European Commission

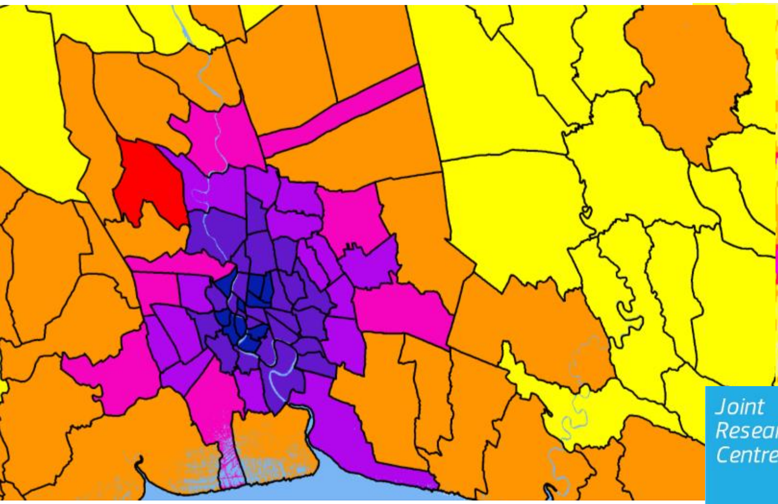
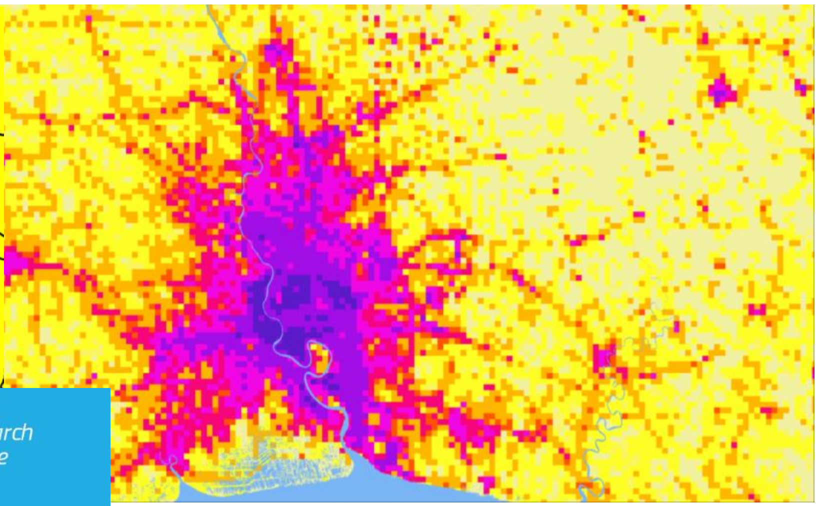
JRC SCIENTIFIC INFORMATION SYSTEMS AND DATABASES

POP2G User Guide

Population To Grid Tool
User Guide
Version 1.0

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

2020

Joint Research Centre

This publication is a Scientific Information Systems and Databases report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

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

JRC120164

PDF ISBN 978-92-76-17301-4 doi:10.2760/305540

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Maffenini, L., Schiavina, M., Freire, S., Melchiorri, M., Pesaresi, M. and Kemper, T., POP2G User Guide, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17301-4, doi:10.2760/305540, JRC120164

Contents

Authors

Maffenini, Luca[b]; Schiavina, Marcello[a]; Freire Sergio[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

Abstract

The Population To Grid tool (POP2G) –version 1.0- is an information system developed in the framework of the Global Human Settlement Layer (GHSL) to produce geospatial grids of population counts at different spatial resolutions.

The POP2G v1.0 is a flexible tool to produce geospatial population grids in GeoTIFF format from census data. The tool operationalises the workflow developed for the production of the Global Human Settlement Layer Population Grid layers (GHS-POP). The POP2G v1.0 tool allows the creation of population grids at 50 m, 250 m, and 1 km spatial resolutions, handling census data stored as point or polygon vector data (in latter case requires additional covariate as input for dasymetric disaggregation). The principal purpose of the tool is the production of the population grid used as input for the Degree of Urbanisation Grid (DUG) also produced in the GHSL framework. However the potential uses of the tool and population grids go far beyond this main application.

The tool is a capacity enhancement asset in the framework of the multi-stakeholder effort to develop a people-based harmonised definition of cities and settlements that helps the assessment of the feasibility of applying a global definition of cities/urban areas in support of global monitoring of SDGs and the New Urban Agenda urban targets. The POP2G, 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. So far, however, no method or international standard has been established at the global level to delineate these areas. 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 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.

Data on global population distribution are a strategic resource currently in high demand in an age of new Development Agendas that call for universal inclusiveness of people. Regular grids (raster or vector) are now well established and widely used spatial structures to model and report population attributes, due to their flexibility and advantages. Making population data available as fine grids facilitates analyses, increases flexibility and interoperability with other spatial datasets, and mitigates the modifiable areal unit problem (MAUP). Census campaigns and other ground surveys are usually the main source of population and other demographic data. In georeferenced censuses, counts of resident population are assigned to census reporting units or zones, or geocoded to dwellings.

The sub-section below includes elements of population grids, and the two principal methods that are used to produce them. The POP2G v1.0 is designed to produce population grids both through aggregation and disaggregation methods. The methodology primarily depends on the typology of input geospatial census data used, i.e. census counts attached to points or to polygons.

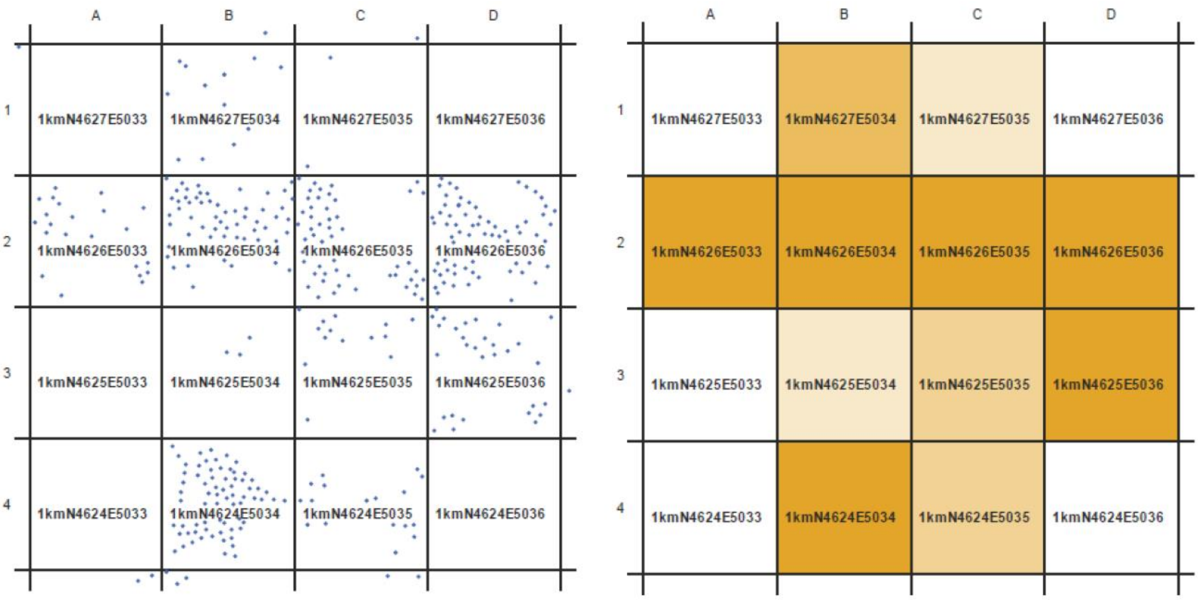
1.1 Population grids

A population distribution grid is a geospatial layer composed of regular grid cells (usually square) containing population counts per each cell. A population grid is created by re-allocating population counts from points and/or polygons to gridded surfaces of regular and standardised grid cells or pixels. This grid is produced through geospatial and geo-statistical processing of georeferenced population counts, such as that originating from a population census or register. The population grid can be produced in different ways depending on the type of data available.

1.1.1 Method through aggregation of point data

Population distribution grids can be produced by aggregation of detailed population counts. If census data is provided with the coordinates of respective dwellings (i.e. point data), no inference on the potential location of these residents is needed and all that is further required is to add up or aggregate the population of dwellings located within each grid cell. In particular for *micro-census data* that is generally restricted information and it is structured in a way so that have several records/features (points –building locations) fall into each 1km grid cell as displayed in Figure 1. Point-based micro-census is generally conducted at the building or block level, this high level of spatial detail should be the only one for which the aggregation method can be deployed.

Figure 1. Example of point-based data overlaid on a statistical geocoded grid of 1 km (left); and Population counts in shades of orange according to population density per 1 km cell (unpopulated grid cells in white) for aggregated point-based information (right). *Image from: Methodological manual on territorial typologies, p. 14 and p.15*

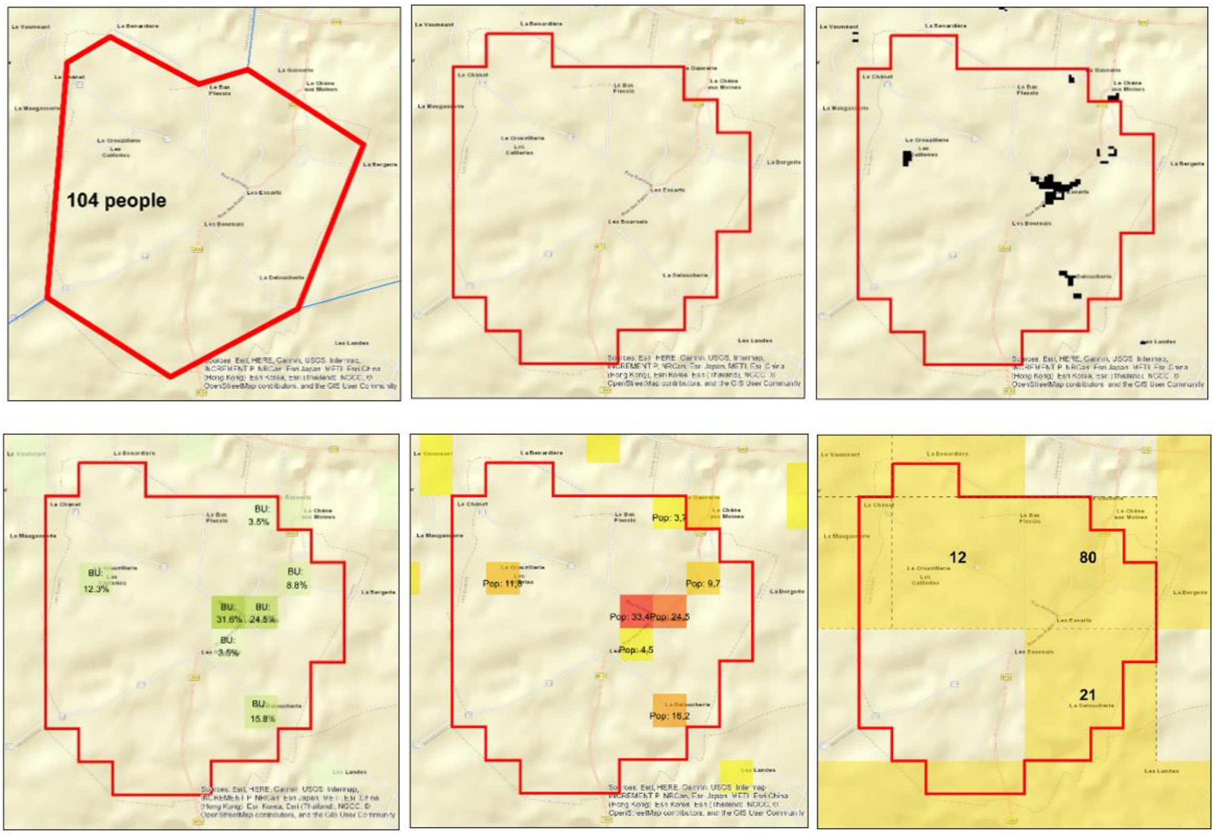


1.1.2 Method through disaggregation from census units

The second method to produce a population grid is that of disaggregation. If the exact location of each household is not available, the actual population distribution can be inferred or estimated by disaggregating counts available by census reporting units or administrative areas (e.g. municipalities, districts, provinces). In this case a population disaggregation grid can be created by combining the population of census or administrative units with spatial ancillary data mapping likely inhabited areas, such as urban land cover or built-up areas. Such ancillary data can be derived from remote sensing data, such as satellite imagery. Several organisations offer such free global layers, including the Global Human Settlement Layer produced by the European Commission's Joint Research Centre.

The GHS-POP R2019A (documented in Florczyk et al. 2019) is produced through disaggregation (population input is the Gridded Population of the World v4.10 produced by CIESIN). The disaggregation in GHSL environment is driven by the density of built-up areas as proxy for locations of residential population. Based on Freire et al. 2016, an explanatory workflow is schematised in Figure 2.

Figure 2. Example of the process used to generate the GHS-POP layer (extract from a location in France). In the first image the population of a census unit reports the population of the area (p); second the administrative unit is outlined with a 250 m grid; third the built-up areas (b%) are mapped at 30 m resolution; fourth the built-up areas are expressed as share of a 250 m grid cell; fifth the population is assigned to the 250 m grid cells where built-up areas are present proportionally to built-up area density with the equation POPcell = p X b% this step produces a 250 m population grid; sixth the 250 m population grid is aggregated in a 1 km grid (in yellow) and the resident population is expressed as density and count per 1 km^2. *Image from: Atlas of the Human Planet 2019, p. 21*



2 Installation

2.1 System requirements

The Population To Grid tool (POP2G) is graphical tool developed in MATLAB and created for Windows, its system requirements are listed below (Table 1)

Table 1. System requirements

Requirements	**Operating system**	**Processor**	**RAM**	**Disk space**
Minimum	Windows 7	Any Intel or AMD x86-64 processor	8 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)

POP2G operates on raster grids and shapefiles: in general the larger the extent, or the higher the output maximum grid resolution, the higher the required memory. Few examples of RAM requirements are below in Table 2. The input raster size (in number of 1 km grid cells) is intended per each of the raster layer(s). However, using low maximum pixel resolution as output (e.g. 1000) does not guarantee lower memory consumption. When input shapefile contains very high resolution polygons (e.g. smaller than 1 km^2) the memory consumption could increase due to the complex rasterization procedure required (see section 3.3 Box 2, for examples on how to reduce memory consumption).

Table 2. Requirements by input raster size

Population layer extent	Population number of features	Built-up raster size	Output pixel resolution	Process RAM needed	Disk space for outputs
1500x1500km	~ 14000 polygons	1700x1700 pixels	1000	1 GB	< 10 MB
1500x1500km	~ 14000 polygons	6800x6800 pixels	250	2 GB	< 10 MB
1500x1500km	~ 14000 polygons	34000x34000 pixels	50	> 32 GB	< 100 MB
1500x1500km	~ 14000 polygons	34000x34000 pixels	50 (stepwise)	28 GB	< 100 MB

2.2 Installation procedure

2.2.1 Install MATLAB Runtime

POP2G 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

2.2.2 Install POP2G

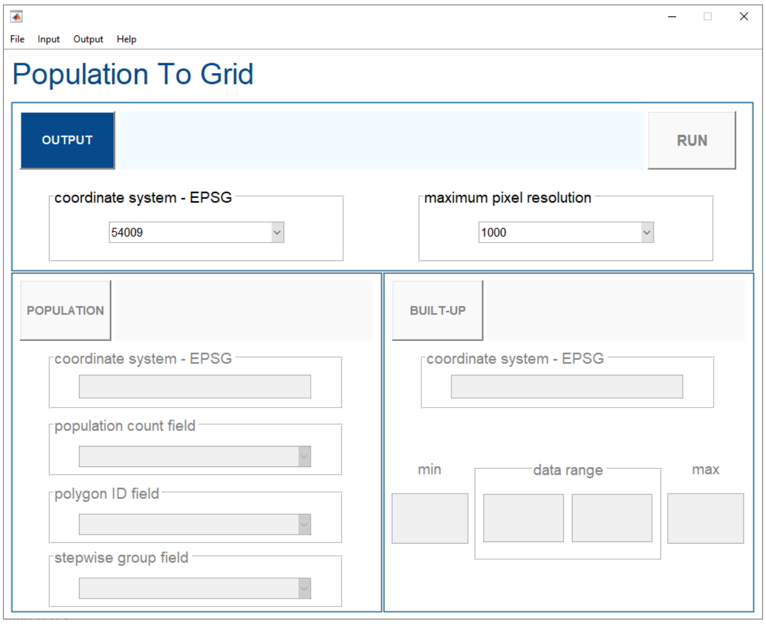
Once the MATLAB Runtime is installed, if you have not done so already, download the POP2G 1.0 tool from the GHSL website (Tools section):

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

Unzip the POP2G.zip file and run the executable directly by double click the *POP2G.exe* file.

N.B. The first time POP2G is launched on a new computer it performs several checks and initialisation steps, therefore it can take up to few minutes to start.

Figure 3. First startup



2.2.3 Remove POP2G

To uninstall POP2G simply delete the entire folder containing the *POP2G.exe* file.

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

3 The graphical user interface

The POP2G graphical user interface is created in MATLAB and distributed as a standalone application.

After the start, the user is guided in the process of input selection, starting from the output folder, continuing with the population and the built-up sections.

3.1 Menu bar

The menu bar is organized as in four sections: File, Input, Output and Help.

3.1.1 File menu

Table 3. File menu bar

Item	Action	Shortcut
Run	Starts the Population Grid computation	CTRL + R
Clear all	Clear all inputs	
Quit	Exit POP2G and close the main window	CTRL + Q

3.1.2 Input

Table 4. Input menu bar

Item	Action	Shortcut
Population	Select and load the Population shapefile	CTRL + P
Built-up	Select and load the Built-up area grid	CTRL + B

3.1.3 Output

Table 5. Output menu bar

Item	Action	Shortcut
Output	Set the output folder	CTRL + O

3.1.4 Help

Table 6. Help menu bar

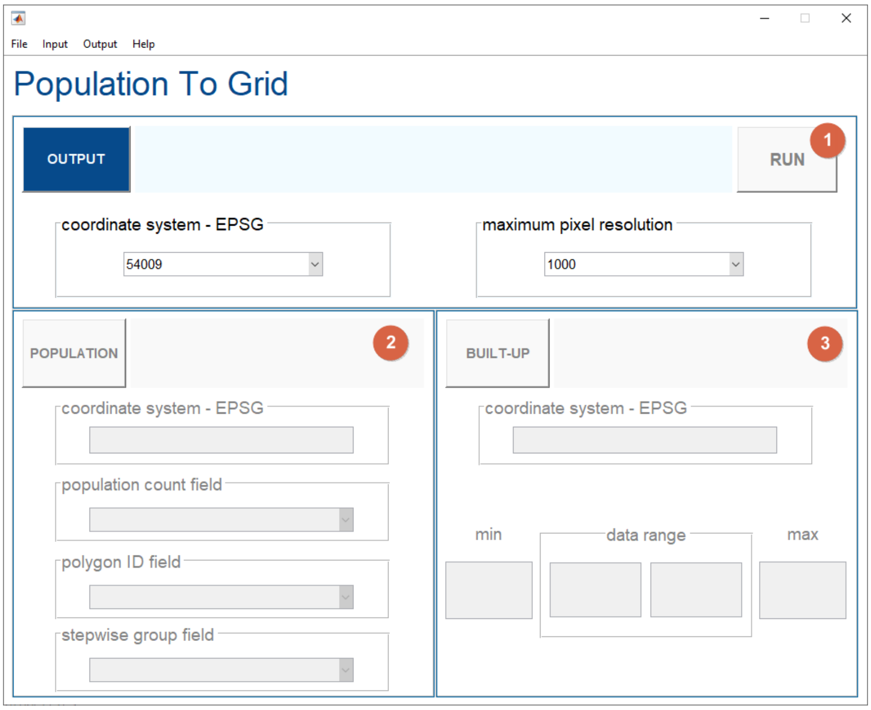
Item	Action	Shortcut
Online help	Open a web page to the POP2G documentation website using the default browser	CTRL + H
About POP2G	Open a new window containing POP2G version number, release date and contact information	

3.2 Control panels

The main window is divided in three areas:

1. Top panel, which contains the buttons to select the output folder, choose the output projection and grid cell resolution, start the process;
2. Left panel, which contains the button to select the shapefile population input layer and the fields to use in the grid computation
3. Right panel, which contains the button to select the built-up raster and set the data rescaling range limits.

Figure 4. POP2G main windows and areas



3.2.1 Output panel

In the output panel the user choose the output folder and the desired output coordinate system and pixel resolution.

Table 7. Output panel buttons

Item	Action	Shortcut
OUTPUT	Set the output folder	CTRL + O
coordinate system - EPSG	Select the output coordinate system / projection using EPSG code	
maximum pixel resolution	Select the output max pixel resolution	
RUN	Starts the Population Grid computation	CTRL + R

3.2.2 Population panel

In the population panel the user choose the vector population file to convert to grid and check the detected coordinate system expressed as EPSG code: if the value is incorrect or missing he can manually change it. He also sets which field to use for the population count.

When the population is represented in polygons instead of points some advanced option can be specified, like the field to use to differentiate the polygons and the one to group the computation in smaller steps.

Table 8. Population panel buttons

Item	**Action**	**Shortcut**
POPULATION	Select and load the Population shapefile	CTRL + P
coordinate system - EPSG	Display the population coordinate system: it can be manually changed if the value is incorrect or missing.	
population count field	Select the shapefile field containing the population count	
polygon ID field	Select the shapefile field to be used as index in the objects rasterization. Used only for polygons. By default is set to AUTO	
Stepwise group field	Select the shapefile field to be used to split the rasterization process in smaller steps (useful with low memory machine). Used only for polygons. By default is deactivated.	

3.2.3 Built-up panel

In the built-up panel the user choose the built-up grid and check the detected coordinate system expressed as EPSG code: as for the population dataset, if the value is incorrect or missing he can manually change it.

There are also two editable text fields 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 (where 0 represents the absence of built-up in cell; and 1 represent a cell fully covered by built-up) in order to have normalized data before starting the population grid computation. The expected min and max values are precomputed internally, but can be adjusted by the user in case the heuristic guessed a wrong value.

Box 1. Example of rescaling

The input built-up layer should range 0 to 100. If the user built-up area raster contains values ranging, for example, from 0 to 80 (see Figure 11), POP2G heuristic choose 0 and 100 as minimum and maximum possible values. 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.

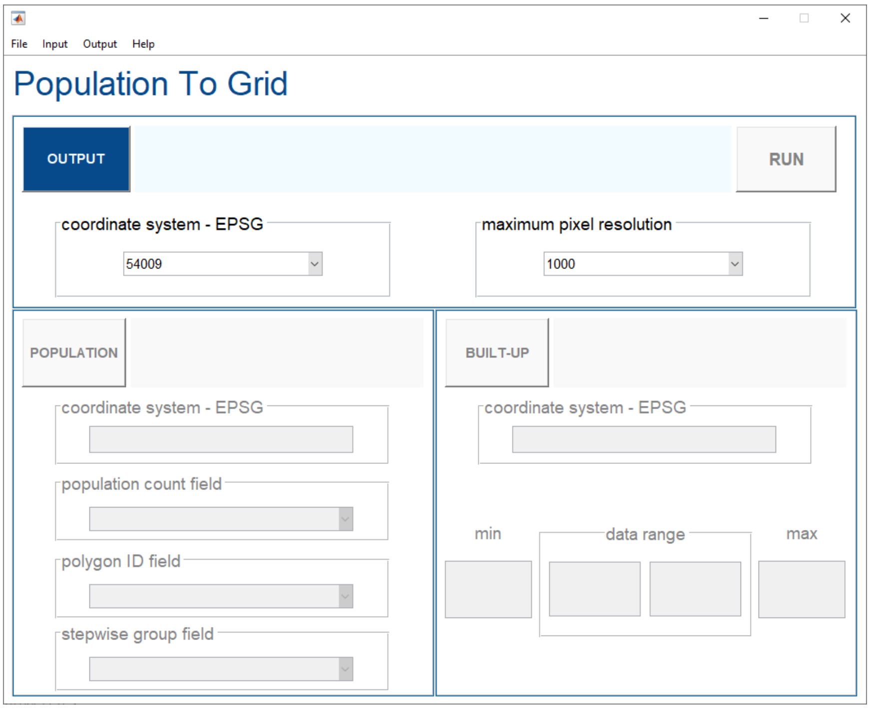
Table 9. Built-up panel buttons

Item	**Action**	**Shortcut**
BUILT-UP	Select and load the Built-up area grid	CTRL + B
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	

3.3 Guided run

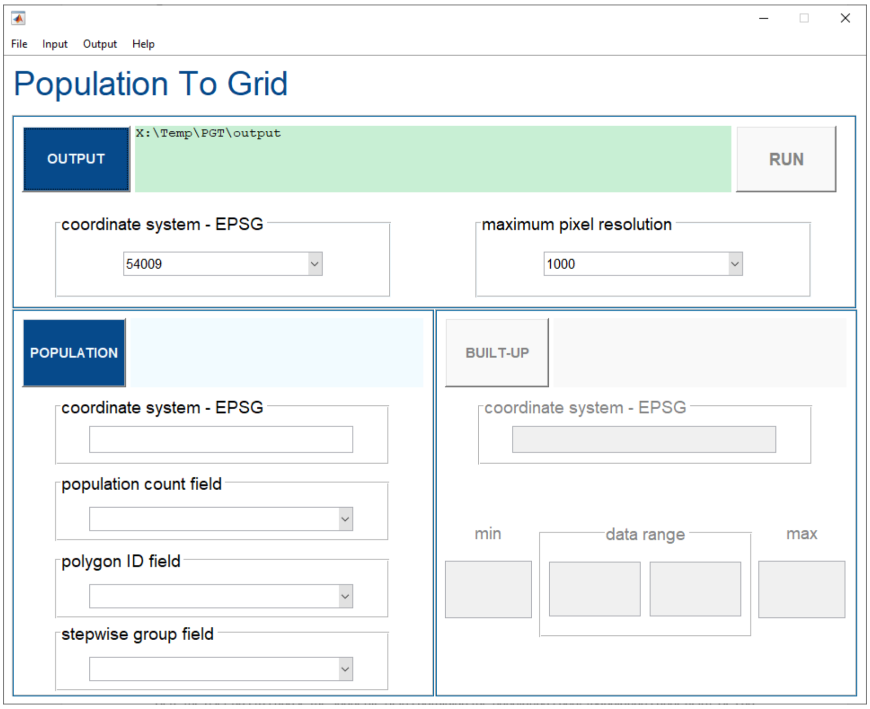
After starting POP2G, the only active command is the output button (Figure 5): this is to set the output folder (path) first in order to create the log file and check disk write permission for results. Below in the shaded italic text lines the actions and commands the user should prompt to run the POP2G Tool.

Figure 5. POP2G startup



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

Figure 6. Output folder selected



Once output is set, the user can choose the output coordinate system and pixel resolution.

The list of supported coordinate systems includes:

- Mollweide projection (EPSG:54009)
- ETRS89 / LAEA Europe projection (EPSG:3035)
- Several UTM zone projections

The supported maximum output grid cell resolution are 1000, 250 and 50 meters, where 50 meters is the most computationally expensive option. If the user select 250 the tool computes to outputs: the grid at 250 and then the results aggregated at 1000 as well. If the user select 50 the tool computes the grid at 50 meters and then aggregate the results at both 250 and 1000 meters resolution.

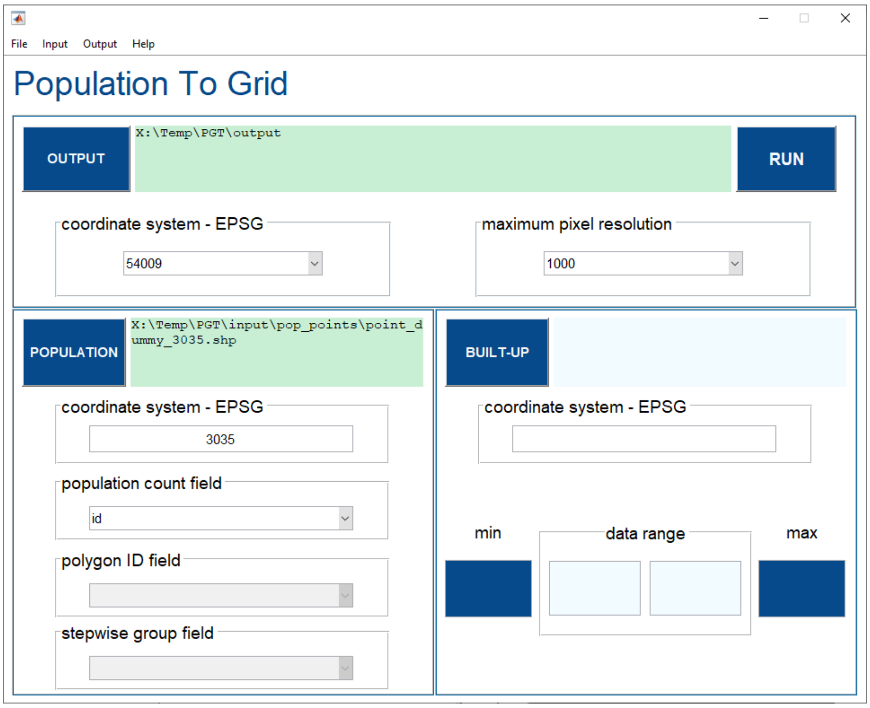
The POPULATION panel become active.

The user chooses the population shapefile. (Figure 7)

Once population is set, the shapefile metadata are read and the pop menu are filled with the shapefile attribute fields' names.

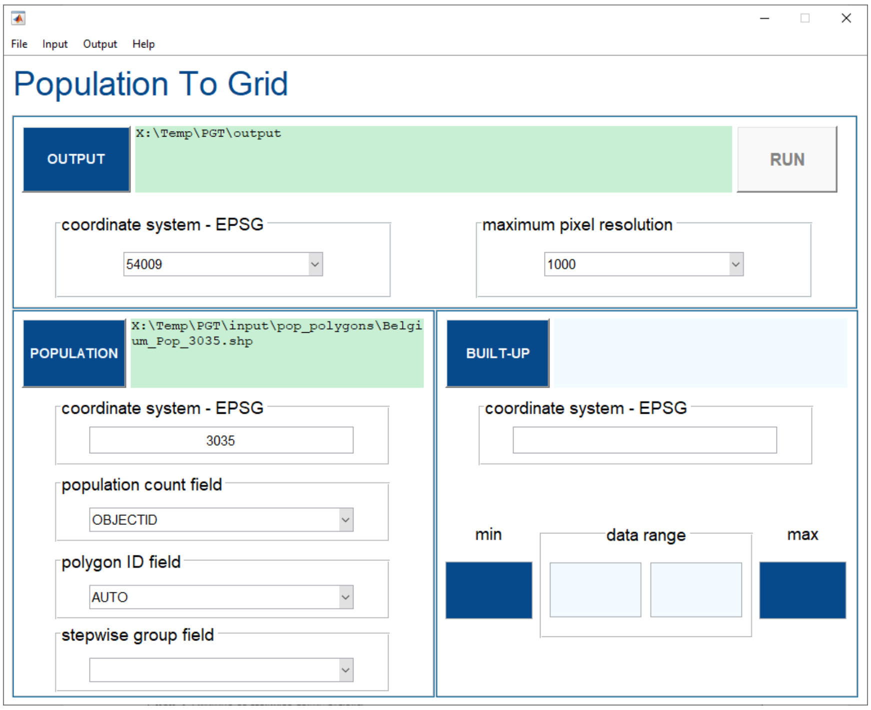
The population shapefile can contains points or polygons. When the population shapefile is made by points the built-up layer is activated but not needed. The user has to choose the shapefile field containing the population count (*population count field*) only and then can start the computation.

Figure 7. Population layer with points



When the population shapefile is made by polygons the user has more options: first he has to choose the shapefile field containing the population count (*population count field*) as before. Then he can also select the field to be used as index in the objects rasterization (*polygon ID field*). By default the *ID field* is set to AUTO, each attribute table row is treated as a single object (the selected field must be a unique identifier of the each table row). The user can also select another field to be used to split the rasterization process in smaller steps (*stepwise group field*) if computing power is not enough to generate the entire grid at once. By default this option is deactivated and the process is run at once on the entire population layer.

Figure 8. Population layer with polygons



Box 2. Example of stepwise computation

Suppose the input shapefile is made of polygons with an overall extent covering an entire country. Each polygon contains the population count and another field indicating it belongs to a country sub region (e.g. county or region). If the computation for the overall country is too demanding the user can choose the stepwise approach which compute the grid in every sub region indicated by the stepwise group field and then aggregate the single results. To ensure a reduction of memory consumption, the selected field should group the polygons in sets having smaller extents. The smaller the groups extents the lower the memory required. In extreme demanding cases the *polygon ID field* could be used to reduce to the minimum the memory consumption of the process.

The user chooses the built-up raster grid

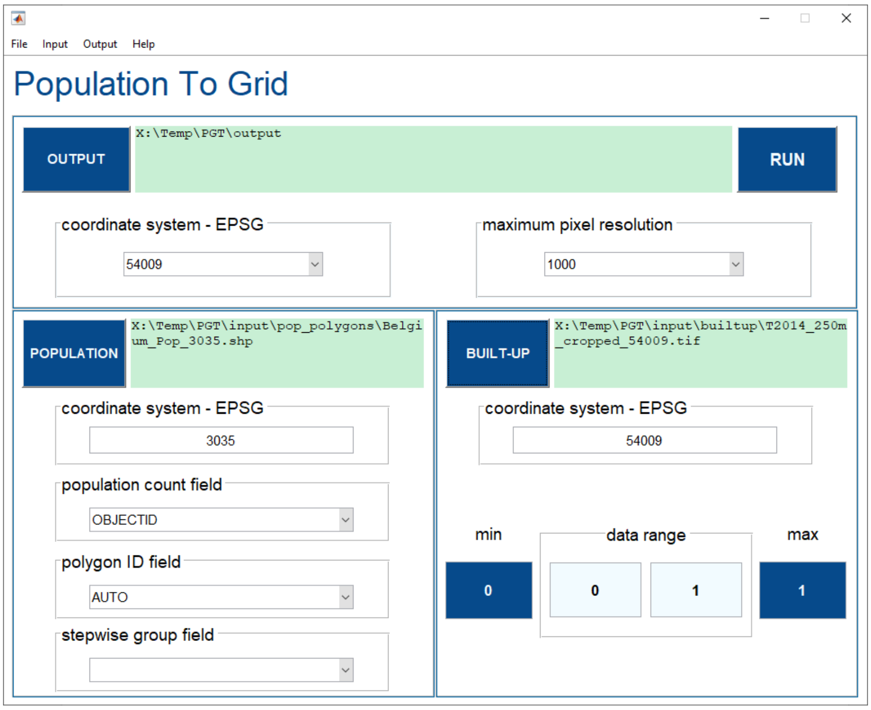
When the population layer is made by points if no built-up area grid is provided the population grid is computed using the shapefile population layer extent and the grid resolution chosen by the user. When the user provide a built-up area grid, the same computation happens but the population grid is aligned with the built-up area grid and the overall extent is adjusted to fully contain the new grid.

The built-up area grid is always needed when the population layers is made by polygons.

Once built-up area is loaded, minimum and maximum raster values are displayed in the data range.

Internal heuristics set 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 – 1 for the example in Figure 9).

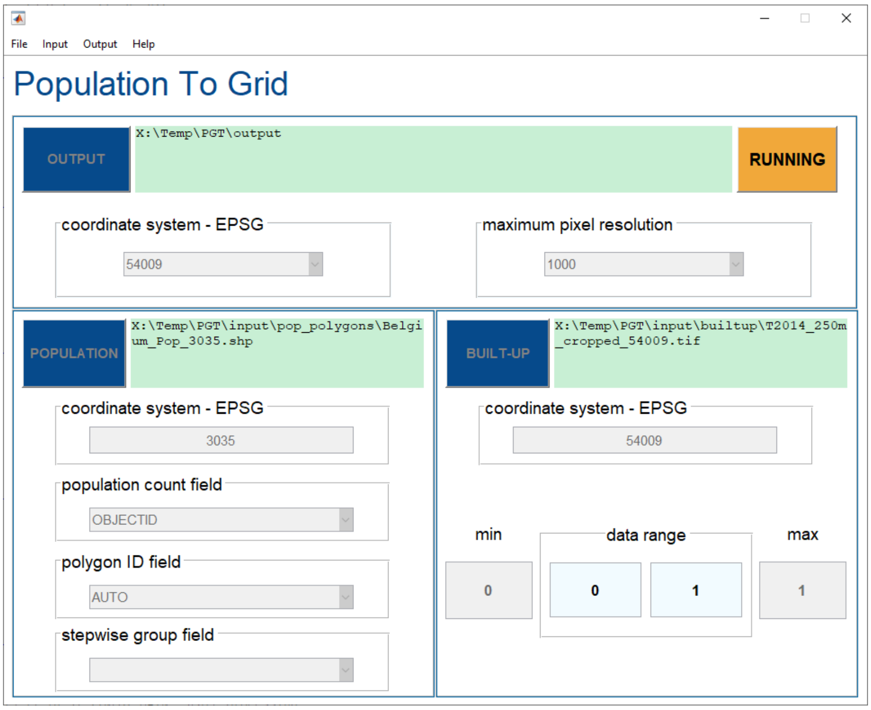
Figure 9. Land and Built-up raster selected



The user hits RUN and start the computation

All commands are disabled when running, the RUN button becomes orange with label RUNNING and the mouse pointer transform into wheel (Figure 10). Once the run completed everything is reactivated again.

Figure 10. POP2G running



4 Input and output files

4.1 Input

The input files requested by the tool depend on the population shapefile geometry: if the population attribute is contained in polygon features, the built-up area raster is always required; if the population attribute is attached to point features, the built-up area raster is optional (Table 10). The built-up area grid in this case is only used to align (snap) the resulting population grid (output) to the built-up area one.

Table 10. Required input

Population shapefile	**Population geometry**	**Built-up grid**	**Operation**
Mandatory	Polygons	Mandatory	Disaggregation
Mandatory	Points	Optional	Aggregation

Both population shapefile and built-up raster can be in any coordinate system supported by GDAL, as they will be re-projected internally by the tool.

Box 3. Supported projections

Check the projection supported by POP2G through GDAL at:

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

4.1.1 Population shapefile

The population shapefile is the mandatory input to compute the population grid. It can contains population data represented as points or polygons but it has to be in the shapefile format and composed of at least these 4 files:

1. .shp — shape format; the feature geometry itself
2. .shx — shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly
3. .dbf — attribute format; columnar attributes for each shape, in dBase IV format
4. .prj — projection description, using a well-known text representation of coordinate reference systems

The shapefile attribute table must contain at least one field with the population count of the given feature (polygon or point) expressed as a numeric value, integer or float.

The shapefile projection can be of any GDAL supported coordinate system.

Table 11. Population shapefile requirements

File format	**Files list**	**Projection**	**Population count datatype**
shapefile	1. .shp 2. .shx 3. .dbf 4. .prj	Any supported by GDAL	Any numeric (integer or float)

4.1.2 Built-up area grid

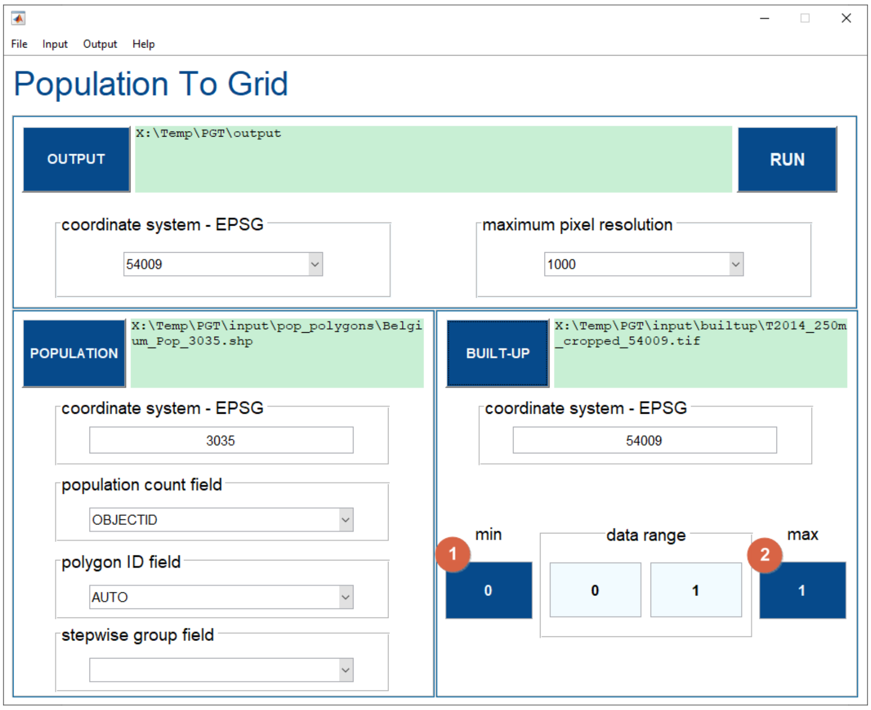
The built-up area grid is required to create the population grid when the input shapefile has polygon geometry. The characteristics of the layer are displayed in Table 12. The built-up area grid shall contain the share of built-up area per grid cell (density), usually expressed as a value in the range 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 12. Built-up area grid requirements

File format	**Data type**	**Projection**	**Pixel size**
GeoTIFF	Any numeric (integer or float)	Any supported by GDAL	Any, it will be resized at the maximum pixel resolution set

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 11).

Figure 11. Built-up area expected min and max



4.2 Output

The output of the POP2G tool is the population grid. Depending on the maximum grid cell size resolution chosen it creates one, two or three grids.

The computation is performed at the higher chosen pixel resolution and then aggregated to the lower ones.

Table 13. Output files

Maximum pixel resolution [meter]	**Output grids**
1000	Pop_grid_1000m.tif
250	Pop_grid_1000m.tif, Pop_grid_250m.tif
50	Pop_grid_1000m.tif, Pop_grid_250m.tif, Pop_grid_50m.tif

5 Conclusion

The POP2G Tool 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 POP2G Tool presented in this report the user can produce population grids at multiple spatial resolutions. 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 assess the feasibility of applying a global definition of cities/urban areas in support of global monitoring of SDGs and the New Urban Agenda urban targets towards the 51st Session of the United Nations Statistical Commission.

References

European Commission, and Statistical Office of the European Union. 2019. Methodological Manual on Territorial Typologies: 2018 Edition. http://dx.publications.europa.eu/10.2785/930137.

European Commission, Joint Research Centre. 2019. Atlas of the Human Planet 2019: A Compendium of Urbanisation Dynamics in 239 Countries. Publications Office of the European Union. 10.2760/445233.

Leyk, Stefan, Andrea E. Gaughan, Susana B. Adamo, Alex de Sherbinin, Deborah Balk, Sergio Freire, Amy Rose, et al. 2019. 'The Spatial Allocation of Population: A Review of Large-Scale Gridded Population Data Products and Their Fitness for Use'. Earth System Science Data 11 (3): 1385–1409. https://doi.org/10.5194/essd-11-1385-2019.

Florczyk, Aneta; Corban, Christina; Ehrlich, Daniele; Freire, Sergio; Kemper, Thomas; Maffenini, Luca; Melchiorri, Michele; Pesaresi, Martino; Politis, Panagiotis; Schiavina, Marcello; Sabo, Filip; Zanchetta, Luigi (2019) GHSL Data Package 2019. JRC117104. Publications Office of the European Union, Luxembourg.

Freire, S., K. MacManus, M. Pesaresi, E. Doxsey-Whitefield, and J. Mills. 2016. "Development of New Open and Free Multi-Temporal Global Population Grids at 250 m Resolution." In *Proc. of the 19th AGILE Conference on Geographic Information Science*. Vol. 250. Helsinki, Finland 14–17 June.

Freire, S., M. Schiavina, A. J Florczyk, K. MacManus, M. Pesaresi, Christina Corbane, Olena Borkovska, et al. 2018. 'Enhanced Data and Methods for Improving Open and Free Global Population Grids: Putting "Leaving No One behind"into Practice'. International Journal of Digital Earth, 1–17. DOI: 10.2760/062975 http://publications.jrc.ec.europa.eu/repository/handle/JRC117104

List of boxes

List of figures

List of tables

KJ-02-20-184-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:10.2760/305540

ISBN 978-92-76-17301-4