016 with the explicit aim to organise global consultations and present the new method to the UN Statistical Commission for endorsement. The 51st Session of the United Nations Statistical Commission recommended this people-based harmonised definition of cities and settlements called "*Degree of Urbanisation*" as the method to delineate cities and rural areas for international statistical comparison.

The goal is to facilitate international statistical comparisons of the performance of cities, urban and rural areas across a selection of global indicators. This method is meant to complement and not replace existing national definitions. National definitions typically rely on a much wider set of indicators and can be adjusted to take into account specific national characteristics. This richness makes national definitions more suitable for national policies, but less suitable for international or global comparisons.

The "*Degree of Urbanisation*" (DEGURBA) is a statistical classification method originally introduced by the European Commission, Directorate of Regional and Urban Policy and successively refined by the joint work of the EU, the Food and Agriculture Organization of the United Nations (FAO), the International Labour Office (ILO), the Organisation for Economic Co-operation and Development (OECD), UN-Habitat and the World Bank. The DEGURBA aims to identify the spatial extents of "Urban Centres", "Urban Clusters" and "Rural Areas" based on resident population density grids at a scale (spatial detail or resolution) of 1 square kilometre (Dijkstra and Poelman, 2014). Following the methodology indicated by the statistical office of the European Union (EUROSTAT), the population grids classified by the DEGURBA method can be used for statistical characterisation of local administrative units[1].

The *GHSL Settlement Model* (GHSL SMOD) is the porting of the DEGURBA in the Global Human Settlement Layer (GHSL) framework developed by the European Commission, Joint Research Centre[2]. The GHSL SMOD supports the international multi-stakeholder discussion on the DEGURBA operationalization parameters and on the DEGURBA derived metrics and indicators using the GHSL baseline information as common global data frame (European Commission , Joint Research Centre, 2018) (Melchiorri et al., 2018) (Corbane et al., 2018) (Melchiorri et al., 2019)

The "*Degree of Urbanisation Grid*" GHS-DUG 4 described in this report is a standalone free software tool incorporating the latest developments of the GHSL SMOD refining the DEGURBA classification in seven distinct settlement classes sub-characterizing at a second hierarchical level the urban cluster and the rural spatial domains of DEGURBA. The GHS-DUG 4 is designed as a scalable tool allowing the application of the GHSL SMOD method to the input data available to the user to generate at the second hierarchical level the settlement entities (Urban Centres, Dense Urban Clusters, Semi-Dense Urban Clusters and Rural Clusters) and the settlement classification grid documented in the GHSL Data Package 2019 (Florczyk et al., 2019).

Compared to DUG 3.1, the GHS-DUG version 4 introduce usability improvements (e.g. ArcGIS toolbox version) and options simplifications.

[1] https://ec.europa.eu/eurostat/web/degree-of-urbanisation/background
[2] https://ghsl.jrc.ec.europa.eu/

1.1 GHSL Settlement model (GHSL SMOD)

The "*Degree of Urbanisation Grid*" GHS-DUG 4 generates output grids and spatial entities by classifying 1 km^2 grid cells, on the basis of population density, size and contiguity, with the GHSL SMOD classification schema organized in two separate hierarchical levels.

The mandatory input for the GHSL SMOD is a 1 km^2 population grid in selected coordinates systems (see section 4.1.1). Other sources are optionally required according to their availability and user needs. Each grid cell has the same shape and size, thereby avoiding distortions caused by using units varying in shape and size. This is a considerable advantage when compared to methods based on the population size and density of local administrative units.

At the first hierarchical level, the GHSL SMOD classifies the 1 km^2 grid cells by identifying the following spatial entities: a) "Urban Centre", b) "Urban Cluster" and classifying all the other cells as "Rural Grid Cells".

The criteria for the definition of the spatial entities at the first hierarchical level are:

- ***"Urban Centre" (also "High Density Cluster" - HDC)*** - An urban centre consists of contiguous grid cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km^2 of permanent land, and has at least 50,000 inhabitants in the cluster with smoothed boundaries (3-by-3 majority filtering) and <15 km^2 holes filled[3];
- ***"Urban Cluster" (also "Moderate Density Cluster" - MDC)*** - An urban cluster (or moderate density clusters) consists of contiguous grid cells (8-connectivity cluster) with a density of at least 300 inhabitants per km^2 of permanent land and has a population of at least 5,000 in the cluster. (The urban centres are subsets of the corresponding urban clusters).

The ***"Rural grid cells" (also "Mostly Low Density Cells" - LDC)*** are all the other cells that do not belong to an Urban Cluster. Most of these will have a density below 300 inhabitants per km^2 (grid cell); some may have a higher density, but they are not part of cluster with sufficient population to be classified as an Urban Cluster.

The settlement grid at level 1 represents these definitions on a layer grid. Each pixel is classified as follow:

— ***Class 3: "Urban Centre grid cell"***, if the cell belongs to an Urban Centre;

— ***Class 2: "Urban Cluster grid cell"***, if the cell belongs to an Urban Cluster and not to an Urban Centre;

— ***Class 1: "Rural grid cell"***, if the cell does not belong to an Urban Cluster.

The second hierarchical level of the GHSL SMOD is a refinement of the DEGURBA set up to identify smaller settlements. It follows the same approach based on population density, population size and contiguity with a **nested classification** into the first hierarchical level. At the second hierarchical level, the GHSL SMOD classifies the 1 km^2 grid cells by identifying the following spatial entities: a) "Urban Centres" as at the first level; b) "Dense Urban Cluster" and c) "Semi-dense Urban Cluster" as parts of the "Urban Cluster", classifying all the other cells of "Urban Clusters" as "Suburban or peri-urban grid cells"; and identifying d) "Rural Cluster" within the "Rural grid cells". All the other cells belonging to the "Rural grid cells" are classified as "Low Density grid cells" or "Very Low Density grid cells" according to their cell population (Figure 1 and Table 1).

Here are reported the definition of the spatial entities at the second hierarchical level (Figure 1):

- ***"Urban Centre" (also "Dense, Large Settlement")*** - An urban centre consists of contiguous grid cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km2 of permanent land, and has at least 50,000 inhabitants in the cluster with smoothed boundaries (3-by-3 majority filtering) and <15 km2 holes filled[3];
- ***"Dense Urban Cluster" (also "Dense, Medium Cluster")*** - A Dense Urban Cluster consists of contiguous cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km^2 of permanent land and has at least 5,000 and less than 50,000 inhabitants in the cluster;
- ***"Semi-dense Urban Cluster" (also "Semi-dense, Medium Cluster")*** - A Semi-dense Urban Cluster consists of contiguous grid cells (8-connectivity cluster) with a density of at least

[3] In a few countries with relatively low-density urban development and a strong separation of land use functions, the Degree of Urbanisation generates multiple urban centres for a single city. Creating urban centres using both cells with a density of at least 1,500 inhabitants and cells that are at least 50% built-up on permanent land resolves this issue (Optional flag is provided in GHS-DUG). Such highly built-up cells typically contain office parks, shopping malls, factories and transport infrastructure.

300 inhabitants per km^2 of permanent land, has at least 5,000 inhabitants in the cluster and is at least 3-km away from other Urban Clusters[4];

- ***"Rural cluster" (also "Semi-dense, Small Cluster")*** - A Rural Cluster consists of contiguous cells (8-connectivity cluster) with a density of at least 300 inhabitants per km^2 of permanent land and has at least 500 and less than 5,000 inhabitants in the cluster.

The ***"Suburban or peri-urban grid cells" (also Semi-dens grid cells)*** are all the other cells that belong to an Urban Cluster but are not part of a Urban Centre, Dense Urban Cluster or a Semi-dense Urban Cluster.

The ***"Low Density Rural grid cells" (also "Low Density grid cells")*** are Rural grid cells with a density of at least 50 inhabitants per km^2 of permanent land and are not part of a Rural Cluster.

The ***"Very low density rural grid cells" (also "Very Low Density grid cells")*** are cells with a density of less than 50 inhabitants per km^2 of permanent land.

The GHSL SMOD classifies as ***"Water grid cells"*** all the cells with more than 0.5 share covered by permanent surface water that are not populated nor built.

The settlement grid at level 2 represents these definitions on a single layer grid. Each pixel is classified as follow:

— ***Class 30: "Urban Centre grid cell"***, if the cell belongs to an Urban Centre spatial entity;

— ***Class 23: "Dense Urban Cluster grid cell"***, if the cell belongs to a Dense Urban Cluster spatial entity;

— ***Class 22: "Semi-dense Urban Cluster grid cell"***, if the cell belongs to a Semi-dense Urban Cluster spatial entity;

— ***Class 21: "Suburban or per-urban grid cell"***, if the cell belongs to an Urban Cluster cells at first hierarchical level but is not part of a Dense or Semi-dense Urban Cluster;

— ***Class 13: "Rural cluster grid cell"***, if the cell belongs to a Rural Cluster spatial entity;

— ***Class 12: "Low Density Rural grid cell"***, if the cell is classified as Rural grid cells at first hierarchical level, has more than 50 inhabitant and is not part of a Rural Cluster;

— ***Class 11: "Very low density rural grid cell"***, if the cell is classified as Rural grid cells at first hierarchical level, has less than 50 inhabitant and is not part of a Rural Cluster;

— ***Class 10: "Water grid cell"***, if the cell has 0.5 share covered by permanent surface water and is not populated nor built.

[4] Measured as outside a buffer of three grid cells of 1 km^2 around dense urban clusters and urban centres.

Figure 1 Schematic overview of GHSL SMOD entities workflow logic. "xpop" represents the population abundance per grid cell; "xpop_dens" represents the population density on permanent land; "xbu" represents the built-up density per grid cell; "xbu_dens" represents the built-up density on permanent land. "DENSITY ON LAND" process fill built-up cells on water with max between 0.5 and built-up surface value and population on water with global average built-up per capita. (*) this procedure of enforcement logic allows the delineation of Urban Clusters Entities which contains by definition the Urban Centres and all 2X classes. Each entity has a corresponding vector boundary.

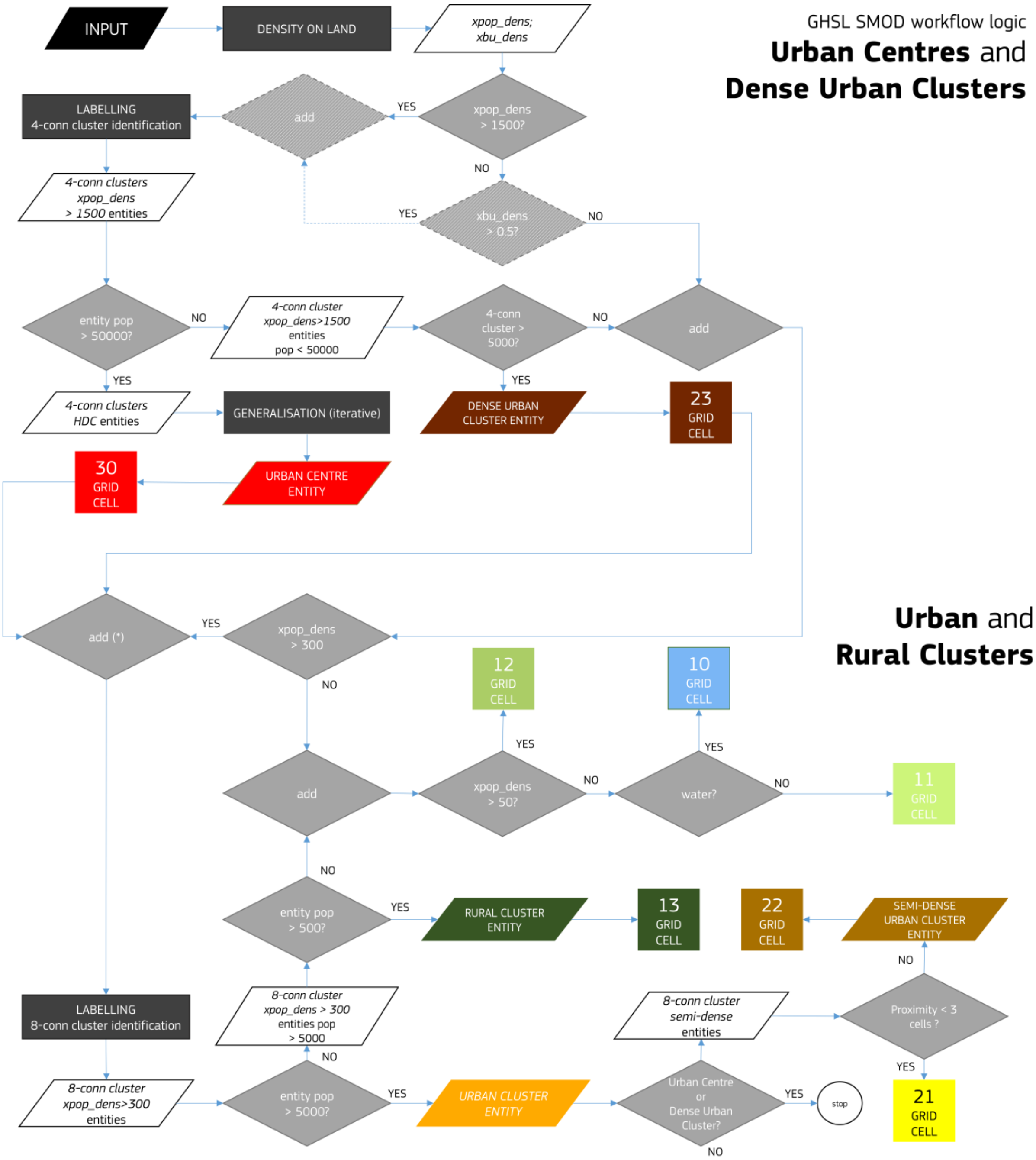
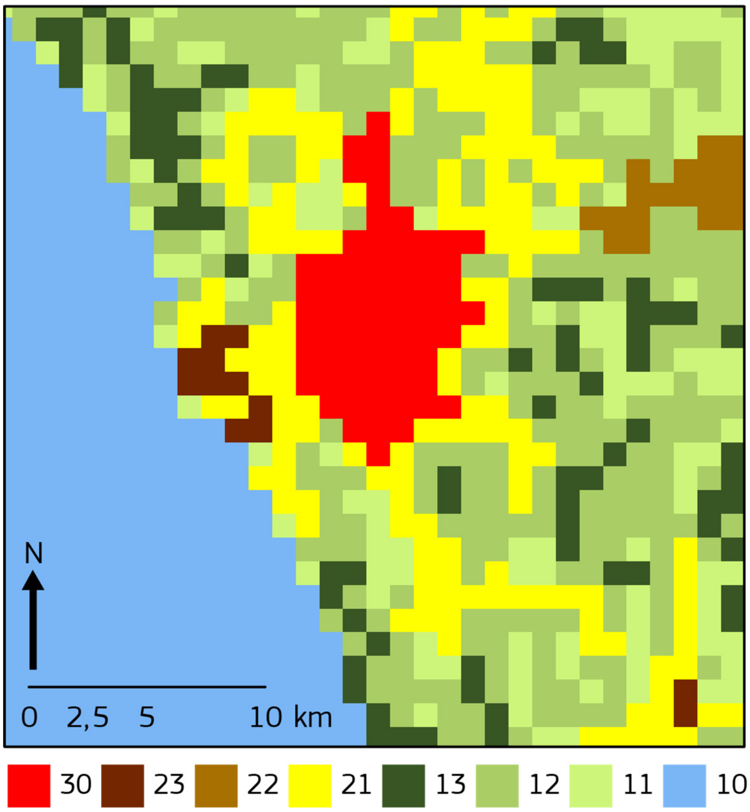


Table 1 Logic of the GHSL SMOD second hierarchical level

Population density of cells, inhabitants per km^2	Minimum population size of the cluster of cells (settlement size): >50,000	50,000 - 5,000	5,000 - 500	No minimum population size criterion (not an entity)
>1500	Urban Centre	Dense Urban Cluster	Rural Cluster	
300 - 1500		Semi-Dense Urban Cluster	Rural Cluster	Suburban or peri urban grid cells
300 - 50				Low Density Rural grid cells
<50				Very Low Density Rural grid cells

Figure 2 Output of the GHSL SMOD implemented by the GHS-DUG 4 in the north of Malaysia displaying settlement class per grid cell



2 Standalone tool

2.1 Installation

2.1.1 System requirements

The Degree of Urbanisation Grid (GHS-DUG) is graphical tool developed in MATLAB and created for Windows, and its system requirements are listed below (Table 2)

Table 2 System requirements

Requirements	Operating system	Processor	RAM	Disk space
Minimum	Windows 7	Any Intel or AMD x86-64 processor	6 GB	150 Mb (+ 700 Mb for Matlab Runtime)
Recommended	Windows 10	Any Intel or AMD x86-64 processor	16 GB	150 Mb (+ 700 Mb for Matlab Runtime)

GHS-DUG operates on raster grids: the larger the grid the higher the required memory. Few examples of RAM requirements in the most computational expensive scenario are below in Table 3. The input raster size (in number of 1 km grid cells) is intended per each of the raster layer(s).

Table 3 Requirements by input raster size

Input raster size	RAM needed	Disk space for outputs
1000x1000 pixels	6 GB	< 10 MB
5000x5000 pixels	12 GB	< 50 MB

2.1.2 Installation procedure

2.1.2.1 Install MATLAB Runtime

GHS-DUG is developed in MATLAB and the MATLAB Runtime is required to run the tool.

The specific version to be installed is **R2018b (9.5)**, Windows 64-bit version.

Check for the official MATLAB website to download and install:

https://it.mathworks.com/products/compiler/matlab-runtime.html

2.1.2.2 Install GHS-DUG

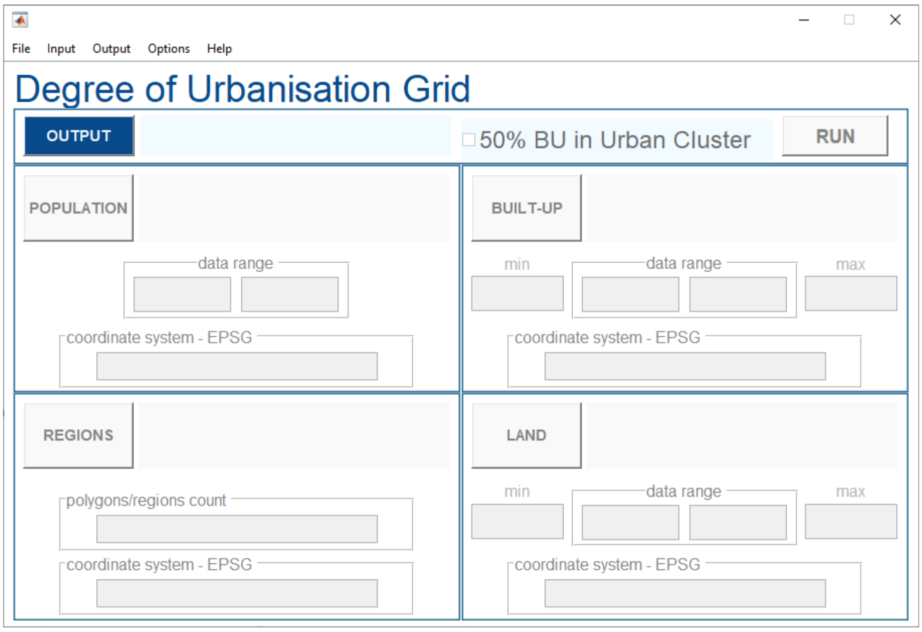
Once the MATLAB Runtime is installed, if you have not done so already, download the GHS-DUG 4 tool from the GHSL website (Tools section):

https://ghsl.jrc.ec.europa.eu/tools.php

Unzip the GHS-DUG.zip file and run the executable directly by double click the *GHS-DUG.exe* file.

The first time GHS-DUG is launched on a new computer it performs several checks and initialization steps, therefore it can take up to 2 minutes to start.

Figure 3 First startup, GUI



2.1.2.3 Remove GHS-DUG

To uninstall GHS-DUG simply delete the entire folder containing the *GHS-DUG.exe* file.

MATLAB Runtime can also be removed if not necessary for other applications.

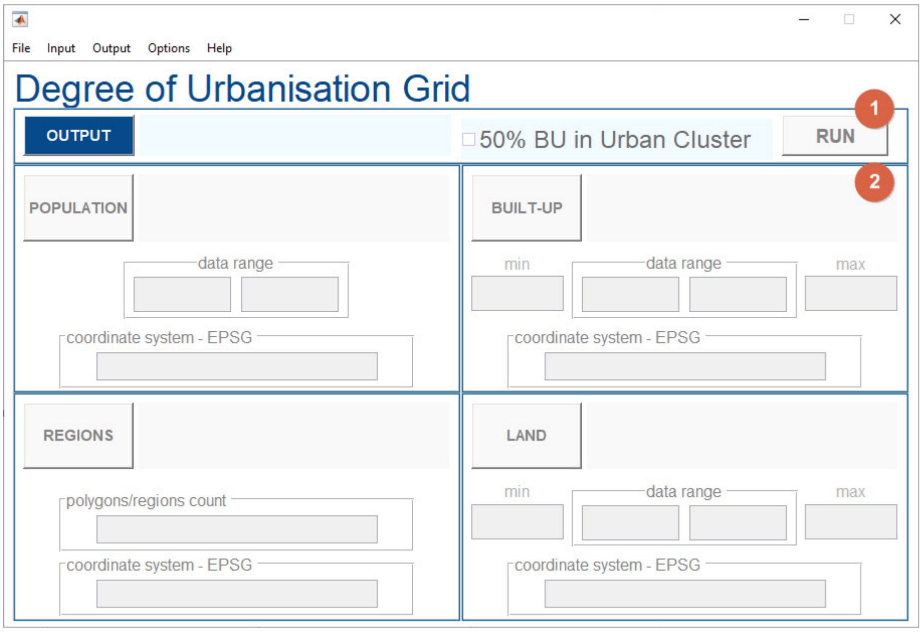
2.2 The graphical user interface

The GHS-DUG graphical user interface is created in MATLAB and is distributed as a standalone application.

The main window is divided in two areas:

1. Top panel, which contains the buttons to select the output folder, activate the 50% built-up rule and start the process;
2. Bottom panel, which contains four areas in a 2x2 grid, which allow to select the input layers: population, land, built-up area and region data;

Figure 4 GHS-DUG main windows and areas



Not all areas are active, some are enabled through the guided run while other explicitly by the user if needed.

After the start, the user is guided in the process of input selection, starting from the output folder, continuing with the population raster and the eventually other grids.

2.2.1 Menu bar

The menu bar is organized as in five sections:

1. File
2. Input
3. Output
4. Options
5. Help

2.2.1.1 File menu

Table 4 File menu bar

Item	Action	Shortcut
Run	Starts the GHSL SMOD computation	CTRL + R
Clear all	Clear all inputs	
Quit	Exit GHS-DUG and close the main window	CTRL + Q

2.2.1.2 Input

Table 5 Input menu bar

Item	Action	Shortcut
Population	Select and load the Population grid	
Built-up	Select and load the Built-up area grid	
Regions	Select and load the Regions file (grid or vector)	
Land	Switch between Land or Water grid	

2.2.1.3 Output

Table 6 Output menu bar

Item	Action	Shortcut
Output	Set the output folder	CTRL + O

2.2.1.4 Options

Table 7 Options menu bar

Item	Action	Shortcut
LAND: Use custom land layer	Activate the land panel to allow the user to provide a custom land layer in input	
Use LAND layer(0 = water --> 100 = land)	Change the values meaning to the LAND format. The value 0 represents 0% land, therefore a pixel full of water; the value 100 means 100% land	
Use WATER layer(0 = land --> 100 = water)	Change the values meaning to the WATER format. The value 0 represents 0% water, therefore a pixel full of land; the value 100 means 100% water	
REGIONS: compute statistics for each polygon	Activate the regions panel to allow the user to provide polygons where statistics are computed	
Open output folder when process completers	Open the output folder in windows explorer at the end of the computation. Default: activated	

2.2.1.5 Help

Table 8 Help menu bar

Item	Action	Shortcut
Online help	Open a web page to the GHS-DUG documentation website using the default browser	CTRL + H
About GHS-DUG	Open a new window containing GHS-DUG version number, release date and contact information	

2.2.2 Buttons

All buttons are located in the main window. Some of them provide the same functionality of the single menu entries and are an alternative way to set input data and start the GHSL SMOD computation.

Table 9 Buttons

Item	Action	Shortcut
OUTPUT	Set the output folder	CTRL + O
50% BU in Urban Cluster	When selected it applies the 50% Built-up in Urban Cluster rule. It needs a Built-up raster provided by the user	
RUN	Starts the GHSL SMOD computation	CTRL + R
POPULATION	Select and load the Population grid	
BUILT-UP	Select and load the Built-up area grid	
REGIONS	Select and load the Regions grid or vector file	
LAND / WATER	Select and load the Land or Water grid	

2.2.3 Editable text fields

The land and built-up area panels contain two editable text field that allow the user to setup the expected minimum and maximum value of the input data. These values are used to rescale the input data in the range 0 - 1 in order to have normalized data before starting the GHSL SMOD computation.

The expected min and max values are precomputed internally, but can be adjusted by the user in case the heuristic guessed an incorrect value.

Box 1 Example of rescaling

The user built-up area raster contains values ranging, for example, from 0 to 80 (see Figure 20). GHS-POP2G heuristic choose 0 and 100 as minimum and maximum possible values assuming the layer represents built-up density as saturation percentage of a pixel and ranges between 0 and 100. If the real maximum value is 80 instead, e.g. 80 means 100% built-up, the user must change it and then rescaling will be performed from 0 to 80 instead of 0 to 100.

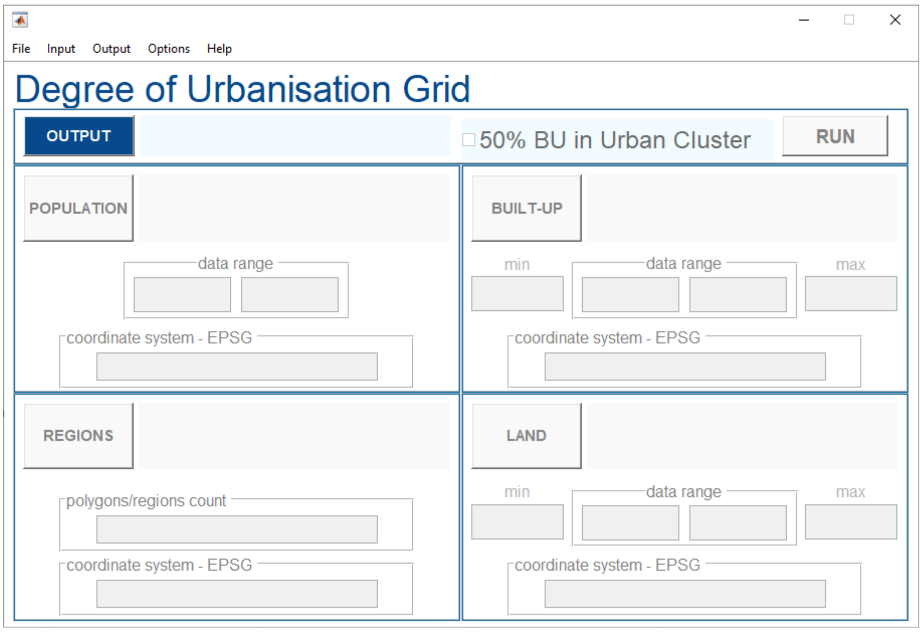
Table 10 Editable text fields

Item	Action
Built-up min	Set all built-up area pixels that are less than the specified value to this value before scaling
Built-up max	Set all built-up area pixels that are more than the specified value to this value before scaling
Land min	Set all land pixels that are less than the specified value to this value before scaling
Land max	Set all land pixels that are more than the specified value to this value before scaling
Coordinate system - EPSG	Display the dataset coordinate system in EPSG format. The user can overwrite this value if it’s incorrect or not automatically detected

2.2.4 Guided run

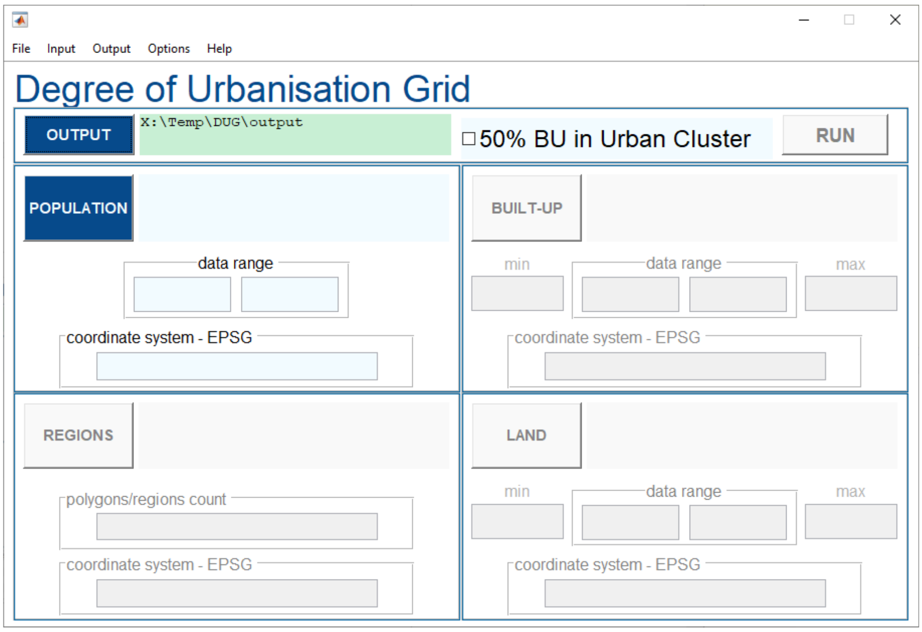
After starting GHS-DUG, the only active command is the output button (Figure 5): this is to set the output folder (path) first in order to create the log file and check disk write permission for results. Below in the shaded italic text lines the actions and commands the user should prompt to run the GHS-DUG Tool.

Figure 5 GHS-DUG startup



The user clicks on OUTPUT and selects the output folder (Figure 6)

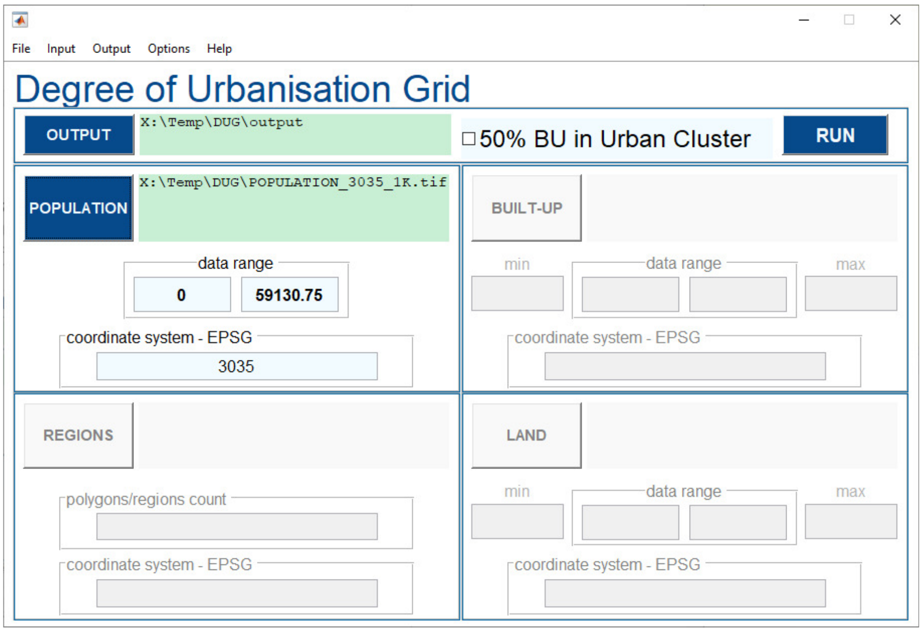
Figure 6 Output folder selected



Once output is set, the POPULATION becomes active.

The user chooses the population raster (Figure 7)

Figure 7 Population selected

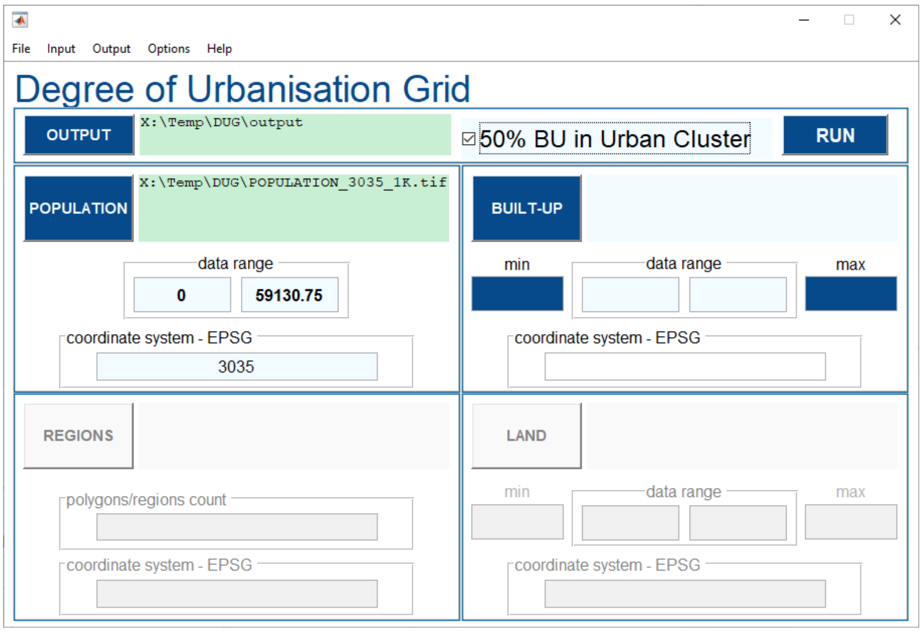


Once population is set, minimum and maximum raster value are displayed in data range and the coordinate system box reports the EPSG value. Now the user can click the RUN button or choose to apply the 50% BU rule in Urban Cluster or activate other options.

Selecting the 50% BU in Urban Cluster checkbox will activate the built-up area panels (Figure 8)

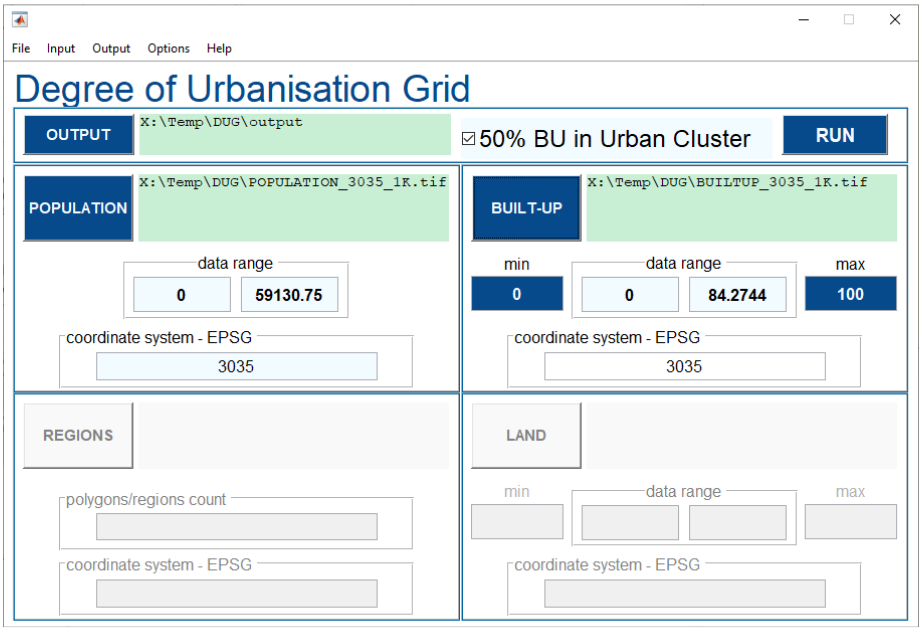
The user can load built-up area raster as already done for the population.

Figure 8 Checkbox 50% BU in Urban Cluster selected



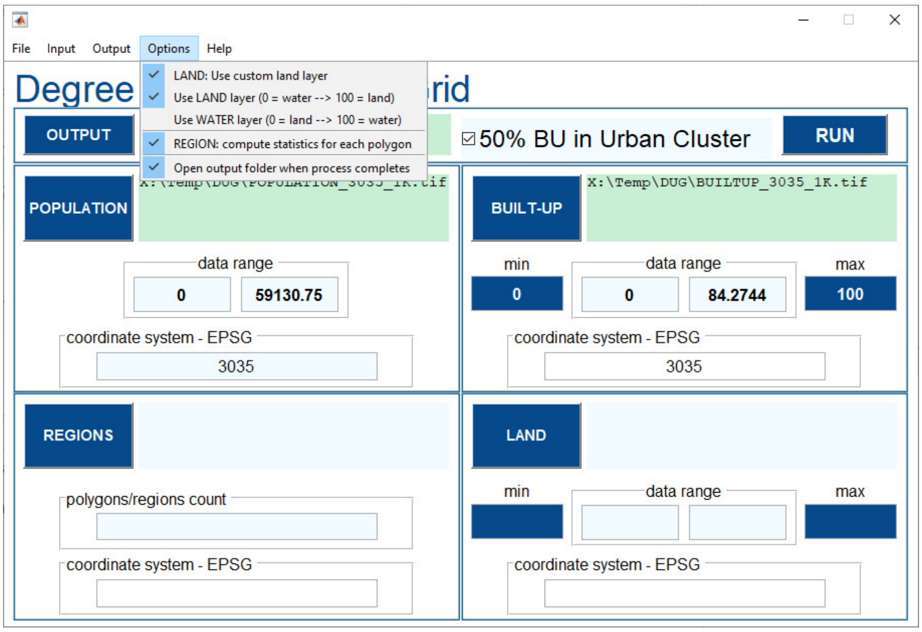
Once built-up area and/or land grids are loaded, minimum and maximum raster values are displayed in the data range. Internal heuristics set also the expected minimum and maximum value used to rescale the data, these values can be changed by the user (blue rectangles can be edited, limits are set as 0 – 100 for the example in Figure 9).

Figure 9 Built-up raster loaded and min max values set



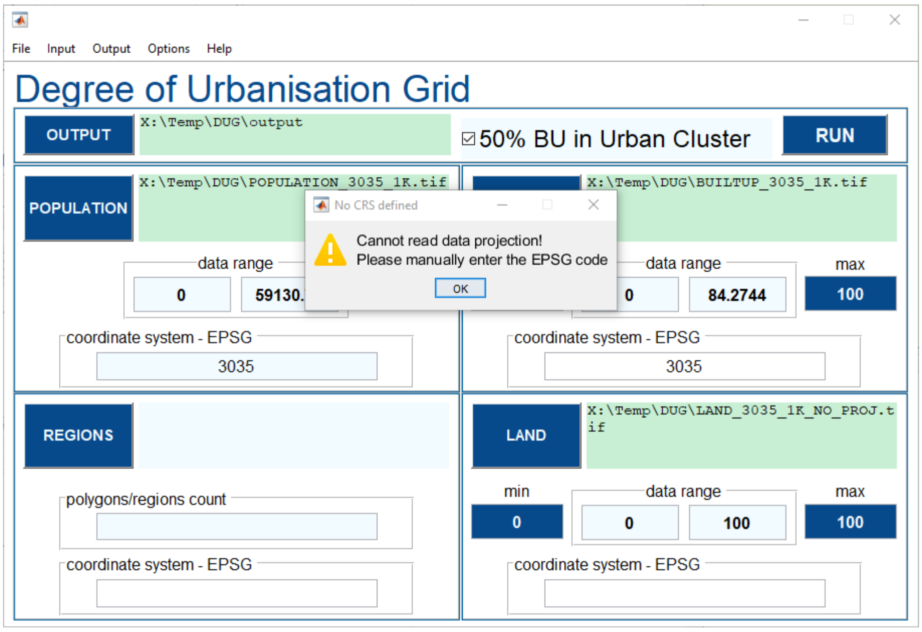
The other panels can be activated using the Options menu and allow the user to use a custom land layer and to provide a regions file in form of raster or shapefile; this file will be used to compute statistics in every distinct area designed by a polygon.

Figure 10 Regions and Land panel activate via Options menu



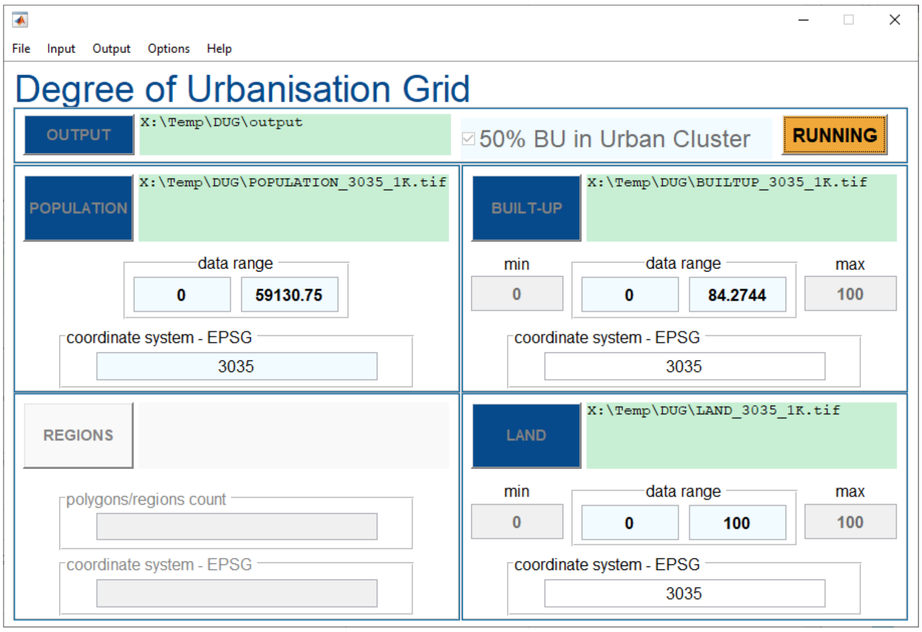
In case the dataset projection is missing or cannot be read correctly a warning is displayed. The user can then manually enter the EPSG code value in the corresponding "coordinate system - EPSG" box.

Figure 11 Missing EPSG code



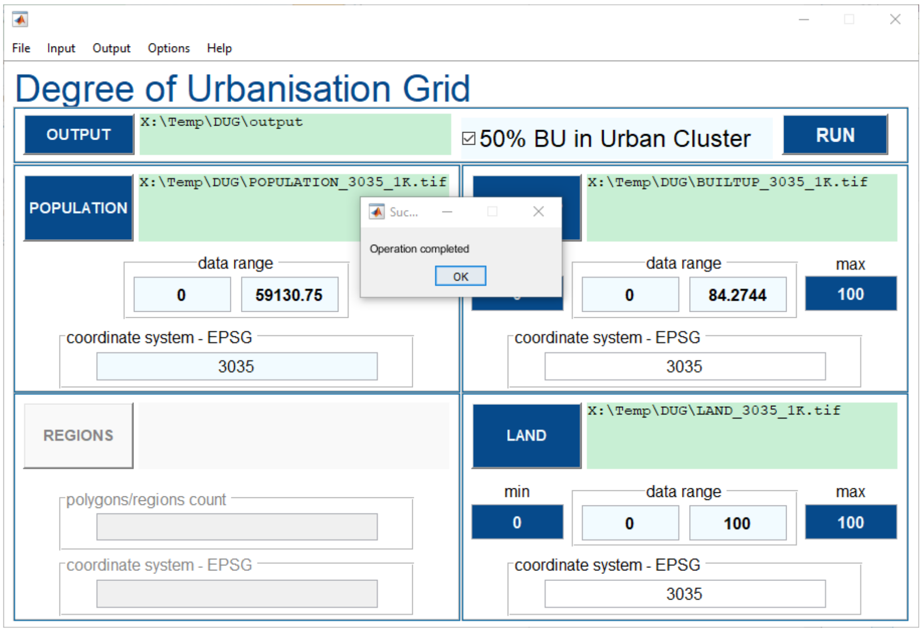
All commands are disabled when running, the RUN button becomes orange with label RUNNING and the mouse pointer transform into wheel (Figure 12).

Figure 12 GHS-DUG running



Once the run completed everything is reactivated again, a message is displayed and the output folder is opened in windows explorer, if the corresponding option is active.

Figure 13 Run completed



3 ArcGIS toolbox

The ArcGIS toolbox version of the tool allows the user to perform the classification directly from the ArcGIS software (Esri). Under the hood the main algorithm is still run through the same MATLAB code used for the standalone version to ensure equivalent results from both tools. The ArcGIS toolbox version has been developed in ArcGIS 10.6 environment and tested for back compatibility in ArcGIS 10.3.

3.1 Installation

3.1.1 System requirements

The requirements are the same as listed in 2.1.1.

3.1.2 Installation procedure

3.1.2.1 Install MATLAB Runtime

GHS-DUG is developed in MATLAB and the MATLAB Runtime is required to run the tool.

The specific version to be installed is **R2018b (9.5)**, Windows 64-bit version.

Check for the official MATLAB website to download and install:

https://it.mathworks.com/products/compiler/matlab-runtime.html

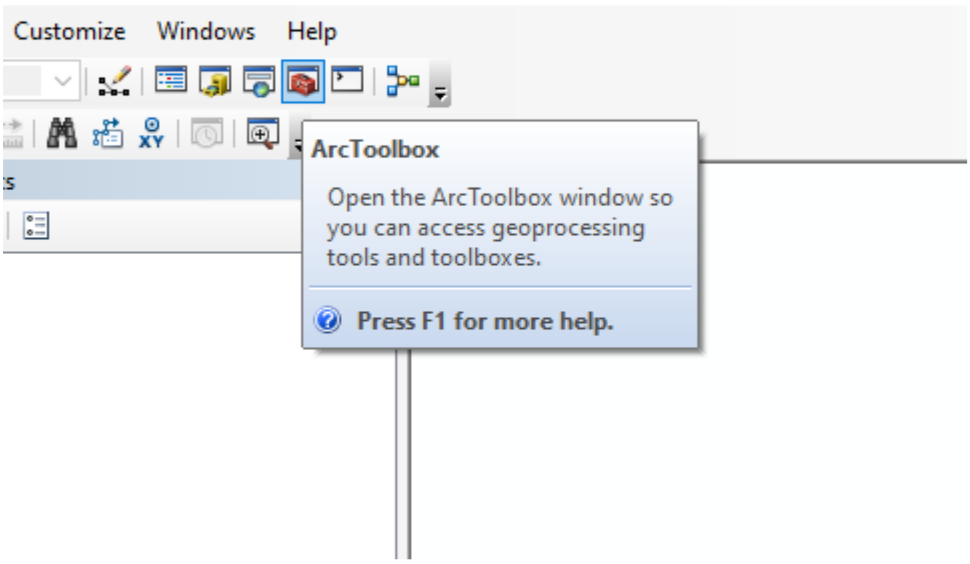
3.1.2.2 Install GHS-DU-TUC ArcGIS toolbox

Once the MATLAB Runtime is installed, if you have not done so already, download the GHS ArcGIS toolbox from the GHSL website (Tools section):

https://ghsl.jrc.ec.europa.eu/tools.php

Unzip the file, then open ArcGIS and from the Tools panel open the ArcToolbox window:

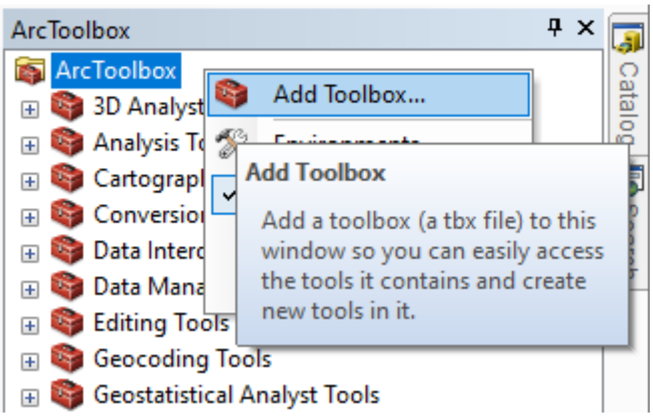
Figure 14 Show ArcToolbox window



From the ArcToolbox window right click, press Add Toolbox and select the *GHS-DUG.pyt* file.

The toolbox will be listed under GHS Tools section.

Figure 15 Add toolbox to ArcGIS



3.1.2.3 Remove GHS-DU-TUC ArcGIS toolbox

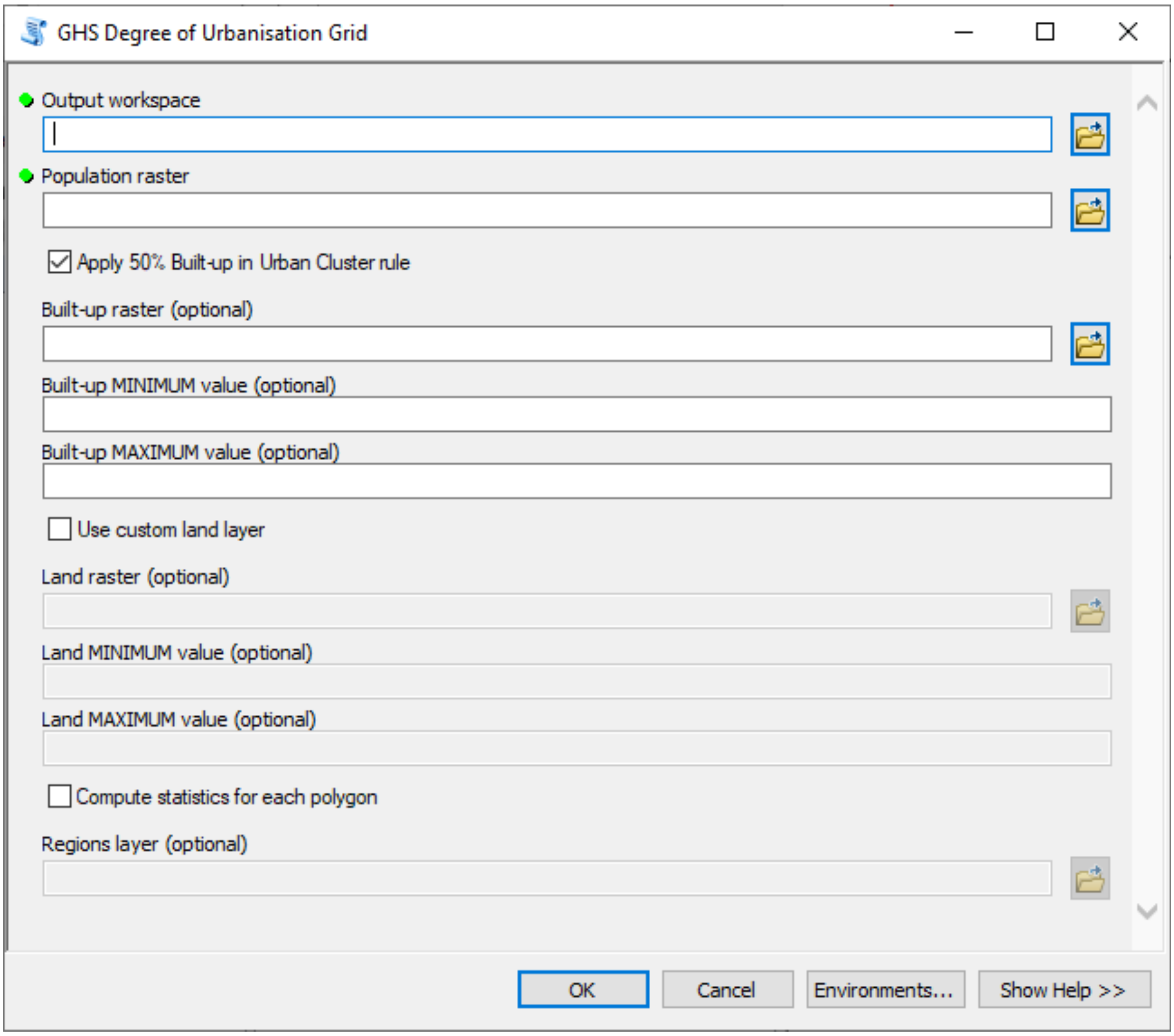
To uninstall GHS ArcGIS toolbox simply remove the toolbox from ArcGIS and delete the entire folder containing the toolbox files.

MATLAB Runtime can also be removed if not necessary for other applications.

3.2 The toolbox user interface

The toolbox user interface includes the same input and parameters than the standalone version.

Figure 16 ArcGIS toolbox window



The main windows cannot be split in sections, therefore all fields are together but disposed in a logical order. The user should start to fill them from top to bottom, but that's not mandatory.

Table 11 Toolbox items description

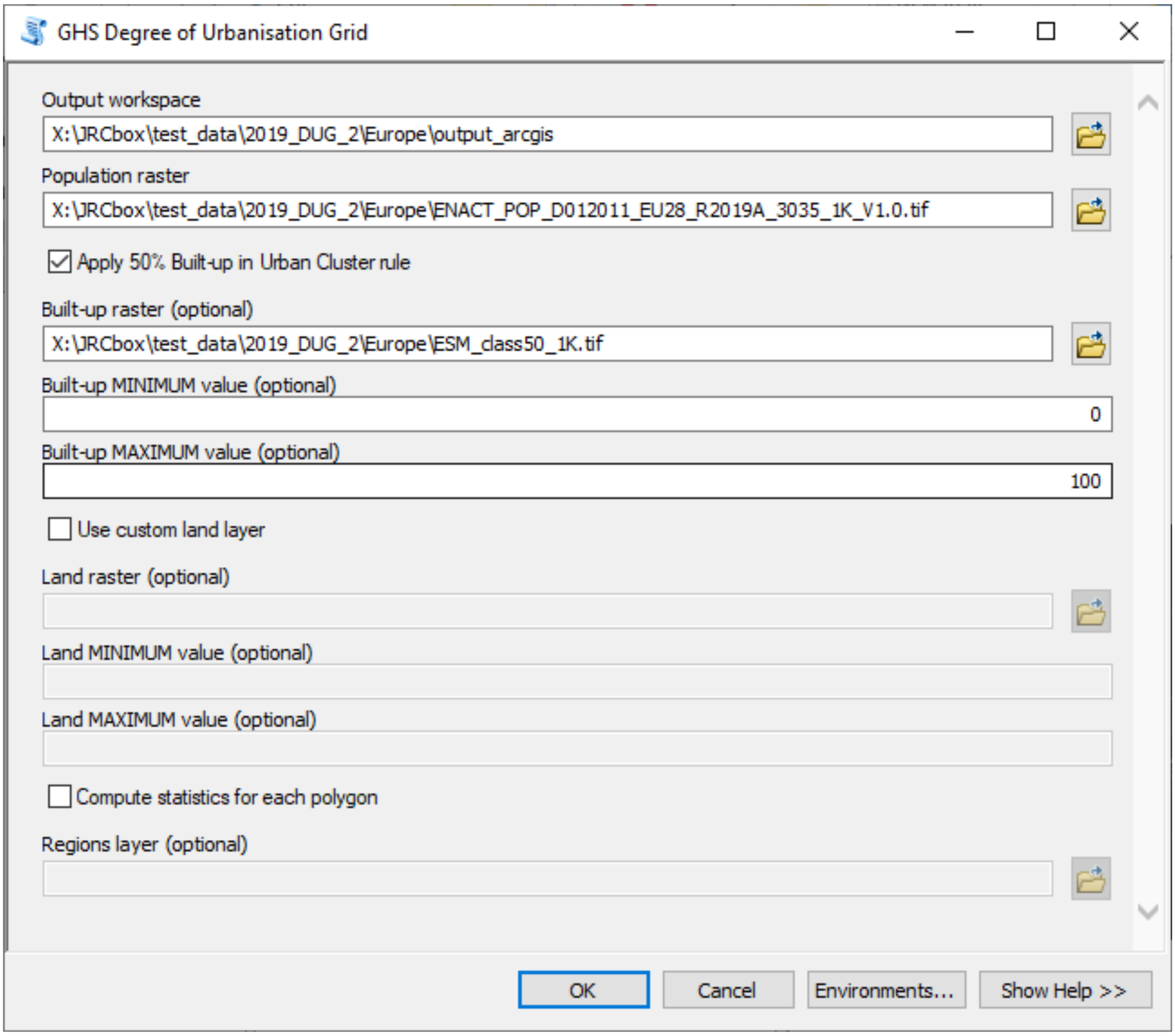
Item	**Description**
Output workspace	Set the output folder. This is the output path where results and intermediate files are saved. The user needs to have write permission to this folder
Population raster	Select and load the Population raster
Apply 50% Built-up in Urban Cluster rule	When selected it applies the 50% Built-up in Urban Cluster rule. It needs a Built-up raster provided by the user
Built-up raster	Select and load the Built-up raster
Built-up MINIMUM value	Set all built-up area pixels that are less than the specified value to this value before scaling
Built-up MAXIMUM value	Set all built-up area pixels that are more than the specified value to this value before scaling
Use custom land layer	When selected the user can provide a custom Land layer. By default is deactivated and an internal value is used.
Land raster	Select and load the Land raster
Land MINIMUM value	Set all land pixels that are less than the specified value to this value before scaling
Land MAXIMUM value	Set all land pixels that are more than the specified value to this value before scaling
Compute statistics for each polygon	When selected it computes the GHSL SMOD statistics for each polygon in the provided Regions shapefile
Regions layer	Select and load the Regions shapefile

For a detailed description of input data and their constraints check 4.1.

3.2.1 Run the toolbox

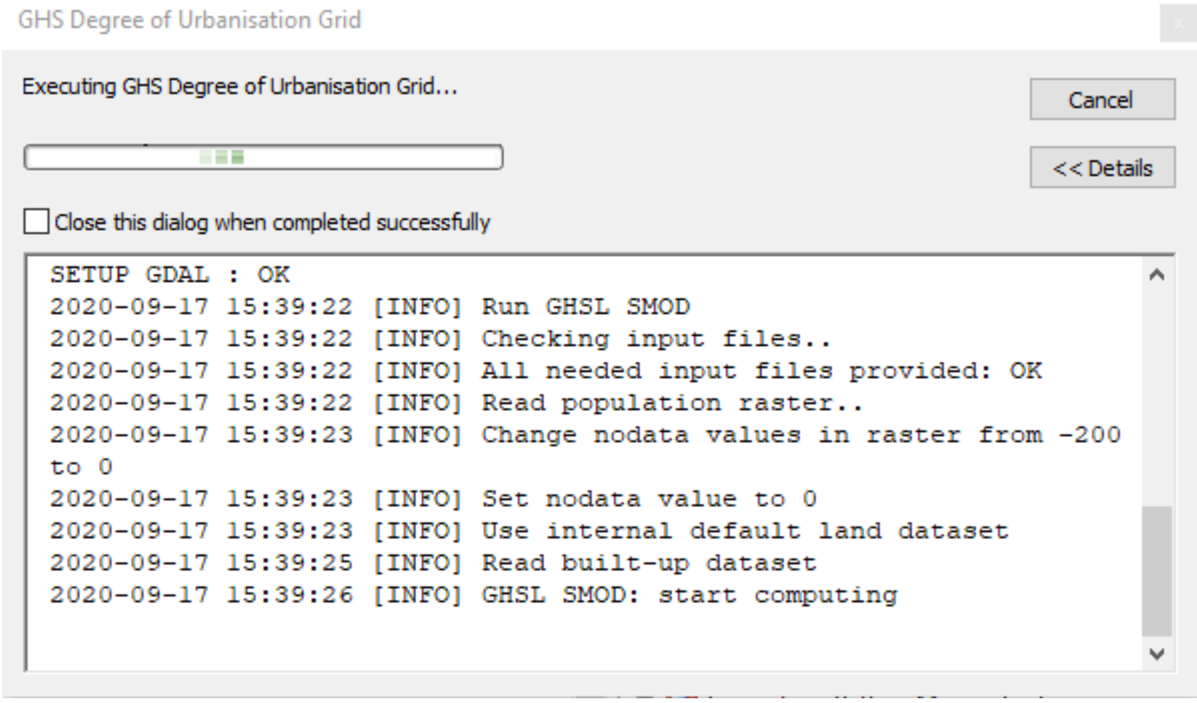
Before running the toolbox all inputs and parameters must be filled. The toolbox performs internal checks to ensure the values are correct. In case of doubt, the help section contains a short description about each field.

Figure 17 Toolbox setup with all inputs and parameters



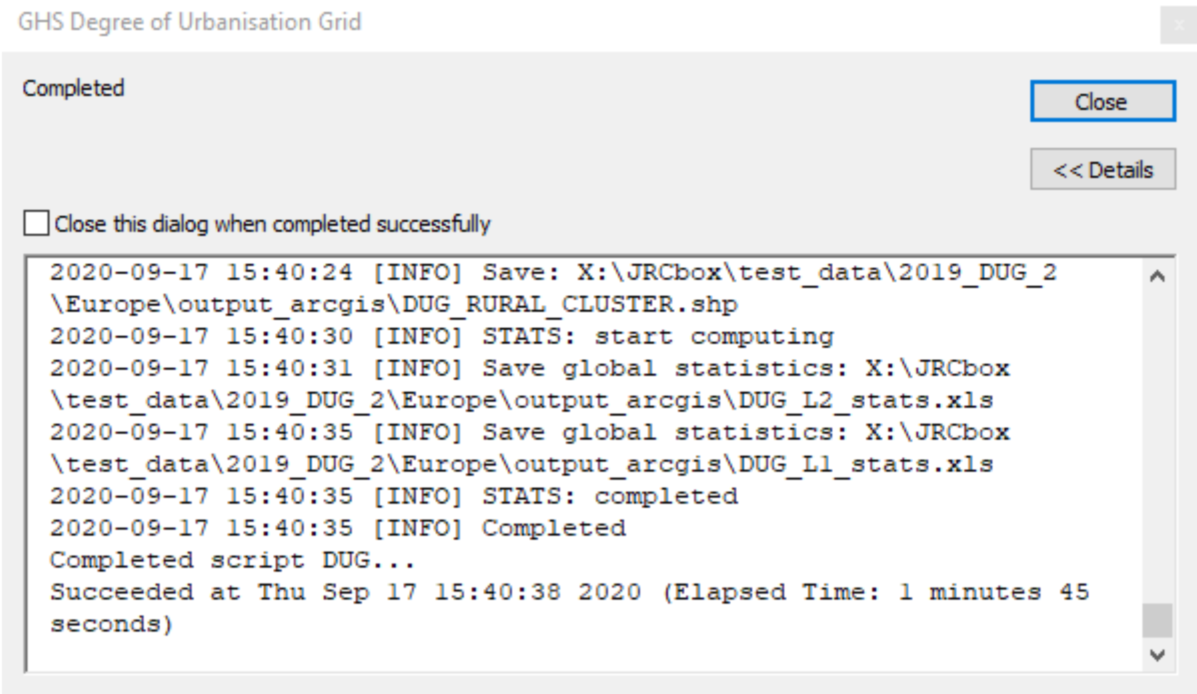
If no error messages appears while select the input data and the parameters the user can hit OK and start the computation. The user can follow the run progress by checking the computation logs.

Figure 18 Toolbox running



Once the process is completed the windows can be closed and the results can be loaded in ArcGIS or any other GIS to be inspected.

Figure 19 Toolbox run successfully completed



4 Input and output

4.1 Input

The input files requested by the tool depend on the selected GHSL SMOD parameter setting (Table 12).

Table 12 Required input (marked by green squares) and optional input (hollow) by setting

Input	without "50% BU in Urban Cluster"	with "50% BU in Urban Cluster"
Population grid	■	■
Land / Water grid	□	□
Built-up area grid		■
Regions	□	□

The Land / Water grid file is optional as the GHS-DUG tool comes with a default Land layer. However, it is highly recommended to provide a specific land layer align with the Population grid layer (see section 4.1.4)

The population has to meet key requirements, see all details in the Population grid sub-paragraph (4.1.1).

The land, built-up area and regions grids can instead be in any projection supported by GDAL (box 2 for details) and can have any pixel size. If those values do not match with the values from the population grid, those grids will be resampled and re-projected in order to align them to the population raster. A sign of the transformations is compiled in the log file.

Box 2 Supported projections

Check the projection supported by GHS-DUG through GDAL at:

https://proj.org/operations/projections/index.html

4.1.1 Population grid

The population grid is the mandatory input to compute GHSL. The population grid has to be a GeoTIFF file, with integer or float data type, in one World Mollweide or LAEA Europe projection, and with a 1 km x 1 km pixel size (Table 13). Each pixel shall contain the population count expressed as absolute number (e.g. "3125.9" means 3125.9 people). If the input population grid does not comply with these requirements the process cannot start and the tool throws an error. In such case the user needs to re-project and/or resample the grid in order to comply with these requirements or provide another grid. The re-projection of geospatial layers requires specific technical knowledge. No responsibility is taken for workflows performed by users.

The best way to be sure that the grid comply with all requirements is to generate it using the **GHS-POP2G tool** (Maffenini et al., 2020).

Be careful: re-projecting population grid is a sensitive task, as the resampling method can affect population distribution and total count.

Table 13 Population grid requirements

File format	Data type	Map projection	Pixel size
GeoTIFF	Any numeric (int or float)	— World Mollweide (EPSG:54009) — ETRS89 / LAEA Europe (EPSG:3035)	1 km x 1 km

4.1.2 Built-up area grid

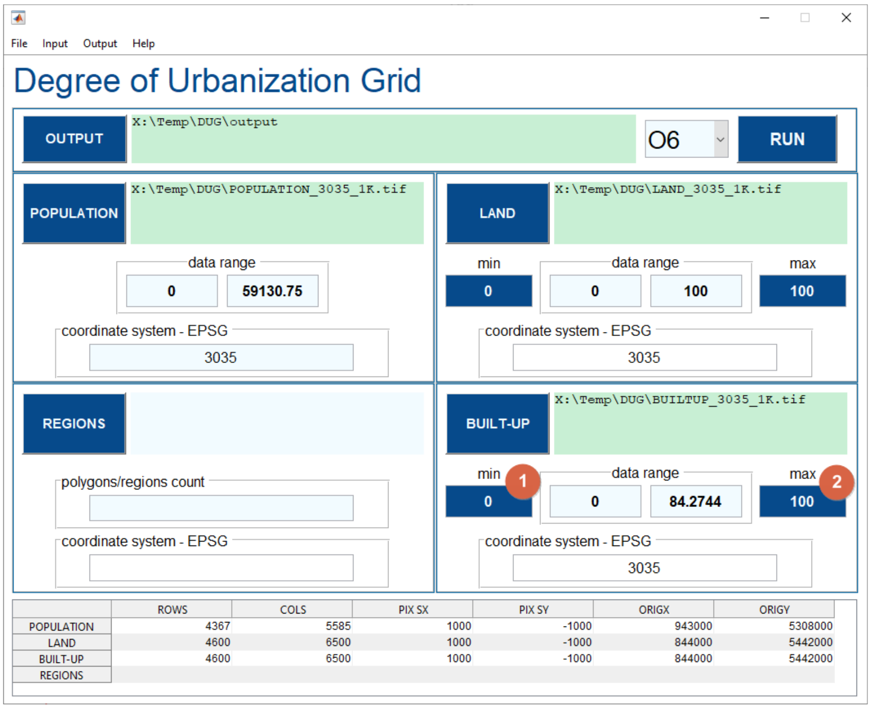
The built-up area grid is required when using 50% Built-up in Urban Cluster rule. The built-up area grid shall contain the pixel share of built-up area, usually expressed as a value in the range 0 - 1: "0" means 0% built-up area, "1" means 100% built-up area. The user can provide a grid with a different data range: in this case the tool will internally rescale those value to the standard range 0 - 1.

Table 14 Built-up area grid requirements

File format	Data type	Map projection	Pixel size
GeoTIFF	Any numeric (int or float)	Any supported by GDAL, it will be reprojected to population's projection	Any, it will be resized at 1 km x 1 km

Once the built-up area grid is loaded, its actual value range is displayed and the expected minimum and maximum values are guessed by the internal heuristics. The user can change these values to rescale the land grid using a different minimum and maximum values (operating on buttons highlighted with marker 1 and 2 in Figure 20).

Figure 20. Built-up area expected min and max



4.1.3 Regions

The regions file is optional and is used to delineate the zones in which the tool computes the GHSL SMOD values and their statistics. These areas could be of any sort bigger than 1 grid cell (administrative areas, countries, major regions of the world). The regions file can be in the form of raster grid or vector data (all the characteristics of the regions file are presented in the following subsections).

4.1.3.1 Raster grid

The regions raster grid shall contain integer values: 0 is considered as the background or no region. All positive values are considered as distinct areas.

Table 15 Regions raster grid requirements

File format	Data type	Map projection	Pixel size
GeoTIFF	Any numeric (int or float)	Any supported by GDAL, it will be re-project to population's projection	Any, it will be resized at 1 km x 1 km

4.1.3.2 Polygons vector data

The regions file can be also provided in the shapefile format. In this case it will be internally rasterized using the population grid as target grid. Each polygon will get a unique identifier, ranging from 1 to the maximum number of polygons. GHS-DUG will also read the polygons attribute table and create a new sheet that combine these values with the GHSL SMOD statistics computed for each rasterised polygon.

Table 16 Regions vector file requirements

File format	Data type	Map projection	Geometry
Shapefile	Any numeric (int or float)	Any supported by GDAL, it will be re-project to population's projection	Polygons

4.1.4 Land / Water grid

The land grid is optional, when not provided a default standard layer will be used internally. The layer is expected to contain the pixel share of permanent land surface, usually expressed as a value in the range 0 - 1: "0" means 0% land, "1" means 100% land.

Instead of the land grid a water grid can be used. This is the logical inverse of land: it contains the pixel share of permanent water surface, usually expressed as a value in the range 0 - 1: "0" means 0% water, "1" means 100% water. This option can be chosen using the land menu.

The user can provide a grid with a different data range: in this case the tool will internally rescale those value to the standard range 0 - 1.

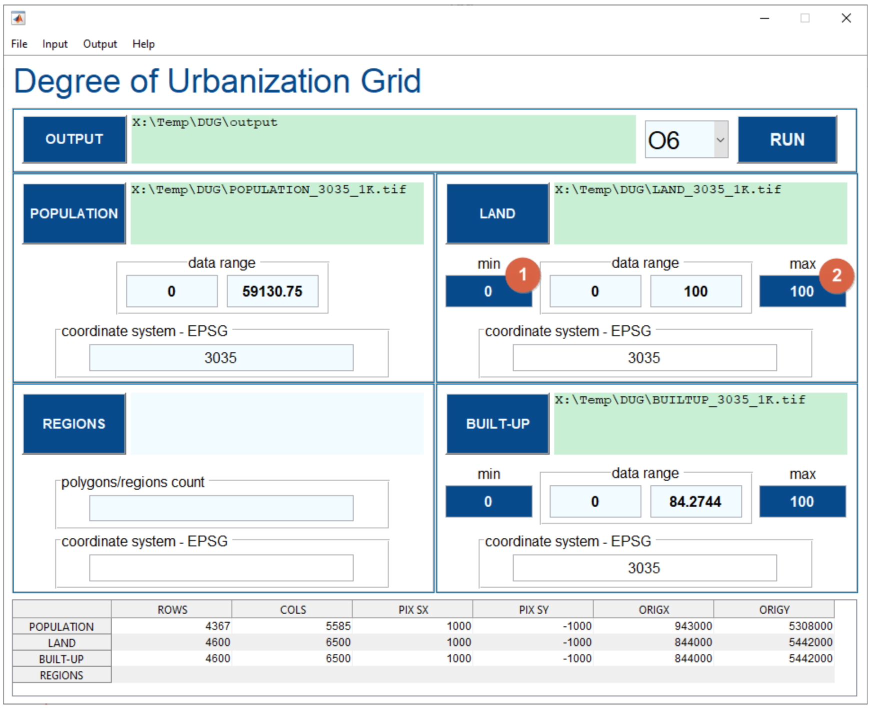
Table 17 Land / Water grid requirements

File format	Data type	Map projection	Pixel size
GeoTIFF	Any numeric (int or float)	Any supported by GDAL, it will be reprojected to population's coordinate system	Any, it will be resized at 1 km x 1 km

Once the land grid is loaded, its actual value range is displayed and the expected minimum and maximum values are guessed by the internal heuristics. The user can change these values to rescale the land grid using a different minimum and maximum values (operating on buttons highlighted with marker 1 and 2 in Figure 21).

The default land layer has been generated with global purposes without any country borders specification. Therefore, to avoid small biases in the classification grid for delimited areas of interest (e.g. a single country), it is highly recommended to provide a specific raster land layer for the population used. This raster land layer should be aligned with Population grid and should specifically represent the share of land available for the people depicted in each pixel of the Population grid. For example, while the default layer is quite accurate in representing the land shares on coastal areas (both inland and sea waters), it does not represent share of land for a specific country along the border with other countries (i.e. cross-border pixels with population only from one country will have distorted densities as they won't use the specific share of land in the given country).

Figure 21 Land grid expected min and max



4.2 Output

The GHS-DUG 4 tool generates two types of output. First, the geospatial outputs, a combo of raster and shapefile layers that classify grid cells into settlement classes (as described in section 1.1 p.6) or delineate settlement entity. Second, the statistics including population count and areal extent per settlement class.

4.2.1 GHS-DUG geospatial output

The GHSL SMOD geospatial output data consists in several files. The following outputs are produced both as raster grid (GeoTIFF) and shapefile: Urban Centre entities, Dense Urban Cluster entities, Semi Dense Urban Cluster entities, and Rural Cluster entities; the settlement grid at second hierarchical level is delivered as raster grid (Table 18).

Table 18 GHS-DUG output files (outputs available marked in green)

Output	**Raster**	**Shapefile**
GHS-DUG_GRID_L1	■	
GHS-DUG_GRID_2	■	
GHS-DUG_URBAN_CENTRE	■	■
GHS-DUG_DENSE_URBAN_CLUSTER	■	■
GHS-DUG_SEMI_DENSE_URBAN_CLUSTER	■	■
GHS-DUG_RURAL_CLUSTER	■	■

4.2.1.1 GHS-SMOD GRID L1

The GHSL SMOD raster file contains the GHS-DUG grid L1 classification, where each pixel is classified in one of these settlement classes: 1, 2 and 3 (see paragraph 1.1).

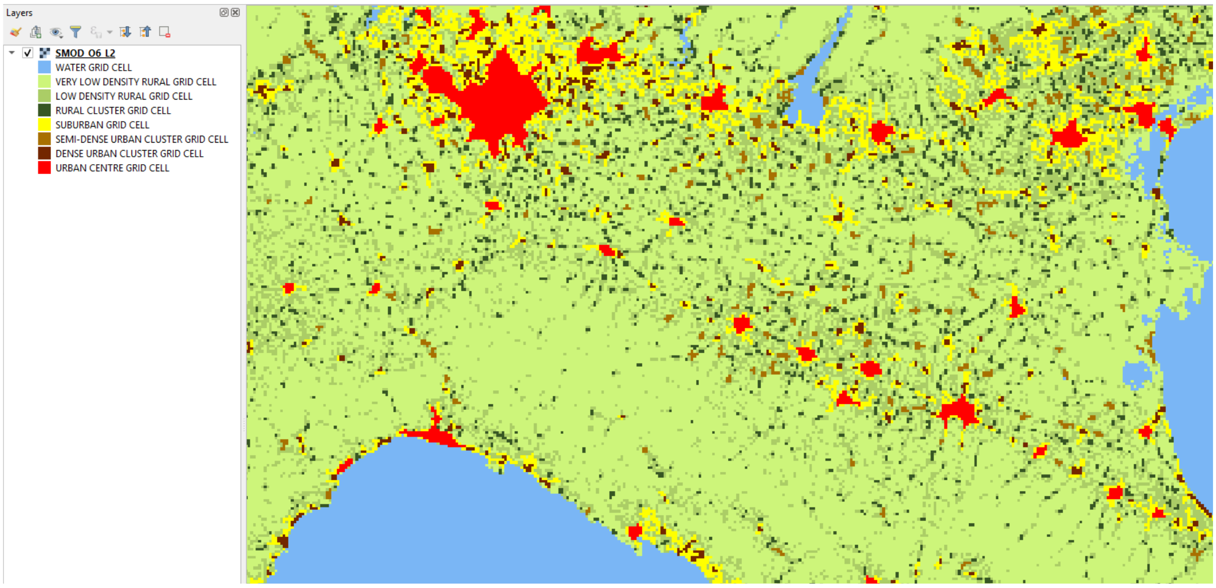
The filename of the output is *GHS-DUG_GRID_L1.tif*.

4.2.1.2 GHS-SMOD GRID L2

The GHSL SMOD raster file contains the GHS-DUG grid L2 classification, where each pixel is classified in one of these settlement classes: 10, 11, 12, 13, 21, 22, 23 and 30 (see paragraph 1.1).

The filename of the output is *GHS-DUG_GRID_L2.tif*.

Figure 22 Screen capture of a QGIS session displaying the GHS-DUG Grid L2 output of the GHS-DUG tool. Example in northern Italy



4.2.1.3 GHS-DUG spatial entities

The GHS-DUG vector files contain spatial entities that are derived from GHSL SMOD grid L2 classification raster. The results are stored in shapefiles, where each entity has a unique identifier ranging from 1 to the maximum number of entities. The attribute table of the shapefile contains basic statistics (Table 19). See paragraph 4.2.2.3 for more information about the statistics stored into the entity shapefiles.

Table 19 GHSL SMOD spatial entities

GHS-SMOD spatial entity	Shapefile filename	Shapefile statistics
URBAN CENTRE	GHS-DUG_URBAN_CENTRE.shp	Area, population, built-up area (1)
DENSE URBAN CLUSTER	GHS-DUG_DENSE_URBAN_CLUSTER.shp	Area, population, built-up area (1)
SEMI DENSE URBAN CLUSTER	GHS-DUG_SEMI_DENSE_URBAN_CLUSTER.shp	Area, population, built-up area (1)
RURAL CLUSTER	GHS-DUG_ RURAL_CLUSTER.shp	Area, population, built-up area (1)

(1) Only when built-up area grid is provided by the user

4.2.2 GHS-DUG output statistics

4.2.2.1 Summary GHS-DUG statistics

GHS-DUG computes basic statistics about the generated GHS-DUG raster, representing population counts per each GHSL SMOD class (Population_count in Figure 23), number of grid cells per class (Area_km2), and the share of the total population accounted in each GHSL SMOD class (Population_share[5]). The input for the statistics is the raster grid *GHS-DUG Grid L2* output. The same statistics are also aggregated at the first hierarchical level without generating any additional raster (Figure 24)

[5] The population share represents the ratio between the selected class and the total sum of population in each class. Its value is in the range 0 - 1.

Figure 23 Screen capture of a MS Office Excel session displaying an example of GHS-DUG L2 summary statistics

	A	B	C	D
1	L2_class	Population_count	Area_km2	Population_share
2	10	0	37948	0
3	11	898054.9227	407999	0.033099738
4	12	4707707.131	34161	0.173512629
5	13	3531490.086	4626	0.130160631
6	21	1956236.935	2438	0.072101302
7	22	807628.1617	778	0.029766866
8	23	3664146.405	978	0.135049964
9	30	11566520.1	1709	0.426308871

Figure 24 Screen capture of a MS Office Excel session displaying an example of GHS-DUG L1 summary statistics

	A	B	C	D
1	L1_class	Population_count	Area_km2	Population_share
2	1	9137252.14	484734	0.336772998
3	2	6428011.502	4194	0.236918131
4	3	11566520.1	1709	0.426308871

The filenames are GHS-DUG_L2_stats.xls and GHS-DUG_L1_stats.xls (with X from 1 to 7).

4.2.2.2 Statistics by region

When a region file is provided, then the population count per GHSL SMOD class is also disaggregated per region (Figure 25). Depending on the extent of the regions the total sum per class can be different than the one computed on the entire raster: if the regions area covers a smaller extent compared to the population grid input, statistics will be computed within the sole regions extent.

In the case the region file is a vector file, these statistics are also combined with the original shapefile attributes table. The table structure is straightforward, the population counts per class are organised in columns with the header containing the corresponding GHSL SMOD class code. Statistics for some polygons may not be calculated due to the nature of the rasterization process used to compute them. Such polygons will be listed in “Sheet2” of the output spreadsheet.

Figure 25 Screen capture of a MS Office Excel session with an example of GHS-SMOD statistics by region

	A	B	C	D	E	F	G	H
1	Region_label	L2_class_10_pop_count	L2_class_11_pop_count	L2_class_12_pop_count	L2_class_13_pop_count	L2_class_21_pop_count	L2_class_22_pop_count	L2_class_23_pop_
2	3	0	84865.57207	424810.667	298744.4918	180203.8431	89337.49265	4026
3	14	0	94036.58292	545256.1956	728789.8971	309583.8826	230469.7552	34852
4	24	0	690226.0424	3500291.402	2325814.735	1395197.853	431035.0007	26566
5	27	0	28108.40367	231611.6033	177030.4639	71251.35648	56785.91312	25635

The filename is GHS-DUG_L2_L1_stats_regions.xls.

4.2.2.3 Statistics by spatial entity

GHS-DUG computes basic statistics for each entity (available entities are listed in Table 18). The statics computed are:

1. Entity enclosed area in km^2;
2. Entity enclosed population count;
3. Entity enclosed built-up area in km^2 (only when built-up area grid is provided by the user).

These values are saved in the attributes table of each entity shapefile.

Figure 26 Screen capture of a QGIS session with an example of spatial entity statistics

	ID_UC	Area_km2	Pop_count	Builtup_km2
1	7	18	58776.895050	7.395387
2	2	13	63921.917725	5.159350
3	1	13	54107.244202	2.862250
4	4	35	144246.621674	8.977819
5	3	29	147927.440063	10.950325
6	14	22	142461.287720	9.845706
7	13	12	78188.381165	3.496750
8	16	12	88048.113525	4.831875
9	15	14	81811.688126	2.164825
10	10	64	428807.582523	24.946881
11	9	42	184213.866428	12.078563
12	12	42	232087.884926	11.897681

5 Conclusion

The GHS-DUG Tool 4 adds to the family of GHSL open and free tools to enact the "open input, open method, open output" paradigm of the Global Human Settlement Layer framework. With the GHS-DUG Tool presented in this report the user can produce settlement classification grids and related statistics for specific areas of interest and input data. This capability supports the endeavours of the European Commission and the partner Organisations of the voluntary commitment to develop a harmonised global definition of cities and settlements launched at the Habitat III conference in 2016. The tool specifically supports the capacity enhancement activities currently carried out jointly by the European Commission Directorate General for Regional and Urban Policy, the Directorate General Joint Research Centre and the United Nations Human Settlements Programme to apply the Degree of Urbanisation in support of global monitoring of SDGs and the New Urban Agenda urban targets as recommended by the 51st Session of the United Nations Statistical Commission.

References

Corbane, C., Pesaresi, M., Politis, P., Florczyk, A.J., Melchiorri, M., Freire, S., Schiavina, M., Ehrlich, D., Naumann, G., Kemper, T., 2018. The grey-green divide: multi-temporal analysis of greenness across 10,000 urban centres derived from the Global Human Settlement Layer (GHSL). International Journal of Digital Earth 1–18. https://doi.org/10.1080/17538947.2018.1530311

Dijkstra, L., Poelman, H., 2014. A harmonised definition of cities and rural areas: the new degree of urbanisation (Working Papers), Regional Working Paper 2014.

European Commission , Joint Research Centre, 2018. Atlas of the Human Planet 2018, a World of Cities. Publications Office of the European Union.

Florczyk, A.J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., Zanchetta, L., 2019. GHSL Data Package 2019. Publications Office of the European Union.

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

Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M., Kemper, T., 2018. Unveiling 25 Years of Planetary Urbanization with Remote Sensing: Perspectives from the Global Human Settlement Layer. Remote Sensing 10, 768. https://doi.org/10.3390/rs10050768

Melchiorri, M., Pesaresi, M., Florczyk, A.J., Corbane, C., Kemper, T., 2019. Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator—SDG 11.3.1. ISPRS International Journal of Geo-Information 8, 96. https://doi.org/10.3390/ijgi8020096

List of boxes

List of figures

List of tables

Annexes

Annex 1. Error messages

Error	**Description**	**Solution**
Cannot create log file! Check you have write permission	The tool has no write access to the selected folder and cannot create log file	Check and change folder permission or chose another folder with write permission
The selected population raster has no spatial metadata	The population raster is not a valid GeoTIFF, it does not contain georeferencing information in its tags	Use a population raster with valid georeferencing information embedded in tis tags
The selected population raster has no kilometric grid	The population raster pixel size is different than 1000 meters	Use a population raster with 1000 meters pixel resolution
The selected population raster has unsupported projection "EPSG: XXXX"	The population raster has a spatial reference that is not supported by the tool	Use a population raster that has either World Mollweide (EPSG:54009) or ETRS89 / LAEA Europe (EPSG:3035) spatial reference
Cannot read data projection! Please manually enter the EPSG code	The tool cannot detect the Spatial Reference of the provided layer	Manually enter the layer EPSG code in the dedicate box

KJ-04-20-537-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:10.2760/916217

ISBN 978-92-76-22416-7